Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

A Hand Gesture Recognition System for Deaf-Mute Individuals

Rugia Said Kamaleldeen[1] and Dr. Ebtihal H.G. Yousif[2]

[1] Sudan University for Sciences and Technology

Received: 5 April 2021 Accepted: 4 May 2021 Published: 15 May 2021

Abstract

A deaf-dumb individual always uses gestures to convey his/her ideas to others. However, it is hard for people to understand this gesture language. The purpose of the project is to develop a computer-based system to recognize 26 gestures from American Sign Language (ASL) using MATLAB, which will enable deaf-dumb individuals significantly to communicate with all other people using their natural hand gestures. The proposed system in this project is composed of five modules, which are prepared datasets for ASL which was self-collected using hand gestures from both male and female volunteers, who have alternative ages and skin color in different backgrounds and postures by an ordinary phone camera in total the dataset was 260 images preprocessing, hand segmentation, feature extraction, sign recognition, and text of sign voice conversion. Segmentation is done by converting the image to Hue-Saturation- Value (HSV) format and using color threshold APP. Blob features are extracted by using (BOF) which used the Speed up Robust Features (SURF) algorithm. Furthermore? the K- Nearest Neighbor (KNN) and Support Vector Machine (SVM) algorithms are used for gesture recognition. The Recognized gesture is converted into voice format.

Index terms— bag of feature (BOF), HSV, KNN, SURF, SVM, gesture.

1 Introduction a) Introduction

ommunication between normal and handicapped person such as deaf people, dumb people, and blind people has always been a challenging task. They found it really difficult at times to interact with normal people with their gestures, as only a very few of those are recognized by most people. Since people with hearing impairment or deaf people cannot talk like normal people so they have to depend on some sort of visual communication in most of the time. Sign Language is the primary means of communication in the deaf and dumb community [1]. The deaf people become neglected by society because ordinary people never try to learn ASL nor try to interact with the deaf people. They mostly remain uneducated and isolated. So the only way to enhancing the communication between mute people and ordinary people is recognition of sign language and converting it to the corresponding voice [2]. Sign language recognition developed in the ”90s.

Research related to hand gestures classified into two parts. In the first part, electromagnetic gloves and sensors are introduced, which consist of the hand shape, movements and, orientation of the hand. These have limitations such as cost and not suitable for practical use. The second one is a computer visionbased gesture recognition system, which consists of image processing techniques, which required only a camera and computer or mobile device like phones or tablets, which are very common among people [3].

2 b) American Sign Language

American Sign Language (ASL) is a complete, complex language that employs signs made by moving the hands combined with facial expressions and postures of the body. It is the primary language of many North Americans who are deaf and is one of several communications. [4]

3 c) Proposed Model

The dataset used in the proposed model was collected and created by ourselves by using phone camera without perfect aligned with camera. Firstly, the system will take the dataset images and apply skin detection algorithm on it and detect the skin color pixels from it. Then it will make the image binary. Secondly for feature extraction and evaluation we are using Bag of features in category classification. It splits the image grid by grid and takes number of image patches from it. The strongest features are identified by using SURF algorithm and K-means clustering is used for vector quantization. The features or bag of words are stored in the feature vector After that for multi-class SVM (Support Vector Machine) classifier has been used for categorize training and testing set for evaluation. Thirdly, the system will take the tested image and apply skin detection algorithm on it for purpose of skin detection, using SURF algorithm to extract features and K-mean clustering for vector quantization finally it was compared with features in dataset to find the corresponding letter. This proposed model is shown in figure.1 below.

4 Medical Research

5 ii. Quantization and Distance Measures

Vector Quantization (Clustering) is used to build the visual vocabulary in Bag of Features algorithms. Nearest-neighbor assignments are used not only in the clustering of features but also in the comparison of term vectors for similarity ranking or classification. Many BOF implementations are described as using K-means cluster.

With K-means clustering process we want to minimize sum of squared Euclidean distances between points xi and their nearest cluster centers mk. [5] Algorithms:

1. Select initial centroids at random, 2. Assign each object to the cluster with the nearest centroid, 3. Compute each centroid as the mean of the objects assigned to it, 4. Repeat previous 2 steps until no change.

