IJEW 国际工程期刊 International Journal of Engineering Works انجینئرنگ کاموں کا بین الاقوامی جرنل

International Journal of Engineering Works
ISSN-p: 2521-2419
ISSN-e: 2409-2770
Vol. 7, Issue 09, PP. 305-311, September 2020
https://www.ijew.io/
https://doi.org/10.34259/ijew.20.709305311

Hindko Sign Language (HSL) Recognition using Convolutional Neural Network

Ali Raza[1], Dr.Syed Irfan Ullah[2]
[1,2] Department of Computing, Abasyn University Peshawar Pakistan
ali.raza@abasyn.edu.pk[1], syed.irfanullah@abasyn.edu.pk[2]
Received: 30 August, Revised: 14 September, Accepted: 20 September

***Abstract*—** Language has always played a significant role in Human to Human communication. In case of not knowing someone else's language, one can use hand gestures for communicating crudely but still be able to convey the message. Other than not knowing someone else's language there are millions of people in the world who have hearing or speaking disability. According to the World Health Organization (WHO), it has been estimated to be a population of 436 million (5% of the total world's population) people in the world who have hearing disabilities. Deaf people cannot use oral languages for communicating with other people. The source of their communication is Sign Language (SL) that conveys the message to the other person. In Computer Vision, there are different algorithms, which are used to interpret gestures and recognize them. The Deaf community of Pakistan uses its SL, like any other country in the world i.e., Pakistan Sign Language (PSL). There are around 60 local languages that are spoken in Pakistan including Hindko. Hindko and many other local languages spoken by minorities in Pakistan are on the brink of being endangered as the amount of research done on these languages is almost negligible. In this paper Convolutional Neural Network (CNN) is used for the recognition of Hindko Sign Language (HSL). Furthermore, we examine and analyze the recognition based on prediction to evaluate the efficiency of the utilized CNN. The methodology developed in this research work achieved an accuracy rate of 99.98%.

***Keywords*—** World Health Organization (WHO), Sign Language (SL), Gestures, Pakistan Sign Language (PSL), Convolutional Neural Network, Hindko Sign Language (HSL).

I. INTRODUCTION

The use of gestures and its applications are vast in number and are utilized in domains like gesture recognition, entertainment purposes, Video Games, Sports, etc. It has been observed that in recent years, Human-Computer Interaction (HCI) has moved to another level of popularity in such a way that humans can now interact with computers not only with mouse and keyboard, but with voice, body movements, hand gestures, etc. The basic aim of this type of interaction is to make the experience natural (without any additional hardware attached with the user's body) and user friendly, for which different types of environments have been developed like web browsers, video games, and Virtual Reality (VR) environments. This type of interface is known as Natural User Interface (NUI), where sensors are used to capture the user's interactions, and the system performs actions accordingly.

Gesture recognition systems have also proved to be very effective in the development of assistive technologies for old and handicapped people. For visually impaired or blind people, a lot of applications have been developed for indoor and outdoor navigation to improve the quality of their lives, e.g. on-device sensors to detect dead-reckoning in support with a web-based architecture, to easily create indoor maps for navigation and localization [1]. Several factors make the process of interpretation of gestures a critical one including type of gestures, background, illuminance, number of hands, and space. Gestures can be performed by humans, robots, or any other device. There can be different types of gestures one-handed, two-handed or in some cases you might need to wear gloves [2]. There are times when different parts of the body other than hands being used for making gestures [3]. Usually, arms and hands are the main focus of gesture recognition systems while they disregard the movement of the whole body.

Deaf people cannot use oral languages for communicating with other people. The source of their communication is SL (making different movements and shapes with hands). Facial expressions and body gestures also come into play when communicating through SL. It is considered to be a natural language by Linguists because it shares many commonalities with the oral language.

One misconception about the SL is that people consider it to be universal and deaf people belonging to any part of the world can communicate with people from other parts of the world. SL is mostly subjected to the country and within a country many local SLs' may exist. The exact number of existing SLs' is unknown, but according to a study conducted in 2015, Ethnology has listed them to be 137, which are known [4].

