Original Research Paper

OFFLINE GESTURE RECOGNITION SYSTEM FOR YORÙBÁ NUMERAL COUNTING

ABSTRACT

Aims: To determine the specific features responsible for recognition of gestures, design a computational model for the process, implemented the model and evaluated its performance.

Place and Duration of Study: Department of Computer Engineering, Federal Polytechnic, Ede, between August 2017 and February 2018.

Methodology: Samples of hand gesture were collected from the deaf school. In total, 40 samples containing 4 gestures for each numeral were collected and processed. The collected samples were pre-processed and rescaled from 340 X 512 pixels to 256 X 256 pixels. The samples were examined for the specific characteristic responsible for recognition of gestures using edge detection and histogram of oriented gradient as feature extraction techniques. The model was implemented in MATLAB using Support Vector machine (SVM) as its classifier. The performance of the system was evaluated using sensitivity, recall and accuracy as metrics.

Results: It was observed that the system showed a high classification rate for the hand gestures considered. For numeral 1, 3, 5 and 7, 100% accuracy, numeral 2 and 9 have 90% accuracy, numeral 4 has 85.67% accuracy, numeral 6 has 93.56%, numeral 8 has 88% while numeral 10 recorded 90.72% accuracy. An average recognition rate of 94% on tested data was recorded over a dataset of 40 hand gestures.

Conclusion: The study has successfully classified hand gesture for Yorùbá Sign language (YSL). Thus, confirming that YSL could be incorporated into the deaf educational system. The developed system will enhance the communication skills between hearing and hearing impaired people.

1. INTRODUCTION

Sign language is an avenue through which hearing-impaired people communicate. Sign language uses gesture instead of speech to convey messages and meaning [7]. The design of the gesture recognition system has become an important part of everyday life among the deaf people and their quality is of primary importance for numerous applications. The movement of part of the body especially the hand or the head in a view to express an idea or meaning mostly refers to gesture. Gesture can originate from any bodily motion or state but usually from the face and hand. This study developed a gesture recognition system for Yorùbá numerals. This is with the view to make the computer understand hand gesture demonstration and develop a user-friendly system for communication flow between hearing and hearing impaired people. There are multitudes of sign language in the African continent which evolved outside the context of the deaf educational system. These include Yorùbá sign language (YSL), Hausa sign language and other indigenous sign languages. YSL is known to be an indigenous sign language found in the community of deaf people in Yorùbá speaking area of Southwestern Nigeria. The speaking of fluent Yorùbá language requires a large amount of gesture demonstration [8] while communicating, this allows minimal communication between the deaf and the hearing. YSL incorporates many Yorùbá gestural signs and many mouthing of Yorùbá words [8]. This sign language is yet to be incorporated into the deaf education system and therefore depriving the deaf people the opportunity to learn Yorùbá language. This act has restricted the students to a certain language in their educational system. Also, most of the work in sign language has focused on the recognition of American Sign Language (ASL) because of its widespread use and thus less complexity is required [7]. The YSL are more complex as it involves double hands and even foot in some cases. This study developed a gesture recognition system for Yorùbá numeral counting from 1 to 10.

2. RELATED WORK

Various gesture recognition system involving hand have been reported to date. A number of techniques used includes data gloves, stereo camera, time of flight (TOF) camera and Microsoft's Kinect sensor [12]. Basically two main methods have been identified in sign language recognition [1] and each aim at recognising gestures. These two approaches are glove-based and vision-based of which the former requires the signer to wear a colour or sensor glove for the segmentation process. The vision-based approach employs image processing techniques for the recognition process. These two approaches are mostly used by many authors who worked in the gesture recognition area. The data glove approach has recorded a better performance than some other systems and the limitations of its popularity are due to expensive and uncomfortable nature of its use [2]. The wearer of the data glove is roughly affected due to the installation of highly intense sensors [13] and also one person cannot use the glove of another person [6].

