DOI: 10.36108/ujees/2202.40.0110

UNIOSUN Journal of Engineering and Environmental Sciences.Vol. 4 No. 1.March. 2022

Real-Time Large Vocabulary American Sign Language Recognition System for Mobile Devices

Jimoh, K.O., Ogunbiyi, D.T., Kasali, A.A. and Salami, I.O.

Abstract: The American Sign Language (ASL) is the only major language used in the educational system of hearing-impaired people in Nigeria. The automatic recognition system of the signs has not been currently used for the teaching of hearing-impaired students. This study developed a real-time large vocabulary sign language for ASL implemented on android devices. Samples of static and dynamic hand gestures were collected from the primary school of handicap, Osogbo. The specific objectives include the collection of hand gestures, examining the specific features for the recognition process, designing a model for the specific features examined, implementing the model, and evaluating the performance of the system. The real-time vocabulary sign language was recognized using Convolution Neural Network (CNN) implemented using Python programming language. The developed system was evaluated using precision, recall and accuracy as metrics. The model prediction carried out using the test image has an overall accuracy of 92.98%. The obtained result showed that the system will enhance the learning skills and provide adequate learning platform for both students and the teachers of hearing-impaired schools.

Keywords: Data Augmentation, Gesture Recognition, Overfitting, Sign Language, Tensor flow,

I. Introduction

Recently, there is great emphasis on human-computer-interface (HCI) research to gain the most efficient method of interfaces between manipulation skills of humans and computer hardware [1]. Now, technology and the interactions with computing devices have progressed significantly, as it has become a necessity to humans in their day to day life activities such as work, shopping, communicating, entertainment and so on [2]. Human-computer-interaction (HCI) as an interface between the human and computer (machine), has played a major role in the emergence of sign language. Gestures are a major form of human communication among hearing-impaired people. It is a means by which they relate with one another and also with hearing people. The use of gestures as a sign language has greatly enabled hearing-impaired people to communicate normally and effectively with hearing people [3].

In Nigeria, ASL is the only available language found to be used in the hearing-impaired educational system and has been manually used to teach the students. The development of an automated mobile application for the teaching of this language is crucial as the significance of the use of mobile devices increases. The emergence of the automated system is with the view to enhancing the learning process of the hearing-impaired student. Hence, gestures are found to be an appealing way to interact with computers and mobile devices, as they are already a natural part of how we communicate [4]. Sign language is a way through which

Jimoh, K.O., Ogunbiyi, D.T. and Salami, I.O. (Department of Information and Communication Technology, Osun State University, Osogbo, Osun State, Nigeria)
Kasali, A.A. (Department of Computer Engineering, The Federal Polytechnic, Ede, Osun State, Nigeria)
Corresponding author: kudirat.jimoh@uniosun.edu.ng
Phone Number:+2348038574796
Submitted: 11-12-2021
Accepted: 01-02-2022

hearing-impaired people communicate with one another and it is a major means through which they relate conveniently with hearing people. Sign language uses gestures to replace speech in carrying out their daily activities [5]. The design of sign language recognition systems has become imperative in everyday life among hearing-impaired people as these systems use intelligent techniques to efficiently respond to needs and can be easily accessible. The movement of the body parts, especially the hand or the head to express feelings and interpret the environment is mostly referred to as a gesture [6]. Traditionally, gesture emanates from body movement with face and hand as the main tools for gesture demonstration [7]. Literature on sign language recognition systems have shown that hearing-impaired people do not often have access to systems that aid their communication process [3] especially in Nigeria for a person with a hearing disability, mobile devices such as tablets and phones can serve as means of translating and communicating. This study extends existing research to develop a user-friendly system for communication flow between hearing and hearing-impaired people. Hence, the development of a real-time sign language recognition system for large vocabulary that can be used on android devices. The implementation of this system leveraged mobile devices due to its accessibility to most young and old people.

