Template Matching Based Sign Language Recognition System for Android Devices

*1Kudirat O. Jimoh, 2Anuoluwapo O. Ajayi and 1Ibrahim K. Ogundoyin

1Department of Information and Communication Technology, Osun State University, Osogbo, Nigeria
2Department of Computer Science and Engineering, Obafemi Awolowo University, Ile-Ife, Nigeria
kudirat.jimoh@uniosun.edu.ng|anuajayi@oauife.edu.ng|ibraheem.ogundoyin@uniosun.edu.ng

Submitted: 29-DEC-2019; Reviewed: 04-MAR-2020; Accepted: 27-MAR-2020

Abstract- An android based sign language recognition system for selected English vocabularies was developed with the explicit objective to examine the specific characteristics that are responsible for gestures recognition. Also, a recognition model for the process was designed, implemented, and evaluated on 230 samples of hand gestures. The collected samples were pre-processed and rescaled from 3024 x4032 pixels to 245 x350 pixels. The samples were examined for the specific characteristics using Oriented FAST and Rotated BRIEF, and the Principal Component Analysis used for feature extraction. The model was implemented in Android Studio using the template matching algorithm as its classifier. The performance of the system was evaluated using precision, recall, and accuracy as metrics. It was observed that the system obtained an average classification rate of 87%, an average precision value of 88% and 91% for the average recall rate on the test data of hand gestures. The study, therefore, has successfully classified hand gestures for selected English vocabularies. The developed system will enhance the communication skills between hearing and hearing-impaired people, and also aid their teaching and learning processes. Future work includes exploring state-of-the-art machining learning techniques such Generative Adversarial Networks (GANs) for large dataset to improve the accuracy of results.

Keywords- Feature extraction; Gestures Recognition; Sign Language; Vocabulary, Android device.

1 INTRODUCTION

Hand gesturing is a natural and intuitive communication modality for human-computer interaction. It is a tool for a human to human interaction by which they convey information in the day to day activities (Nyiirarugira, 2016). Attention has been paid to the hand gesture recognition system because of its ability to communicate efficiently with computer (Khan and Ibraheem, 2012). The development of efficient and effective human-computer interface is a necessity for a computer to visually identify and recognize hand gestures in a real-time environment in order to allow a comprehensive communication among the hearing-impaired people (Shakunthaladevi, 2014).

Sign language is an important tool in the gesture recognition process as it aids the communication gap between hearing and hearing-impaired people. Sign Language (SL) is a medium which allows communication among the hearing-impaired people that emerges naturally within the impaired community (Kour and Mathew, 2017). Sign language system comprises of manual and non-manual signals in which the former involves fingers, hands, arms; and the later use face, head, eyes, and body (Kour and Mathew, 2017). It is therefore important to create a learning environment where hearing-impaired people can learn signs (Stefano and Beskow, 2016).

Nowadays, most hand-held devices such as phones and tablets are very common among people of all age categories (Boulos et al., 2011). For a disabled person who is unable to speak, these devices can serve as a means of understanding, translating; and a communicating system for these set of people (Kakde and Rawate, 2016). It has been noticed that hearing-impaired people do not often have access to systems that aid their communication process.

Sign language Recognition has been defined as the development of an algorithm and methods to efficiently recognize a series of signs and interpret their meaning (Jalal, 2015). Sign language recognition systems are important tools that aid the teaching and learning process of impaired people. In Nigeria today, most hearing-impaired schools are yet to teach with an automated gesture recognition system. However, this study will aid the method of teaching and also provide a conducive environment for both hearing-impaired and hearing people. An android-based and embedded system is becoming an area of attention as they represent the main prime trends in all applications. These devices can assist people who have difficulty in communicating properly and even in emergency conditions (Kakde and Rawate, 2016). This study developed a gesture recognition system that is Android-based for some selected English vocabularies. Various gesture recognition systems involving hand have been reported to date. The glove-based control interface is the genesis of the hand gesture recognition for the computer control system (Premaratne, 2014). This gradually evolved with the development of many accurate accelerometers, infrared cameras, and even fiber optic bend-sensors (optical goniometry). With these developments in glove-based systems, computer vision-based recognition system requires no sensors attached to the glove (Premaratne, 2014). The data glove method is a wired interface with certain tangible sensory units that were attached to the fingers or joints of the user wearing the glove (Sharma and Sharma, 2015).