The potential of Kinect depth-mapping was examined by [15] for the recognition of sign language. A total of 1000 American Sign Language (ASL) were used and Hidden Markov Model was employed to recognize the gestures. The work of [5] used the colour glove to identify finger-tips and joints of Pakistani sign language. The work employed fuzzy logic as its classifier and the calculated angle between the finger-tip are input into the fuzzy inference system for the recognition process. The drawback of this method is the wearing of the colour glove during the process and the fuzzy logic used suffers from the problem of a huge number of rules required to accommodate all the features of the gestures [1]. In another work of [4], the author employed vision-based technique and applied edge detection as a feature extraction technique and the classification of the gestures was achieved using MATLAB if-then rule. The recognition process was limited to only 6 sample images of English alphabets. The vision-based approach reported by [10] used colour marker to develop an intelligent sign language recognition for English alphabets.

The method is exposed to an accuracy problem because of difficulty in differentiating the skin and the marker from the background during image processing. More attention has been given to American Sign Language as found in [13] with the review of vision-based ASL recognition. Similar work found in [7] discussed various Indian Sign Language (ISL) recognition and the limitation and strength of most of the work reviewed were clearly stated for further improvement. In another work of [11], the hand gesture recognition system on a geometric model was developed for the Arabic language. The system extracts geometric features from the hand gesture using only colour as a specific characteristic which limits the efficiency of the work for large datasets. The real-time gesture recognition [3] is an improvement over static gesture. The author recognized 37 signs of ASL using Artificial Neural Network with feed forward, back propagation algorithm and accuracy of 94.32% was recorded. Also in [9], edge orientation histogram technique was adopted and accuracy of 88.26% was obtained

The reviewed work have used gestures of various languages and it was shown that less attention has been given to YSL. This study, therefore, developed a hand gesture recognition system for Yorùbá numerals. This is with the view to enhance communication skills between hearing and hearing-impaired people in Yorùbá communities.

3. MATERIALS AND METHODS

The materials and methods used in formulating the recognition models are presented in this section.

3.1 Recognition Model

The recognition process was divided into stages which involved the feature extraction and classification stages. The feature extraction stage includes the pre-processing operation where the features were extracted using the histogram of oriented gradient and edge detection techniques. The classification stages were achieved using Support Vector Machine (SVM) classifier. Classification problem involves associating an object with existing classes of objects. Classes may be defined as a set of objects, or by a set of rules, which define how objects are classified into given classes. In this study, the classifications of numerals are considered according to their respective gesture. Figure 1 depicts a flowchart representing an automated gesture recognition model. Captured gesture images are input to this model. The models first convert the image to grayscale and the images were pre-processed using various image processing techniques, then the edge was detected using canny edge

detection algorithm for segmentation process and the feature vectors were extracted using Histogram of Oriented Gradient (HOG), the output of the extracted feature was trained using SVM classifier and the corresponding gesture type was recognised.

3.2 Image Processing

For training and recognition purpose, different samples of hand gesture were collected with a clear background. In total, 40 samples containing four (4) different images for each gesture were collected from the deaf school in Osogbo, Osun state. The samples were captured using a 14.1 Megapixels digital camera and after data capture, the samples were pre-processed and data were rescaled and normalised to reduce redundancy. The pre-processing operation reduces noise and redundant data for the purpose of acquiring an optimum quality image that enhanced the recognition performance. The contrast and brightness of images were changed to enhance the appearance of the image. The colour of the images was modelled and processed in a digital domain. The images were then compressed using JPEG tool to reduce the storage required to save the image. The collected samples were saved in Tagged Image File Format (TIFF) with an image size of 340 ×512. The captured images were converted from RGB (Red, Green and Blue) colour components to a grayscale image in order to extract the discriminatory information about the image and then thresholding using Otsu's method in MATLAB. Figure 2 depicts the sample data of hand gestures collected for the recognition process.
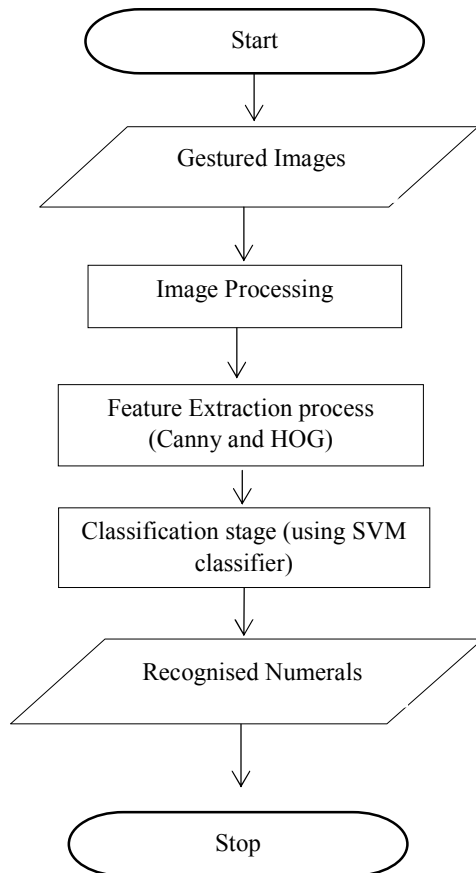