Several works of literature have been reported on the sign language recognition systems for different languages. [8] reviewed various kinds of literature on American Sign Language translators and examined different approaches by which Artificial Intelligence (AI) techniques can be adopted for Sign Language Recognition (SLR). In the study, various feature extraction techniques were also considered and their behaviour towards the SLR system. In [9] various literature on sign language recognition systems reviewed considered two significant approaches namely vision-based and sensor-based. The literature evaluated some techniques of vision-based and evaluated their accuracy result. The study revealed that variation exhibited in the accuracy result is due to the nature of the language considered and the problem addressed. In the work of [10] occlusion problem in sign language was addressed by using the geometric method to segment skin edges and Euclidean distance for the classification process of 30 isolated words collected. Due to the large dataset considered in the study and the nature of signs, detecting edges for the feature extraction process may be insufficient to generate an efficient output.

In another work of [11], recognition of Bahasa Isyarat Indonesia (BISINDO) letters was developed using Robert edge detector and chain code for the feature extraction process and Euclidean distance was adopted for the recognition process. The Robert edge detector adopted in the work usually missed out on some important features of an image because it only looks for big value in the first derivative of an image. In a related work of [12], 3D motion capture was employed to obtain sign images and frames and a graph matching algorithm to recognize signs. Also, problems relating to the generation of 3D variation were addressed with a total number of 350 words captured with a variation of 4 collected from 5 different people. [13] used Euclidean distance as a classifier to recognize the number base sign language and employed a Canny edge detector for feature extraction.

In [14], a gesture recognition system for Yorùbá Numeral was proposed. The work used only 40 static gestures as data and employed Histogram

of Oriented Gradient and Canny edge detector as feature extraction with Support Vector Machine as its classifier. [15] developed an intelligent gesture identification system for domestic navigation using Artificial Neural Network (ANN). The system was proposed to enable old age and disabled people to move around for their daily activities. For commands identification, both static and dynamic gestures were employed.

A template matching algorithm was proposed by [16]for sign language of selected English vocabulary using Oriented FAST and Rotated BRIEF and the Principal Component Analysis (PCA) was adopted for the feature extraction process. For real-time purposes, the model was implemented with OpenCV to recognize the gesture vocabularies using android studio. Some literature worked on the hardware implementation using a Microcontroller kit for the recognition of sign language. The work of [17] used ArduinoAtmega Microcontroller for the building of digital interface and APR voice recorder for the collection of voices of the selected data used. ANN was adopted for the recognition process. In a related work of [18], AVR Microcontroller with a GSM module was adopted for interaction between mobile devices. The work was implemented in a Matlab environment. The use of data glove with flex sensor was examined by [19] and also [20] recruited 121 Polish language using a continuous stream of RGB (Red, Green, and Blue) data and feature vector for recognizing some selected sentence. In another work of [21], a speech to gesture conversion system was developed. The work uses Raspberry Pi to build up the hardware component with a camera to capture the gesture. The collected data were pre-processed and Local Binary Patterns was used as a feature extraction method. Support Vector Machine was adopted for the recognition process.

The literatures reviewed above have focused on hand gestures of various languages and vocabularies implemented mainly on the computer and mobile devices with a limited number of the dataset. In this study, a real-time sign language recognition was proposed and implemented on android devices with a large dataset for more efficiency. This study was carried out with the aim of enhancing the existing sign language recognition system to provide accurate and reliable communication among hearing and hearing-impaired people, making the technology available and easily accessible by considering the use of mobile devices. Also, the consideration of ASL for English vocabularies in the development of the system arises due to its wide use in the hearing-impaired educational system.

II. Materials and Methods

The system development methodology adopted in this research is shown in Figure 1. This development methodology comprises four stages: data acquisition, image pre-processing, feature extraction and image recognition. Data (images and video frames) collected were presented to the model as input and the output was the corresponding hand gesture of the vocabulary displayed in English text with a percentage of accuracy level. The Images were augmented to add variations to the dataset and model trained using CNN. The model was saved and loaded with OpenCV to recognize the sign language vocabulary in real-time.

A. Data Acquisition

Samples of hand gestures for different vocabularies were collected from the school of handicap, Osogbo, Osun state using iPhone 12 Pro Max digital camera. The sign language vocabulary gesture collected were processed into images and video frames. The dataset consists of 345 images of 69 sign language vocabularies as shown in Figure 2. The hand-shaped sign language vocabularies were collected from 21 students of the handicapped school. Each student recorded five (5) images and videos for each hand gesture of the vocabulary. These were recorded from various, angles, positions and with different lighting conditions and backgrounds. The remaining images were reproduced by image augmentation techniques from 345 to 3,450 where each image generate 10 more images.