Two main approaches are employed in the development of an automatic hand gesture recognition system, and these approaches are sensor-based approach and vision-based approach (Kour and Mathew, 2107). The sensor-based systems involve data gloves with accurate positions of hand gestures directly measured (Husain et al., 2013). The device employed in this method is very complicated with many cables connected to a computer.

*Corresponding Author
As a result of this, the vision-based approaches for recognizing gestures are preferred (Mitra and Acharya, 2007). Also, a major disadvantage of the sensor-based approach of gesture recognition is that complicated gestures cannot be represented because of the techniques involved in the recognition process (Kour and Mathew, 2017). The vision-based approaches obtain images from the camera as data input and then process them for gesture recognition. The focus of this method is mainly on the captured image of gesture; and the extraction of the main features for the recognition process (Kour and Mathew, 2017). The geometrical shape of trajectory and its gesture drawing speed are used to extract different features, which are then used by various pattern recognition algorithms to classify the different gestures (Misra et al., 2017).

Various prevailing methods of deaf-mute communication interpreter system were examined by Sunitha (2016). The two broad classifications of communication methodologies used by deaf-mute people are wearable communication device and online learning system. For wearable communication method, Glove-based system, keypad method and Handicom Touch-screen are three techniques used. All these techniques employ various sensors like accelerometer, a suitable microcontroller, a text to speech conversion module, a keypad, and a touchscreen. The need for an external device to interpret messages between deaf-mute and non deaf-mute people can be overcome by the second method i.e. the online learning system. The work used several hardware devices that are too cumbersome for the wearer.

The work presented by Maandep and Amardeep (2015) used Principal Component Analysis (PCA) for the recognition of hand gestures. In the work, a scheme was proposed with a database-driven hand gesture recognition based on skin color model and thresholding approach using a template matching algorithm. The segmentation of the hand region was achieved by applying skin color model in the color space, thresholding was applied for pixel separation, and a template-based system was developed using Principal Component Analysis (PCA) for the recognition process. This work is limited to hearing-impaired people only and cannot be used by the hearing community for their communication process. In a related work, an offline gesture recognition system for Yoruba numeral counting was proposed by Jimoh et al. (2018). The work considered hand gesture for Yoruba numeral from one to ten as data input and used edge detection and histogram of the oriented gradients as feature extraction techniques with support vector machine as a tool for the recognition process. The peculiarity of the work is the type of gesture considered, and its limitation is the static nature of the system.

Pansare et al. (2013) developed a real-time hand gesture recognition system for Devenagari number system. The work employed discrete cosine transform and edge oriented histogram as a comparison for the calculation of feature vector, and least Euclidian distance was used for the recognition of match hand gesture. The accuracy obtained for the recognition process was not presented, and the dataset used for the training and testing of the system are extremely low. In a related work by Jost (2015), a real-time gesture recognition based on skeleton analysis was presented. Feature analysis is performed on skeleton joints using principal component analysis, and distance measure were employed for the recognition process. The work employed a single person for the data collection, which limits the system to a specific gesture. The confusion rate is a bit high as the work considered only gesture with a similar part of the body. The Fuzzy decision tree employed in Gaolin and Wen (2004) generates large computational rules for the investigation of search space in vocabulary sign language. In another work of Lai (2018), a real-time hand gesture sensing and recognition system was developed and its application to a TV channel and volume control process was examined. Rajan and Leo (2019) reviewed various methods of sign language recognition system and observed that camera based approach is more conservative in terms of computational complexity and cost.

The above reviewed literature had considered hand gesture of various language and vocabulary implemented mainly on computer. This study, therefore, developed an android based sign language recognition system for selected gesture implemented on mobile devices. It was developed to enhance the communication process and also to aid the teaching and learning process of the hearing impaired people. The system was designed to recognize words in English Vocabulary for the gesture collected from the School for the Handicapped at Osogbo in Osun State.