Fig.1. Hand Gesture Recognition Model

Fig.2. Samples of Hand Gesture

3.3 Feature Extraction Stage

Feature extraction is an essential step of sign language recognition. This is because the classifier accepts feature vectors obtained from the feature extraction process as an input. The techniques used for feature extraction extract the distinguishing characteristics about the gesture images. Feature extraction method used in this study serves two purposes, (1) to extract properties that can identify the number uniquely (segmentation process) and; (2) to extract features that differentiate between each gesture captured. It is an essential stage in sign recognition as its function improves the recognition rate and reduces the misclassification. Canny edge detection was used to obtain edges of the gesture images and histogram of oriented gradient was employed to obtain the feature vector that was used to train the classifier.

3.3.1 <u>Canny Edge Detection</u>

Canny edge detection is a Gaussian-based technique that searches for the zero crossing in the second derivatives of an image. Canny edge detection method was employed for the segmentation process. This is with the view to extracting features that uniquely identify the gesture images. The algorithm used involved the following steps, (1) a Gaussian filter was applied to smoothen the images so as to obtain a better-segmented edge. (2) The gradient of magnitude and orientation was computed using a finite-difference approximation for the derivatives and; (3) Non-maxima suppression was applied to the gradient magnitude. The image smoothening was computed as input $I_{i,j}$ image and the Gaussian filter was applied which are denoted as G [i, j, σ] where σ is the distribution of the Gaussian and it also controls the degree of smoothing. The result obtained from the convolution of $I_{i,j}$ with G [i, j, σ]] gives a set of smoothed data as shown in Equation 1

.

$$S_{i,j} = G[i,j,\sigma] * I_{i,j} \qquad (1)$$

The gradient was calibrated using the smoothed array $S_{i,j}$ which was then used to generate X and Y partial derivatives $X_{i,j}$ and $Y_{i,j}$ respectively as shown in Equation 2 and 3.

$$X_{i,j} = (S_{i,j+1} - S_{i,j} + S_{i+1,j+1} - S_{i+1,j})/2 \qquad (2)$$

$$Y_{i,j} = (S_{i,j} - S_{i+1,,j} + S_{i,j+1} - S_{i+1,j+1})/2 \qquad (3)$$

The X and Y partial derivatives were derived by averaging the finite difference over 2×2 matrix. The magnitude and orientation of the gradient were computed using Equation 4 and 5 respectively.

$$M_{i,j} = \sqrt{(X_{i,j})^2 + (Y_{i,j})^2} \qquad (4)$$

$$\theta_{i,j} = \tan^{-1}(\frac{Y_{i,j}}{X_{i,j}}) \qquad (5)$$

The pixel values were chosen based on the angle of the magnitude of the pixel and its surrounding neighbourhood. The Non-maxima suppression was evaluated using the magnitude of the image array. The situation with this approach is that the low threshold generates false edges, but a high threshold misses important edges. The threshold value was moderately selected to obtain a better edge orientation for the recognition process. This is because canny edge detection has a problem with the fixed value for thresholding.