The deaf community of Pakistan uses its SL, like any other country in the world i.e. PSL. It has a distinct style of syntax,

grammar, and vocabulary. A point to understand is that SLs' vary depending upon the variability in regions e.g. American Sign Language (ASL) and British Sign Language (BSL). In the same manner, the PSL alphabet gestures are the exact depiction of the Urdu language alphabets (National Language of Pakistan) [5]. There are 37 alphabets in the Urdu language and each of them is represented by a unique PSL sign shown in Fig. 1.

Figure 1. SL Representation of Urdu Language Alphabets

Majority of the people with hearing disabilities belong to third world countries. These countries are mostly financially very unstable which results in poor services to the deaf community. According to the WHO, there are around 436 million people in the world with hearing disabilities [6

There are around 60 local languages that are spoken in Pakistan including Hindko. Hindko and many other local languages spoken by minorities in Pakistan are on the brink of being endangered because of a lack of developmental work to promote languages. In this research Hindko Language (a local language in Pakistan) is selected for research to revive the dying language. Hindko has a total of 50 alphabets, with 37 alphabets the same as in Urdu and the rest of them being new. Some of the alphabets in Hindko language [7], are shown in Fig. 2 for which there are no SL gestures available at present for blind people to understand. In this research not only, we developed gestures for new alphabets but have also generated a system to recognize all of them.

ٹ Tah	ٹھ Tha	ٹ Tah	پھ Pha	پ Pah	آ Aaa
ڻ Nra	کھ Kha	ک Kah	چھ Chhah	چ Chah	ٹھ Tha

Figure 2. Hindko Language Alphabets with No Sign Gestures

The scope of this research work is limited to HSL recognition by creating a recognition and classification system for which a new dataset had to be developed. The organization of the research work is as follows. Section 2 provides background information and research work done in SLR. Section 3 presents the methodology including the classification model and algorithm. Section 4 covers the discussion on results. Finally, section 5 presents the conclusions and future work.

II. Literature Review

Deep learning being a branch of machine learning works on the principle of the brain's neural networks and uses algorithms developed based on the brain's structure and function. The learning occurs either in a supervised or unsupervised manner. Supervised learning is based on making interpretations from labeled data, while unsupervised learning involves making interpretations from unlabeled data. This research work is based on supervised learning. The models in deep learning are defined as Artificial Neural Networks (ANNs'). ANN's structure is created based on a collection of connected units called neurons. Neurons can transmit signals with the help of connections between them. Typically, the organization of neurons is in the form of layers.

Recognition of bodies and body parts have been a widely studied topic, for which a variety of application have been developed. In this context the techniques used in deep learning have enhanced the capability of the systems to understand in an accurate and automated manner. Deep learning is also known as a convenient tool in image processing applications, computer vision, text classification, robotics and control, and so on [8]. CNN also has been popularly used for analyzing images, data analysis, and other related classification problems. CNNs' are known to have been designed in such a specialized manner that they can detect patterns and make sense of them. With the help of pattern detection, CNNs' are very useful for image analysis because they can be used on massive collections of images. CNNs' can learn valuable characteristic features for a huge database of images and outperforms traditional features like Histogram of Oriented Gradients (HOG), Local binary pattern (LBP), or Speeded Up Robust Features (SURF) [9].

An easy way to use CNN in a much more efficient manner to minimize the efforts and training time, is to utilize a pre-trained CNN for feature extraction. In this particular research work, a specific architecture of CNN has been used, named as "GoogLeNet [10]". The reuse of pre-trained networks is also known as transfer learning, in which weights from the pre-trained networks are used. Since, pre-trained networks have already detected the low-level features like edges, lines, and curves, which are frequently computationally expensive in terms of time, skipping over these parts helps the network to achieve an exceptional result in less time rather than starting the training process all over again. In this research paper, pre-trained GoogLeNet is used to recognize 50 different HSL alphabets shown in Fig. 3.