B. Image Pre-Processing

In the pre-processing stage, the images were preprocessed with augmentation techniques by considering the addition of pixels and colours; top and black hat, performed morphological transformation, blurring, saturation, and sharpening to enhance the variations to the dataset as shown in Figure 3. The augmentation techniques performed on a single image had generated 5 different variations of images. The Sign language dataset consisted of 345 images in total. The total image size in the dataset cannot be immediately fed into the CNN. The images were rescaled from 256×256 into 64×64×3 pixels at the time of fetching the images by the deep learning library called TensorFlow during the training of the model. Other transformation methods such as height and width shift,

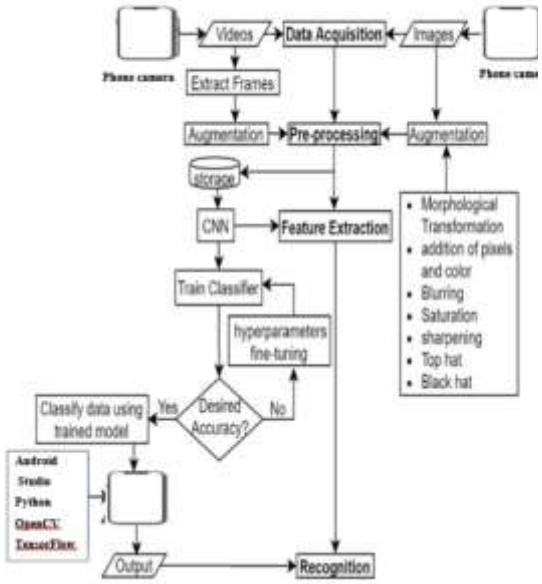

Figure 1: Overview of system methodology

Figure 2: Samples of collected sign gestures

rotation, horizontal flip and shearing, were also performed during the training.

C. Convolution Neural Network and Feature Extraction

Convolutional Neural Networks (CNNs) are among the most successful and widely used deep learning architectures in computer vision [2].