The rest of this paper is arranged as follows. Section 2 describes the materials and methods used to develop the work. Result and discussion are discussed in Section 3. Section 4 concludes the paper.

2 MATERIALS AND METHODS
The materials and methods used for the implementation of this study are discussed in the following subsections.

2.1 VOCABULARY MODEL FOR DATASET USED
The gesture recognition model was divided into three stages, which involved the image acquisition and image processing stage; feature extraction stage, and recognition stage. These stages are broadly classified as the feature extraction stage and the classification/identification stage. The feature extraction stage was achieved using Oriented FAST and Rotated BRIEF (ORB) algorithm with Principal Component Analysis (PCA) while the classification and recognition stage was achieved using the template matching algorithm. The ORB algorithm with PCA was selected because the combination of the two algorithms have been proven to be cost effective and require less computation (Vinay and Natarajan, 2015).

Figure 1 shows a flowchart of the gesture recognition model. The captured gesture images are input into the model as input. The models first convert the image to grayscale, and the images were pre-processed using various image processing techniques, then feature vectors were extracted using the stated feature extraction techniques. The obtained feature vectors were trained
using the template matching algorithm in OpenCV, and the corresponding gesture type was recognized. Gesture images of forty-six (46) selected words in emotional and mental vocabularies were collected. Five different signers were employed for each word making a total of two hundred and thirty (230) samples collected. The two classes of data used are static and dynamic. Some of the static data include wish, usually, stupid, stubborn, sorry, satisfy, sad, respect, remind, reason, love, ignorant, I love you, happy, fool, feel, cry, compare and so on. The dynamic data include goal, laugh, like, memorise, opinion, refuse, surprise, understand, want, wisdom and curious.

2.2 IMAGE PROCESSING

For identification and classification purpose, samples of digital gesture images of selected words in emotional and mental vocabularies were collected from students of the school for the handicapped, Osogbo, Osun State using a Samsung 16 Megapixel digital camera. These data collected were used as input for the recognition process. In total, 230 samples of data were collected and processed in a digital domain. Figure 2 shows some of the samples of data collected for the development of the system. The collected data were rescaled using OpenCV image processing functions to specify the area needed for the gesture recognition, and other irrelevant parts of the data collected were removed. Also, the images were resized from 3024 × 4032 pixels to 245 × 350 pixels using OpenCV image processing functions and also to enhance the quality of the images. The resulting images were saved using the Tag Image File Format (TIFF). All the sample images used for the recognition process were converted from RGB (Red, Green, and Blue) color component to graysc.

2.3 ORB–PCA BASED FEATURE EXTRACTION METHOD

Oriented Fast and Rotated Brief is a fusion of FAST keypoints detector and BRIEF descriptor are designed to enhance the performance of image classification systems while detecting the image features. PCA is a statistical algorithm that extracts the region of interest in an image. The primary aim of this algorithm is to reduce the dimension of a given dataset. Harris Corner measure and pyramid were applied to find the N-points and multi-scale feature, respectively. To calculate the orientation, the pixel intensity weighted centroid of the patch with the central corner was computed using the Harris corner measure (Harris and Stephens, 1988). For an image t with h with horizontal and vertical derivatives, the matrix in Equation 1 can be accumulated.

\[ M = \frac{1}{n^2} \begin{bmatrix} i_x^2 & i_x i_y \\ i_x i_y & i_y^2 \end{bmatrix} \] (1)

Where \( i_x \) is the sum of the derivatives in x-direction, \( i_y \) is the sum of derivative in y-direction for every pixel \( i \) is the summation average. To calculate the orientation of the patch using Equation 1, the eigenvector of Equation 1 is converted to a rotation angle as shown in Equations 2-5.

\[ d = \text{determinant}(M) = M(1,1) \times M(2,2) - M(1,2) \times M(2,1) \] (3)

\[ \text{.angle} = \text{atan2}(M(2,1), e i g - M(2,2)) \] (5)

The value obtained in Equation 4 was found to be the larger value of the two eigenvalues corresponding to the eigenvector shown in Equation 6.