3.3.2 Histogram of Gradient

The histogram of oriented gradients descriptor was used for this study because of its high-performance record in human detection. The local gesture appearance and shape of the image were described by the local intensity changes obtained from the canny edge detection algorithm and the image was divided into small connected regions called cells. The pixels around each cell were used and the histogram of gradient directions was compiled as a feature vector which was later used for the training of the datasets. Figure 3 shows the sample of the image obtained after feature extraction process for numeral one.
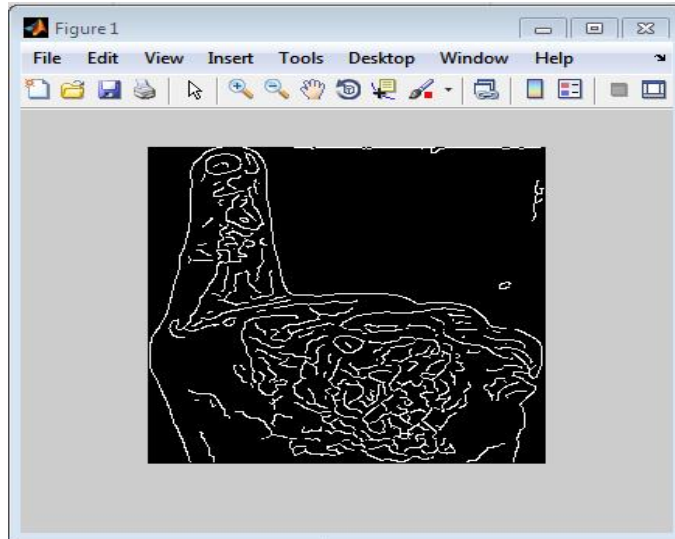

Fig. 3. Feature vector for Hand gesture 1

3.4 Classification Stage

Image classification refers to the process of assigning pixels to an ordered set of related categories in which data are categorised according to its similarities. Image classification step categorises recognised objects into predefined classes by using the suitable algorithm that compares the image patterns with the target patterns. Support Vector Machine (SVM) was employed as a classifier and it is one of the best machine learning algorithms for image classification. The supervised learning algorithm serves to analyse the trained data and gives an inference called classifier. The output extracted from the feature extraction steps was used for the classification process, which classified the hand gestures to their corresponding numerals. For the purpose of the classification process, 60% of the data were used to train the model and 40% were used to test the model. Figure 4 shows the implementation stage depicting the training and testing of the hand gestures recognition. Figure 4(a) and Figure 4(b) show the training and testing stages respectively. For the training stage, the extracted images were fed into the support vector machine and the training variables were saved. For the testing stage, the input image from the testing sets was pre-processed and compared with the saved training variables then output the result.