6 d) Bag of Feature

The past decade has seen the rise of the Bag of Features approach in computer vision. Bag of Features (BOF) methods have been applied to image classification, object detection, image retrieval and even visual localization for robots. iii. Feature Classification Support Vector Machine (SVM) is primarily a classifier that performs classification tasks by constructing hyper planes in a multidimensional space separating cases of different class labels. According to SVM the decision boundary should be as far away from the data of both classes as possible. The linear separating hyper plane is the plane II.

7 Bag of feature

8 Literature Review

In [7], nine gestures are recognized and converted to speech and text in real-time by using MATLAB. YCbCr color transformation used for feature extraction and, PCA algorithm has been used for recognition to captured image using a web camera. In [8], PCA algorithms also have been used to recognize 26 gestures from Indian sign language, morphological filter and, outs algorithms for segmentation to get comparable accuracy. A comparison has been made between using web camera and traditional image processing techniques against the android devices and PCA algorithms the first method is more accurate (90) But it takes time and memory the second method is faster and need less memory but has low accuracy(77) [9]. In [10] does not require the background to be perfectly black to recognize sign language. Image preprocessing, calculating coordinate for feature extraction and, finally the pattern matching algorithm for classifying purpose used to recognize. In terms of machine learning-based approaches, Abdo et al. [11] employed Hidden Markov Model (HMM) for Arabic alphabet and numbers sign language recognition. This system is suitable and reliable compared with other competitive systems but the limitations of this system are required users to wear a color glove. Also, in the study by Dogic and Karli [12], sign language recognition has been applied with accuracy 84; the work has been doing with the use of digital image processing methods providing a system that teaches a multilayer neural network using a back propagation algorithm. Images processed by feature extraction methods (canny edge detector), and by masking method, the data set has been created. Training has been doing using the crossvalidation method for better performance.

Video processing used to translate Real-time Arabic sign language to Arabic text. For example, the method used in [12] includes video segmentation (shot boundary detection, keyframe extraction), pattern construction and discrimination, and feature extraction; the extracted features are intensity histogram and Gray Level Co-Occurrence Matrix (GLCM). To identify English alphabetic sign language without requiring the hand to be perfectly aligned to the camera, an image processing technique (the detection of skin and marker pixels) has been used in [14]. For the purpose of segmentation, the threshold used finally to extract feature, and recognition, the coordinate calculation, color calibration and pattern matching algorithm used ,this system is easy and high accuracy but it requires users to wear the specific color band in their fingers.

9 III.

10 Methodology a) Data collection

The sign language recognition is not a widely researched topic; so we did not find any dataset on any resources. Therefore, we made our dataset. We took images of hands for males and females of vary ages and with alternative positions also different in backgrounds using a camera phone and without perfectly aligned with the camera. We acquired images for 26 alphabets of American Sign Language.

We took images of 10 people's hands in alternative positions. So, in total in the dataset, we have 260 images. For each alphabet, we are getting ten pictures of ten people in different postures and different ages. as shown in figure [3]

11 b) Skin Color Detection Algorithm

Skin detection means detecting those pixels and regions from an image should be contain human skin tone color in a picture. The use of color information as a feature for skin detection enables fast processing and brings robustness to such application [15] Skin color detection is applied to the input image for the detection of hand gestures. This technique is used for separating the skin-colored areas from the non-skin colored regions. The steps used in this skin color detection algorithm are shown in Figure [4]. RGB image converted to HSV color space which used because it is more convenient for research purposes. Conceptually, the HSV color space is a cone. Viewed from the circular side of the cone, the hues are represented by the angle of each color in the cone relative to the 0° line, which is traditionally assigned to be red. The saturation is represented as the distance from the center of the circle. Highly saturated colors are on the outer edge of the cone, whereas gray tones (which have no saturation) are at the very center. The brightness is determined by the colors vertical position in the cone. At the pointy end of the cone, there is no brightness, so all colors are black. At the fat end of the cone are the brightest colors. As hue varies from 0 to 1.0, the corresponding colors vary from red through yellow, green, cyan, blue, magenta, and back to red, so that there are actually red values both at 0 and 1.0. As saturation varies from 0 to 1.0, the corresponding colors (hues) vary from unsaturated (shades of gray) to fully saturated (no white component). As value, or brightness, varies from 0 to 1.0, the corresponding colors become increasingly brighter. The conversion of RGB to HSV is given by the following equations:

People have different color skin so, the value of H, S, and V are different from person to person depending on his color skin; the images in the dataset was taken from various people have alternative color skin, the value of H, S, and V were set depending on the color skin of individual, according to these values the hand was segmented from background after that the images was converted to black and white format by using color threshold APP, this process is shown in figure. [1]. We found the more images in the training set, the more accuracy we have. However, for technical constraints, we cannot increase the number of images more than ten in each class. Consequently, the total number of images in the training set is 260. as shown in table [2] below Table [1]: Comparison between Different Percent

12 iii. Step 3: Training an Image Classifier with Bag of Visual Words

The Train Image Category Classifier Function returned as an image classifier. The algorithm trained a multi-class classifier using the error-correcting output codes (ECOC) framework with binary support vector machine (SVM) classifiers. iv.

Step 4: Classifying an Image or Image Set Finally the Image Category Classifier predicts method is used on the tested image to determine its category. After the BOF has been used to determine the class of picture, the letter which corresponding to the image appears in the workspace,

13 d) Speech Synthesizer

Finally, the letter is converted to the voice by using Speech Synthesizer.

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called a speech synthesizer and implemented in software or hardware. A text-to-speech (TTS) system converts language text into speech; other systems render symbolic linguistic representations like phonetic transcriptions into speech [16]. In this project, we used one which is included in computer operating systems and implemented in MATLAB.

14 e) Graphical user interface

A graphical user interface (GUI) is a pictorial interface to a program. A good GUI can make programs easier to use by providing them with a consistent appearance and with intuitive controls like pushbuttons, list boxes, sliders, menus, and so forth. The GUI should behave in an understandable and predictable manner, so that a user knows what to expect when he or she performs an action. [17] Finally, to make the system friendlier with the user it is converted to a graphical user interface was implemented using MATLAB2016 as shown in Figure [6].

IV.

15 Result a) Evaluating the Model

Train set average accuracy after training the SVM classifier with the train set, we evaluated the trained classifier, on the train set and got 89 percent average accuracy.

16 b) Testing the Model

After training the classifier we tested the program on ten people, so there are 260 images, ten images in each class.

17 i. Determining the Number of Images in Training Set

When ten images have been used, we got the largest accuracy in all class (84.6). When seven images used, the accuracy lessoned in all categories (59.2). When five images used, the accuracy decreased in 17 categories and increased in 9 categories (57.7) it is clear from the above discussion, and as shown in figure [7] and figure [8], the more images in the training set, the more accuracy we have. The average accuracy for the model we received is approximately eighty-five percentages when using ten images for each class in the training set as shown in figure [9] and figure [10].

18 Conclusion

Communication with an ordinary person is always a challenging task for a dumb person. In this research, a system called Sign language recognition is introduced, which is an effective communication aid for a deaf person. It is convenient, comfortable and, cheap; there is no need for wearable to use the device. The system extended to aid the deaf in communication and users friendly with an accuracy of 84.6%.

19 VI.

20 Recommendations

For enhancement of this study, the database can be expanded the numbers of photo in the dataset by more than ten for every letter and take the photos for dataset from more expert people in dumb and deaf language. Also, for further future enhancement, instead of using the graphical user interface, the system can converted to application phone and connect between camera and application to make it easier and friendlier it can convert to real-time application by using video processing. [1]