\[ v = \begin{bmatrix} e i g - d \\ M(2,1) \end{bmatrix} \] (6)

The extracted feature vectors were then used for the recognition process. Figure 3(a) shows the original image of I love, its corresponding grayscale image is given in Figure 3(b) and Figure 3(c) shows the extracted image. The selection of good feature is very important to gesture recognition system. This is because the classification of a
typical hand gesture interprets to a predefined sign language. The non-geometric features extracted for the recognition of hand gesture are motion, shape and the orientation of the gesture. The geometric feature such as finger direction, hand contour and fingertips are explicitly not sufficient and reliable for the recognition process because of lighting conditions (Badi et al., 2015).

For 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 gesture recognition system, respectively. For the training stage, the extracted feature vectors were trained using the template matching algorithm, and the training parameters were saved. In the testing stage, the input image from the testing sets was pre-processed and tested using matching technique and then output the result. Android studio and Software Development Kit (SDK) tools were used to create the application for the recognition system on android platform and then the OpenCV library which contains the recognition algorithm was used to complete the application development process. The XML file was created for the User Interface (UI). The XML file contains the visual description of the manifestation of the android program (apk). The Android Manifest is the file which manages permissions and controls. The camera permission was also added through the AndroidManifest.XML file so that features in this file can use the camera application on the android device.

| Table 1. Template Matching Pseudo-code |
|---------------------------------------|
| **Template Matching Pseudo-code**     |
| 1. Read Input Image \(I_{(x,y)}\)     |
| 2. Convert input image \(I_{(x,y)}\) to grayscale |
| 3. Read the template (feature vector) template = cv2.imread('template', 0) |
| 4. Save width and height of template as \([w,h]\) |
| 5. Compute Template match operations: |
| 6. Specify a threshold, \(\text{thresh} = 0.7\) |
| 7. Save the coordinate of matched area \((R_{(x,y)})\) in array |
| 8. Draw a circle around the matched region for \(tp\) in \(matarea[::1]\): |
| \(cv2.circle(gray\_img, tp, (tp[0] + w, tp[1] + h)(0,255,255),2)\) |
| 9. Show the detected region (final image). |

**2.4 Template Matching Algorithm**

This is a method of identifying or finding areas of an input image that are similar to a template image (patch). In this process, two sets of images are involved; the input or original image \(I_{x,y}\) and the template \(T_{x,y}\). The aim of this process is to identify the highest matching point or area. In order to find the matching area, the template image and the input image were compared using the sliding method that is, the patch was moved one pixel at a time (left to right, up to down). A metric was computed at each location of \(T_{x,y}\) over \(I_{x,y}\) and stored as result matrix \(R_{x,y}\) (match metric). This algorithm was implemented using the OpenCV function matchTemplate with the matching method shown in Equation 7. The pseudo-code is given in Table 1.

\[
R_{(x,y)} = \sum_{x,y} (T_{(x,y)} - I_{x+y})^2
\]  

**2.5 Simulation Model**

At this stage, we created a library to store all the feature vectors obtained from the feature extraction process. A collection of all template images to be matched were stored in the device work area i.e., the hand gesture library created on the android studio application. The captured images were used as data input for the recognition system. After an image is captured, the system receives and compares the data input with the templates. The final step requires matching the images to generate an appropriate output of either match or no match. When a match occurs, the appropriate output is generated on the screen indicating the gesture type.
result for the gesture “Happy”. The effectiveness of the algorithm used in this study is measured by inputting sample hand gesture images into the model, and its performance was evaluated. The performance metrics used are namely, recall rate, precision rate, and accuracy, which are defined in Equations 8, 9, and 10, respectively. We used these metrics to record and note the classification efficiency of the model. Table 2 shows the confusion matrix table for the overall recognition process. The accuracy of the model was recorded for each gesture classified, and the overall recognition rate was obtained using the confusion matrix table. Table 3 shows the classification rate of the recognition system where 100% accuracy was obtained for gestures ‘I Love You’, ‘Happy’ and ‘Stupid’. For the following gestures, ‘Idea’, ‘Fool’ ‘Frustrated’ and ‘Stubborn’, the system recorded the accuracy of 98%, 96%, 90% 90%, respectively. Similarly, 88% accuracy was obtained for ‘Compare’ and ‘Jealous’ while 86% accuracy was obtained for the ‘Anger’ gesture. Consequently, an overall average recognition rate of 87% was obtained on the test data.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{9}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{10}
\]