4.0 RESULT AND DISCUSSION

The hand gesture classification model was simulated in MATLAB R2015a environment. Sample of hand gesture data were rescaled from 340 ×512 to 256 ×256 pixels were processed on an Intel 1.90 GHz and 4GB RAM machine using the model developed in this study. A total of 40 samples containing 4 gestures for each numeral were collected and processed in MATLAB R2015a environment. MATLAB was adopted for programming because of its user-friendliness and scalability it provides. Figure 5 shows the classification result for numeral 1and 2 respectively. Figure 5 is the system design interface which displays the gesture recognition of Yorùbá numeral. The effectiveness of the algorithm used in this study is measured by inputting sample hand gesture images into the model and its performance was recorded. The performance metrics used are recall rate, precision rate and accuracy which are defined in Equation 1, 2 and 3 respectively. Using the metrics defined in Equations above, the classification efficiency of the model was recorded. The accuracy of the model was recorded for each gesture classified. Table 1 shows the classification rate for the YSL for numeral 1 to 10. For numeral 1, 3, 5 and 7, 100% accuracy, numeral 2 and 9 had 90% accuracy, numeral 4 has 85.67% accuracy, numeral 6 has 93.56%, numeral 8 has 88% and numeral 10 which has 90.72%. An average recognition rate of 94% was obtained over tested data.
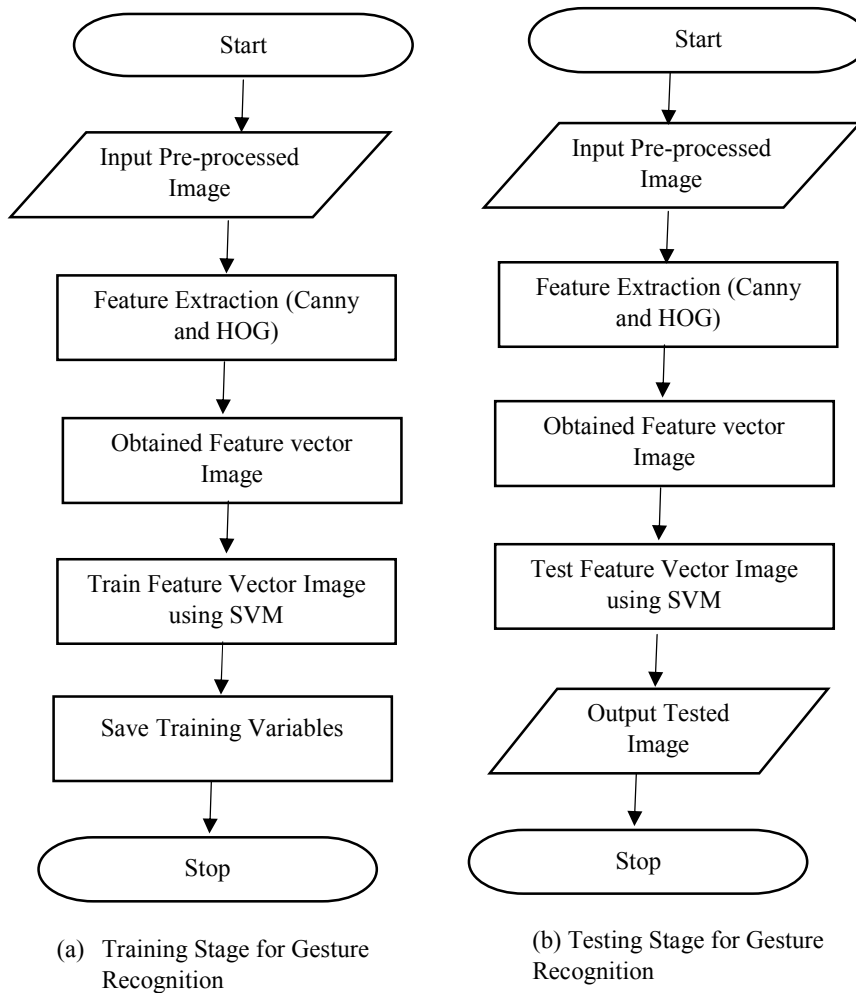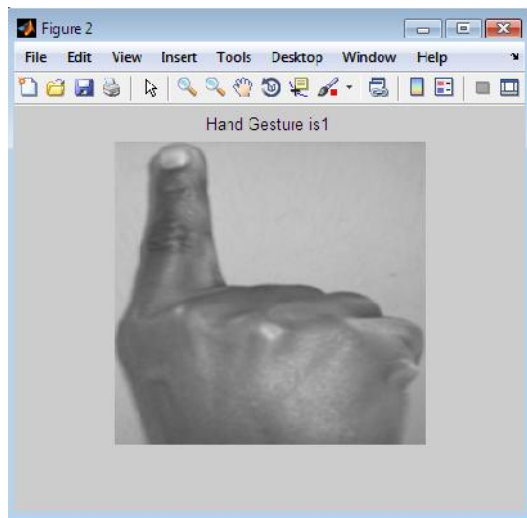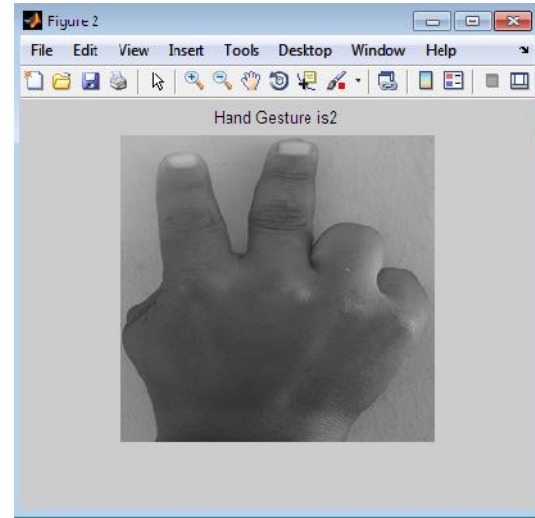