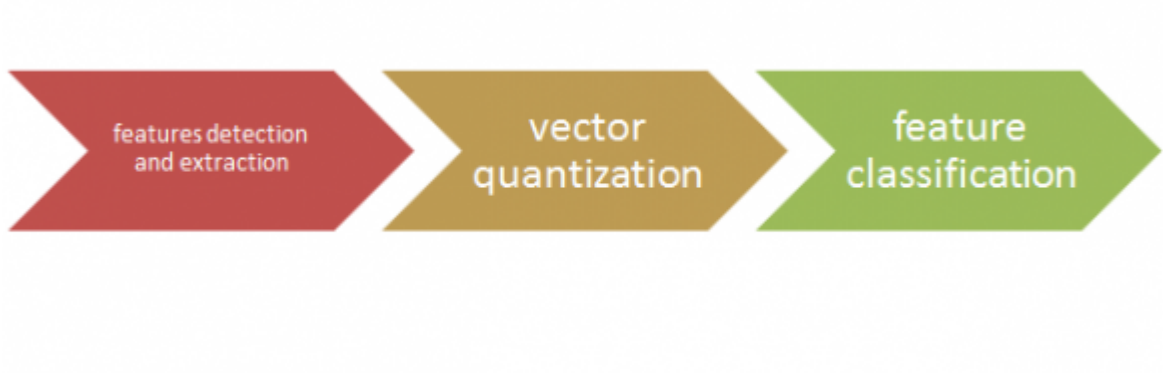
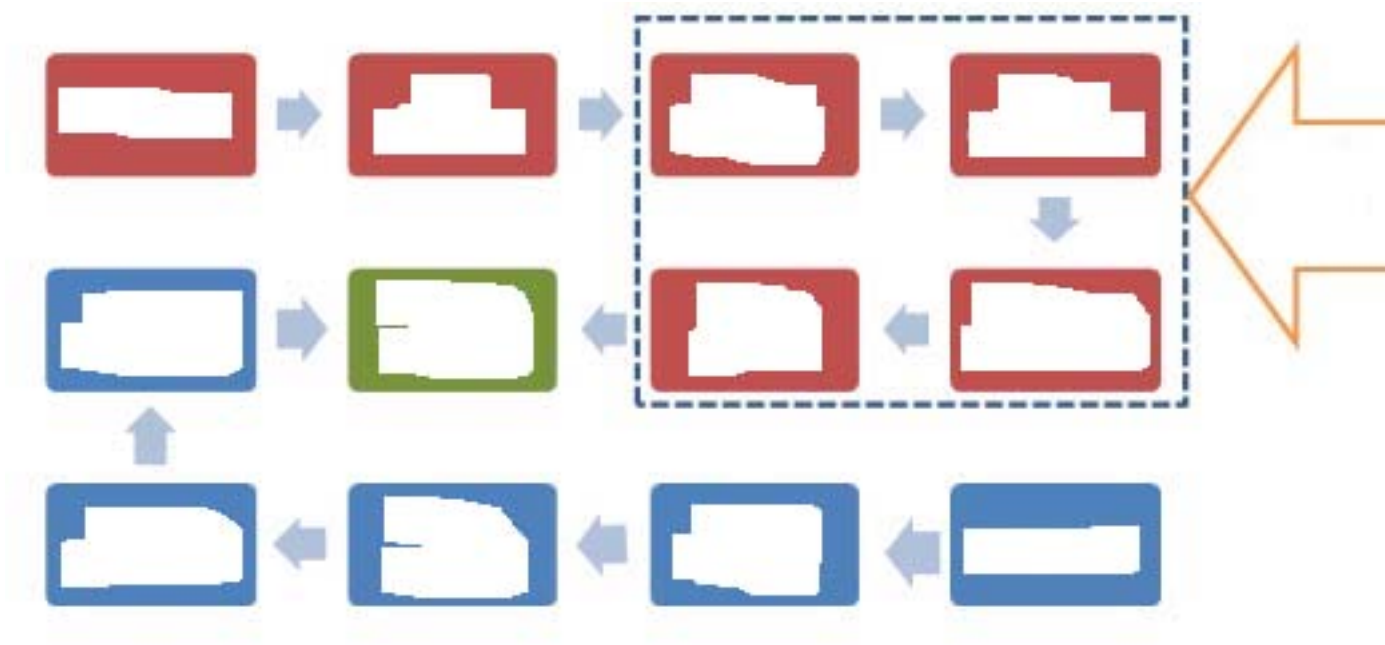


Figure 1: A



2

Figure 2: Figure 2 :

Figure 3:

1

Figure 4: Figure 1 A

3

Figure 5: Figure 3 :A

Figure 6:

4

Figure 7: Figure 4 :A

5

Figure 8: Figure 5 :