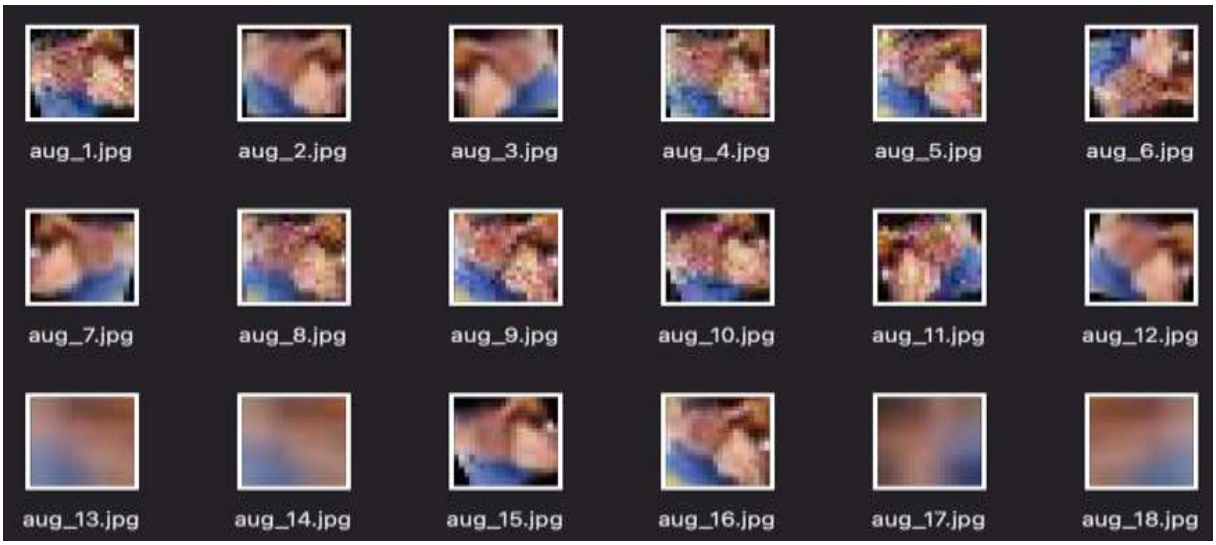

Figure 3: Sample of augmented images

It consists of three layers, namely, the convolutional layers for convolving a kernel of weights to extract features; the nonlinear layers for applying an activation function on feature maps to facilitate nonlinear function modelling and the pooling layers for replacing a small neighbourhood of a feature map with statistical information of the neighbourhood for spatial resolution reduction. Basically, units in the layers are locally connected, i.e., each unit receives weighted inputs from a small neighbourhood of units in the previous layer. Formally, a CNN consists of several convolutional layers, i.e.,$\tilde{X} = C_W(X)$, where C is the number of classes, actingon a p-dimensional input$X(\beta) = \left(X_1(\beta), \dots, X_p(\beta)\right)$by applying a bank of filters$W = (w_h, h'), h = 1, \dots, q;\ h' = 1, \dots, p$ and point-wise nonlinearity φ to give Eqn. (1). The input $X(\beta)$ correspond to the 64×64×3 pixels that were introduced to the CNN model as input.

$$\tilde{X}_h(\beta) = \varphi\left(\sum_{h'=1}^{p} \left(X_{h'} * \omega_{h,h'}(\beta)\right)\right), \qquad (1)$$

The function defined in Eqn. (1) produces a q-dimensional output$\tilde{X}(\beta) = \left(\tilde{X}_1(\beta), \dots, \tilde{X}_q(\beta)\right)$commonly referred to as feature maps, defined in Eqn. (2) as the standard convolution.

$$(X * \beta)(\beta) = \int_{\Psi}^{\beta} X(\beta - \beta')w(\beta')d\beta' \qquad (2)$$

According to the local deformation prior, the filters W have compact spatial support. Additionally, a down-sampling or pooling layer$\tilde{X} = P(X)$may be used, defined as

$$\tilde{X}_h(\beta) = P(\{X_l\beta' : \beta' \in N(\beta)\}), h = 1, \dots., q \qquad (3)$$

where$N(\beta) \in \psi$is in βneighborhood and P, a permutation-invariant function, representing the average, energy, or max-pooling. More so, a convolutional network is constructed by composing several convolutional and optionally pooling layers, obtaining a generic hierarchical representation. The output features obtained enjoy translation invariance/covariance which is a key advantage of CNNs success in numerous tasks is that the geometric priors on which CNNs are based, result in a sample complexity

that avoids the dimensionality challenges. In addition, due to its multiscale hierarchical property, the number of layers grows at a rate O(logD), where D is the number of pixels in an image, to result in a total learning complexity of O(logD) parameters.

D. Image Recognition

The trained model was converted into RGB and grayscale, and rescaled from 256×256 to $64 \times 64 \times 3$ was loaded into an android app using TensorFlow as backend, android mobile phone to read video frames, Android studio, and Anaconda as the editor and Java and Python as the programming languages respectively. A mobile phone with Android OS was used to capture real-time hand-shaped pictures and video frames from the signer and rescaled them into 64×64×3pixels. The model successfully detects and predicts sign vocabulary.

III. Results and Discussion

A total of 345 hand gestures of sign vocabulary collected were augmented to 3,450 images and trained using CNN architecture. The model was processed on a Core i5-5200 CPU of 2.2GHz with an 8GB RAM machine. The percentage of training and testing datasets adopted is 75% and 25%, respectively, using a batch size of 16. The training progress of the proposed CNN model was compared using accuracy and loss, as shown in Figures 4 and 5 respectively. Figure 4 shows the result of a model after data augmentation where the learning rate had greatly improved with no overfitting. The graph plotted after model fitting and data augmentation was carried out to reduce overfitting. The learning process gradually increased for the training and testing and stopped from 30 iterations with a total of 6