(a) Training Stage for Gesture Recognition

(b) Testing Stage for Gesture Recognition

Fig-4

Fig. 4. Implementation Stage

$$\text{Recall} \quad = \quad \frac{T_P}{T_P + F_N} \qquad (1)$$

$$Precision \quad = \quad \frac{T_P}{T_P + F_P} \qquad (2)$$

$$Accuracy \quad = \quad \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \qquad (3)$$

(a) Classified Hand gesture for numeral 1   (b) Classified Hand gesture for numeral 2
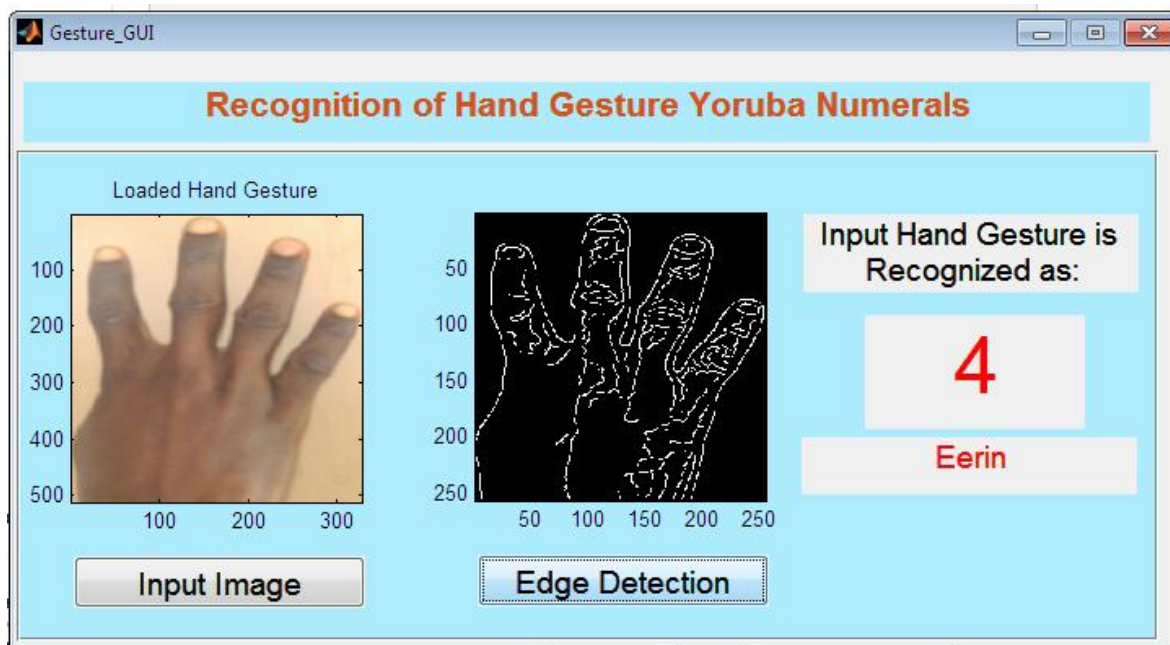
Fig. 5. Classified Hand Gestures