678

Figure 9: Figure 6 :Figure 7 :Figure 8 :

9

Figure 10: Figure 9 :

1011

Figure 11: Figure 10 :Figure 11 :

[

greater scale invariance, SURF algorithm used as a detector and descriptor. s	
Percent Accuracy	
90	0.89
70	0.87
50	0.65
30	0.56
Total Class of images	26
In one class number of images	10
Total images	260
For training the image classifier number of images	234
For testing and evaluation number of images	26

from training Sets. The algorithm iteratively groups the descriptors into k mutually exclusive clusters. The resulting clusters are compacted and separated by similar characteristics. Each cluster center has been represented a feature or visual word. Speeded up robust features (or SURF) detector is used that is provides

Figure 12: Table [2

[American Sign Language (2014)] , *American Sign Language* 4756 February 2014. May 2015. NIH Publication. (11) . U.S Department of Health and Human Services. National Institute of Health (Reprinted)

[Naoreen (2014)] 'A systematic survey of skin detection algorithms, applications and issues'. Naoreen . *Journal of Applied Environmental and Biological Sciences* Sep 2014. 1.

[Gautam and Kaushi (2017)] 'American sign language recognition system using image processing method'. K Gautam , A Kaushi . *International Journal on Computer Science and Engineering (IJCSE)* Jul 2017. 9 (7) .

[Jagdish et al. ()] *Android based portable hand sign recognition system*, S Jagdish , L Raheja , Sadab . 2015. Dept. of Computer Science Truman State University USA

[-R et al. ()] 'Arabic alphabet and numbers sign language recognition'. S A E S -R , Alaa Mahmoud Mahmoud Zaki Abdo , E M Hamdy , Saad . *IJACSA) International Journal of Advanced Computer Science and Applications* 2015. 6 (11) .

[Bhaskar Anand et al. ()] 'Face Recognition using SURF Features and Classifier'. Bhaskar Anand , Mr , Prashant , Shah . *International Journal of Electronics Engineering Research* 2016. 8 (1) .

[Allen et al. ()] *From Text to Speech: The MITalk system*, Jonathan ; Allen , M Hunnicutt , ; Sharon , Dennis Klatt . 1987. Cambridge University Press.

[Kakde and Rawate (2016)] 'Hand gesture recognition system for deaf and dumb people using pca'. M U Kakde , A M Rawate . *International Journal of Engineering Science and Computing* July 2016. 6.

[Pramada et al. ()] 'Intelligent sign language recognition using image processing'. S Pramada , D Saylee , N Pranita , N Samiksha , M Vaidya . *IOSR Journal of Engineering (IOSRJEN)* 2013. 3 (2) p. .

[Ashi and Ameri] *Introduction to Graphical User Interface (GUI) MATLAB 6.5*, R Y A Ashi , A A Ameri . UAE University College of Engineering Electrical Engineering Department

[O'hara and Draper (2011)] *Introduction to the bag of features paradigm for image classification and retrieval*, S O'hara , B A Draper . arXiv:1101.3354. Jan 2011. 1. (arXiv preprint)

[El-Alfi et al. ()] 'Real time arabic sign language to arabic text &sound translation system'. A El-Alfi , R El-Gamal , El-Adly . *Int. J. Eng* 2014. 3 (5) .

[Shinde and Autee (2016)] 'Real time hand gesture recognition and voice conversion system for deaf and dumb person based on image processing'. S S Shinde , D R Autee . *International Journal of Research Publications in Engineering and Technology* Sep 2016. 2.

[Sawant and Kumbhar ()] 'Real time sign language recognition using pca'. S N Sawant , M Kumbhar . *2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies*, 2014. IEEE. p. .

[Singha and Das (2013)] 'Recognition of indian sign language in live video'. J Singha , K Das . *International Journal of Computer Applications* May 2013. 70 p. .

[Mayuresh Keni and Marathe (2013)] 'Sign language recognition system'. S M Mayuresh Keni , A Marathe . *International Journal of Scientific and Engineering Research* Dec 2013. 4.

[Dogic and Karli ()] 'Sign language recognition using neural networks'. S Dogic , G Karli . *TEM Journal* 2014. 3 (4) .