Where True Positive (TP) means predicting “True” for a label and the output is “True” (Correct Classification), True Negative (TN) means predicting “False” for a label and the output is “False” (Correct Classification), False Positive (FP) means predicting “True” for a label but the output is “False” (Incorrect Classification), False Negative (FN) means Predicting “False” for a label but the output is “True” (Incorrect Classification).

[Diagram: Implementation Stage]

(a) Training Stage

(b) Testing Stage

Fig. 4: Implementation Stage
A model for the automatic gesture recognition for sign language was developed on the Android interface using features extracted from sample images using ORB-PCA as feature extraction method and template matching algorithm as its classifier. The system developed in this study tries to mimic the manual process of communication among hearing-impaired communities efficiently and effectively. The system developed will also improve the learning process of hearing-impaired people in their educational system and enhance the communication gap between hearing and hearing impaired. The result obtained from the performance evaluation of the system showed a robust output, which indicated the seamless implementation of the developed model into the deaf and deaf-mute educational system.

An overall recognition rate of 87% on tested data was recorded over a dataset of 230 hand gestures. It was observed that the model was capable of recognizing hand gesture using distinguishing characteristics such as motion, hand shape, and position of the gesture. It was also observed that the gesture recognition process for the sign language can be implemented on android devices. The result obtained from this study showed that hand gesture recognition for sign language can be implemented on an android device and thereby improving the communication skills of hearing- and hearing-impaired people.

The study is limited in that fewer numbers of data are used. However, this study has laid a framework for continued research to develop advanced and robust version using Android application as a platform for its full-scale implementation. Also, we are considering further research in the use of state-of-the-art machine learning techniques such as Generative Adversarial Networks (GANs) to improve the accuracy of results.

### 2 CONCLUSION

In this study, a model for the automatic gesture recognition for sign language was developed on the Android interface using features extracted from sample images using ORB-PCA as feature extraction method and template matching algorithm as its classifier. The system developed in this study tries to mimic the manual process of communication among hearing-impaired communities efficiently and effectively. The system developed will also improve the learning process of hearing-impaired people in their educational system and enhance the communication gap between hearing and hearing impaired. The result obtained from the performance evaluation of the system showed a robust output, which indicated the seamless implementation of the developed model into the deaf and deaf-mute educational system.

An overall recognition rate of 87% on tested data was recorded over a dataset of 230 hand gestures. It was observed that the model was capable of recognizing hand gesture using distinguishing characteristics such as motion, hand shape, and position of the gesture. It was also observed that the gesture recognition process for the sign language can be implemented on android devices. The result obtained from this study showed that hand gesture recognition for sign language can be implemented on an android device and thereby improving the communication skills of hearing- and hearing-impaired people.

The study is limited in that fewer numbers of data are used. However, this study has laid a framework for continued research to develop advanced and robust version using Android application as a platform for its full-scale implementation. Also, we are considering further research in the use of state-of-the-art machine learning techniques such as Generative Adversarial Networks (GANs) to improve the accuracy of results.