Fig. 6. System design Interface showing gesture recognition process

Table 1. Classification Rate for Yorùbá Hand Gesture

Sn	Yorùbá Numeral	English Meaning	Accuracy
1.	Ookan	One	100%
2.	Eeji	Two	90%
3.	Eeta	Three	100%
4.	Eerin	Four	85.67%
5.	Arun	Five	100%
6.	Eefa	Six	93.56%
7.	Eeje	Seven	100%
8.	Eejo	Eight	88%
9.	Eesan	Nine	90%
10.	Eewa	Ten	90.72%

CONCLUSION

In this study, hand gesture recognition for Yorùbá numeral was developed. The study employed canny edge detector and histogram of oriented gradient as feature extraction method and using support vector machine as its classifier. The developed system attempts to enhance the communication gap between hearing and hearing-impaired people in Yorùbá communities. The result produced from the evaluation of the performance of the system showed a robust output which indicates that the work can be incorporated into the deaf educational system. An average recognition rate of 94% on testing data was recorded over a dataset of 40 gestures. The method used in this study will be evaluated on a large dataset and compare with other methods such as linear discriminant analysis, artificial neural network and principal component analysis for efficient performance. Also, development of a real-time system will be considered for future work.

REFERENCES

1. Assaleh K., Al-Rousan M. Recognition of Arabic Sign Language Alphabet using Polynomial Classifier. Journal on Applied Signal Processing. 2005; 2005(13):2136-2145.
2. Hee-Deok Yang, Seong-Whan Lee. Sign Language Spotting with a Threshold Model Based Conditional Random Fields, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2009; 31(7):135-147.
3. Islam A., Siddiqua S., Afnan J. Rea Time Hand Gesture Recognition using Different Algorithms Based on American Sign Language. IEEE International Conference on Imaging Vision and Pattern Recognition. 2017;
4. Kalsh A., Garewal N. S. Sign Language Recognition for Deaf and Dumb. International Journal of Advanced in Computer Science and Software Engineering. 2013. 3(9):10-106.
5. Kausar S., Javed Y. M., Sohail S. Recognition of gestures in Pakistani Sign Language using Fuzzy classifier. International Conference on Signal Processing, computational Geometry and Artificial vision. 2008; pp. 101-105.
6. Keni M., Meher S., Marathe A., Sign Language Recognition System. . International Journal of Scientific and Enginering Research. 2013 4(12): 580-583.
7. Nair A. V., Bindu V. A Review on Indian Sign Language Recognition. International Journal of Computer Application. 2013; 73(22):33-38.
8. Orie O. O. From Conventional Gesture to Sign Language: The case of Yorùbá Sign Language. Selected Proceedings of the 43rd Annual conference on African Linguistaics: Linguistics Interface in African Languages. 2013.
9. Pansare J. R., Ingle M. Vision-based Approach for American Sign Language Recognition using Edge Orientation Histogram. International Conference on Image, Vision and Computing. 2016; pp. 86-90
10. Pramada S., Saylee D., Pranita N., Nerkar S., Intelligent Sign Language using Image Processing. Journal of Engineering. 3(2):45-51.
11. Ren Z., Yuan J., Mong J., Zhang Z., Robust Part-based Hand Gesture Recognition using Kinect Sensor. IEEE Transaction on Multimedia. 2013. 15, 1110-1120.
12. Riad A. M., Elminir H. K., Shohieb S. M., Hand Gesture Recognition System Based on a GeometricModel and Rule Based Classifier. British Journal of Applied Sciences and Technology. 2014. 4(9): 1432-1444.

13. Shivashankara S, and Srinath S. A review on vision-based American Sign Language Recognition, its techniques and outcomes. 7th International Conference on Communication systems and Network Technologies 2017; pp. 293-299.

14. Shujjat K., Gourab S., Gupta, D. B., Serge Demidenko, Chris M., Sign Language Analysis and Recognition: A Preliminary Investigation; 24th International Conference Image and Vision Computing New Zealand (IVCNZ 2009), IEEE, 2009.

15. Zafrulla Z., Brashear H., Starner T., Hamilton H., Presti P., American Sign Language Recognition with Kinect. In proceedings of the International Conference on Multimodal Interfaces Alicante Spain. 2011 pp. 279-286.