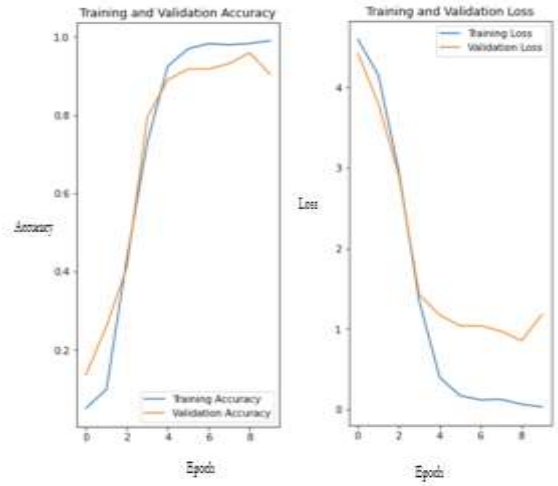

Figure 4: Train and test accuracy before data augmentation: (left) Accuracy vs. epoch, (right) Loss vs. epoch

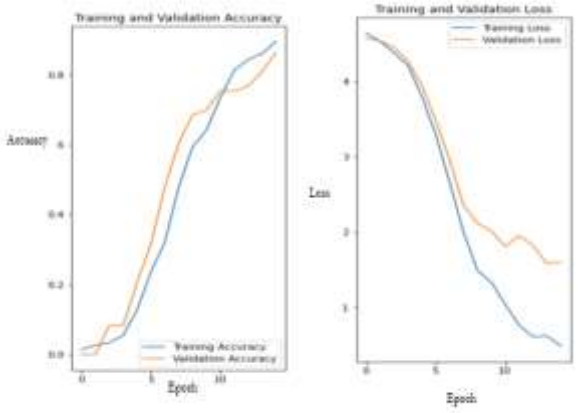

Figure 5: Train and test accuracy after data augmentation: (left) Accuracy vs. epoch, (right) Loss vs. epoch

epochs. The performance of the model was evaluated using precision, recall, and accuracy as metrics as shown in Equations 4, 5 and 6 respectively. Sample of the result is shown in Figure 6 where the hand gesture 'Agree' have an average precision of 81.45%, recall of 87.34% and accuracy of 78.9% respectively.

$$Precision = \frac{T_P}{T_P + F_P} \quad (4)$$

$$Recall = \frac{T_P}{T_P + F_n} \quad (5)$$

$$Accuracy = \frac{T_p + T_n}{T_P + T_n + F_P + F_n} \quad (6)$$

where T_P is True Positive, F_P is False Positive, T_n is True Negative, F_n is False Negative

Figure 6: Real-time recognition of sign "Agree"

IV. Conclusion

An android application for the automatic gesture recognition for sign language was developed using tensor flow library as backend, the image collected was processed using Keras library, and data was augmented to improve the accuracy of the performance of the model, the model prediction carried out using the test image had an overall accuracy of 92.98%.

The adoption of CNN with data augmentation an artificial intelligence tool has greatly enhanced the performance of the system. The augmentation techniques have assisted to generate 3,345 images from 345 images collected from the handicap school in Osogbo. The system developed, when fully deployed into the hearing-impaired educational system will enhance their learning skills and provide an adequate learning platform for both students and the teachers.

References

[1] Andres, Jaramillo-Yanez, Marco, E. andBenalcazer, M.E.M. "Real-Time Hand Gesture Recognition Using Surface Electromyography and Machine Learning a Systematic Literature Review", *Sensors*, vol. 20, no. 9, 2020,pp. 24-67.

[2] Danielle, B., and Oscar Koller, M.B.L.B. "Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective", *21st International SIGASSETS Conference on Computer and Accessibility*, Pittsburgh, PA, USA, 2019, pp. 16-31.

[3] Padmapriya, S. "Hand Gesture Recognition system using android", *International Journal of Research in Science and Technology*, vol. 2, no. 1, 2015, pp 72-76.

[4] Susan, Goldin-Meadow, Diane, B. "Gesture, Sign and Language: The Coming of Age of Sign Language and Gesture Studies", *Behavioral and Brain Sciences*, vol. 51, no. 3, 2015, pp. 1 – 82.

[5] Jerald, S., Hilwa, K. andJinsha, J. "Hand Gesture Recognition",*International Journal of Innovative Technology and Research,* vol. 3, no. 2, 2015, pp. 1946-1949.

[6] Anake, P.M. and Asor, L.J. "Gestures in Guidance and Counselling and their Pedagogical/Androgogical Implications", *Global Journal of Educational Research*, vol. 11, no. 2, 2012, pp 73-78.