### REFERENCES

- Badi H., AbdulKareem S. and Husein S. (2015) Recent Trends in Hand Gesture Recognition, Science Gate Publishing, vol.3 pp. 19-41.
- Boulos, M. N. K., Wheele S., Tavares C. and Jones R. (2011), How Smartphones are changing the face of Mobile and participatory healthcare: an overview with example from eCAALYX. Biomedical Engineering Online 10(24):1-14.
- Gaoli, F. and Wen G. (2004) Large vocabulary sign language recognition based on fuzzy decision trees. IEEE Transactions on Systems, Man and Cybernetics. Vol. 34 No. 3. pp.305-316.
- Harris C. and Stephens M. (1988) A Combined Corner and Edge Detector. Alvey Vision Conference. pp.15.
- Husain, F., Shivani G., Nijhawan T., Agarwal V., Khutun S. and Parveen S. (2013) Gesture Recognition System using MATLAB: A Literature Review. International Journal of Scientific Research and Management Studies. 2(11): 425-432.
- Jalal, A. S. (2015) Automated Recognition of Dynamic Isolated Sign in Video for Indian Sign Language. Minor Research Project, Computer Engineering Institute of Engineering and Technology, Maathura, India.
- Jimoh, K. O., Adepoju A. A and Ayilara A. O. (2018) Offline Gesture Recognition System for Numerical Counting. Asian Journal of Research in Computer science 1(4):1-11.
- Jost C., Stankovic J., Loor P., Nedeele A. and Bevaqua E. (2015), Real Time Gesture Recognition Based on Motion Quality Analysis7th International Conference on Intelligence Technologies for Interactive Entertainment, Tokyo, Italy.
- Kakde, M. U. and Rawate A. M. (2016) Hand Gesture Recognition System for Deaf and Dumb People Using PCA. International Journal of Engineering Science and Computing. 2016; 6(7): 1891-1895.
- Khan R. Z. and Ibraheem N. A. (2012) Hand Gesture Recognition: A Literature Review. International Journal of Artificial Intelligence and Applications. Vol. 3 No. 4. pp. 161-174.
- Kour K. P. and Mathew L. (2017) Literature Survey on Hand Gesture Techniques for Sign Language Recognition. International Journal of Technical Research and Science. 2 (7), pp. 431-434.
- Lai, H., Me H. and Hsu Y. (2018) Real-time Hand Gesture Recognition System and Application. Sensors and Materials 30(4):869-884.
- Mandeep, K. A. and Amardeep S. (2015) Hand Gesture Recognition using PCA. International Journal of Computer Science and Engineering Technology 5(7): 267-285.
Mitra, K. A. and Acharya A. S. (2007) Evolution of Hand Gesture Recognition. *International Journal of Research in Computer and Communication Technology* 4(21): 108-120.

Misra, S., Singh J. and Laskar R. (2017) Vision based hand gesture recognition of alphabets, number, arithmetic operators and ASCII characters in order to develop a virtual text-entry interface system. Article in Neural Computing and Applications 123-145.

Nyirarugira, C., Choi H. and Kim T. (2016) Hand Gesture Recognition using Particle Swarm Movement. Mathematical Problems in Engineering. Vol. 2016 No. 1. pp. 1-8.

Pansare J., Kadam R., Dhawade S. and Chauhan O. (2013) Real Time Static Hand Gesture Recognition System in Simple Background for Devenari Number System. *International Journal of Advanced Research in Computer Engineering and Technology*. 2(3):938-941.

Premaratne, P. (2014) Human Computer Interaction Using Hand Gestures. Cognitive Science and Technology, Science and Business Media, Singapore. 5-14.

Rajan, R. G. and Leo M. J. (2019) Comprehensive Analysis on Sign Language Recognition: System. *International Journal of Recent Technology and Engineering*. 7(6):749-755

Sanchez, J., Monzon N. and Salgado A. (2018) An Analysis and Implementation of the Harris Corner Detector. *Image Processing on Line*. Vol. 8. pp. 305-328.

Sharma, P. K. and Sharma S. (2013) Evolution of Hand Recognition: A Review. *International Journal of Engineering and Computer Science*. 4(1): 9962-9969. (2014) Real time Hand Gesture Recognition using AVR. International Journal of Engineering Research and Technology Vol. 3 No.5 pp. 299-305.

Stefanov, K. and Beskow J. (2016) A Real-time Gesture Recognition System for Isolated Swedish Sign Language. Proceedings of the 4th European and 7th Nordic Symposium on Multimodal Communication 4(3): 18-28.

Sunitha, K. (2016) Deaf Mute Communication Interpreter. *International Journal of Technical Research and Science*. 2(3): 436-445.

Vinay A. and Natarajan S. (2015) ORB-PCA Based Feature Extraction Technique for Face Recognition. *Procedia Computer Science*. Vol. 58 No. 2015. pp. 614-621.