[7] Shailesh, B., Shubham dixit, R.C, and Vinash, B. "Sign Language Recognition using Neural Network", *International Journal of Engineering and Technology*,vol. 7, no. 4, 2020, pp 585 – 586.

[8] Ardiansyah, A., Hitoyoshi,B., Halim, M., Hanafiah, N. and Wibisurya, A. "Systematic

Literature Review: American Sign Language Translator", *Procedia Computer Science*,vol 179, 2021 pp 541–549.

[9] Suharjito, N.T., and Herman, G. "SIBI Sign Language Recognition Using Convolutional Neural Network Combined with Transfer Learning and non-trainable Parameters", *Procedia Computer Science,*vol. 179, 2021, pp 72–80.

[10] Ibrahim, N.B., Mazen, M.S. and Hala, H. Zayed. "An Automatic Arabic Sign Language Recognition System (ArSLRS)", *Journal of King Saud University –Computer and Information Sciences*, vol 30, 2018, pp 470-477.

[11] Purnawansyah, D.I., Madenda, S., Wibowo, EriPrasetyo. "Indonesian Sign Language Recognition Based on Shape of Hand Gesture", *Procedia Computer Science,*vol. 161, 2019, pp 74–81.

[12] Kumar, D. Anil Sastry, A.S.C.S. Kishore, P.V.V. and Kumar, E. Kiran. 3D Sign language recognition using Spatio-temporal graph kernels, Journal of King Saud University – Computer and Information Sciences, https://doi.org/10.1016/j.jksuci.2018.11.008accessed on July 28, 2021

[13] Rajama, P.S. and Balakrishnan, G. "Recognition of Tamil Sign Language Alphabet using Image Processing to aid Deaf-Dumb People", *Procedia Engineering,*vol. 30, 2012, pp 861 – 868.

[14] Jimoh, K.O., Adepoju, T.M. Sobowale, A.A. andAyilara, O.A. "Offline Gesture Recognition System for Yoruba Numeral Counting", *Asian Journal of Research in Computer Science,* vol. 1, no. 4, 2018, pp. 1-11.

[15] Ravindu, H.M., Bandara, T., Priyanayana, K.S., Buddhika, A.G., Jayasekara, P., Chandima, D.P.and Gopura, R.A. An Intelligent Gesture Classification Model for Domestic Wheelchair Navigation with Gesture Variance Compensation. Applied Bionics and Biomechanics, vol. 2020, no. 9160528, 2020, pp.1-11.

[16] Jimoh, K.O., Ajayi, A.O. andOgundoyin, I.K. "Template Matching Based Sign Language Recognition System Devices", *FUOYE Journal of Engineering and Technology*, vol. 5, no. 1, 2020, pp. 42-48.

[17] Aruljothy, S., Arunkumar, S., Ajitraj, G., Yayad, D., Jeevanantham, J. and Subba, M. "Hand Gesture Recognition Using Image Processing for Visually Impaired and Dumb Person", *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 7, no.4, 2018, pp 142 – 148.

[18] Aswathy, M., Narayanan, H., Rajan S., Uthara, P.M. and Jacob, J. "Hand Gesture Recognition and Speech Conversion for Deaf and Dumb using Feature Extraction", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, vol. 6, no. 3, 2017, pp 2093-2098.

[19] Jagadish, R., Gayathri, R., Mohanapriya, R., Kalaivani, R. and Keerthana, S. "Hand Gesture Recognition System for Deaf and Dumb Persons",*Indo-Iranian Journal of Scientific Research*. vol. 2,no. 1, 2018, pp 139-146.

[20] Tomasz, K. and Marian, W. "Recognition of Signed Expressions in an Experimental System Supporting Deaf Clients in the City Office",*Sensors,*vol. 20, no. 2190, 2020, pp 1-19. doi:10.3390/s20082190.

[21] Latha, L. and Kaviya, M. "A Real-Time System for Two Ways Communication of Hearing and Speech Impaired People",*International Journal of Recent Technology and Engineering*, vol. 7,no. 4S2, 2018, pp. 382-385.

[22] Minaee, S., Boykov, Y.Y.,Porikli, F., Plaza, A.J.,Kehtarnavaz, N. and Terzopoulos, D. "Image Segmentation Using Deep Learning: A Survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2021.3059968.