Figure 3. Sign Language Representation of Hindko Alphabets

People with hearing disability live a compromised life, to improve the standard of their lives a lot of different techniques can be implemented, one of them is the conversion of gesture to text and vice versa. A method to convert gesture to text and text to gesture for Hindi Language has been developed, the images of palm gestures were given as an input to extract the region of hand, based on skin color segmentation [11]. A 5-bit binary string to extract features was used to identify 32 different gestures. The algorithm proved to be robust to variability in hand orientation and changes in scale making the algorithm efficient in recognizing images accurately.

Hand Gesture Recognition (HGR) approach is the most appropriate way for recognition and acceptance. A method was proposed for ASL recognition using CNN for feature extraction using Hand Gestures and further classifying them based on Multi-Class Support Vector Machine (MCSVM) [12]. The model shows to perform satisfactorily in terms of classification accuracy, i.e., 94.57%. To support the argument a technique has been proposed in [13], in which the AdaBoost classifier based on Haar is used for the hand segmentation and discrimination of skin color respectively.

To minimize the barrier between deaf and normal people a technique is proposed for recognition of hand gestures for PSL alphabets in unobstructed scenarios [14]. To acquire images, a digital camera was used with a random background. After the images were acquired, they were are pre-processed for hand detection using skin classification filter. For feature extraction Discrete Wavelet Transform (DWT) was used and eventually for sign recognition ANN with backpropagation learning algorithm was used.

An application based on android platform for mobile phones to detect hand motion for the generation of basic mobile commands [15], stored the hand gestures in the application and were re-labelled. Kanade-Lucas-Tomasi (KLT) algorithm was then used for feature point extraction to detect the static gestures of the hand to call applications installed on the phone.

In another method, the authors introduced a system to capture an image using a webcam, removing the hand from the background by segmentation, extracting features using Principal Component Analysis (PCA), and finally classifying PSL gestures with the help of K Nearest Neighbors (KNN) [16].

In comparison to the previous work done in SL detection, the work done in this research opens new dimensions for a pre-trained GoogLeNet for the detection of HSL images and it has proved to be more accurate than other state-of-the-art machine learning algorithms [12, 14, 16, 17, 18, 19]. A detailed comparison is discussed in section 4. There has been a considerable amount of work done on PSL but it is very unfortunate that some of the local languages of Pakistan are yet to be explored. In this research work we have selected "Hindko" as a local language for which an automated gesture recognition system has been developed.

III. METHODOLOGY

CNNs' are different from regular neural networks in terms of architecture. Typically, a CNN comprises 3 architectural layers including convolutional, pooling, and fully connected layers. A feature map of the input data is generated after carrying out a series of steps involving convolution and pooling processes. Every feature map generated through the convolutional layer is combined together to form the final output. The convolutional layer plays an important role in building a CNN, hence results in high time costs for the training process. The convolutional layers have parameters in the form of shared sets of weights, having minimal receptive fields. In the pooling layer, the input is separated into 2 groups i.e. non-overlapping frames and the maximum for each group using nonlinear down sampling (Max pooling). The number of parameters, overfitting, and computational complexity is reduced by max-pooling layers. Hence, the max-pooling layer is normally placed between the convolutional layers. The dropout layer then drops neurons with a certain probability [20] and Non-saturating ReLU (Rectified Linear Unit) helps the network to learn complex patterns. Finally, fully connected layers work as a classifier having all of the neurons in a fully connected layer being fully connected to all of the outputs of the previous layer. Nevertheless, it is very important to mention that CNN training from the beginning is a time-consuming task on a very large training dataset, which is mostly unavailable and can cause overfitting. Hence, in this research paper, a pre-trained GoogLeNet architecture with proper fine-tuning was applied.

There are 3 different sizes of the filters used in GoogLeNet i.e. 1x1, 3x3 and 5x5 on the same image for dimensionality reduction shown in Fig. 4. Finally, for the same image all of the features are combined to produce a robust output. GoogLeNet comprises 22 layers and the number of parameters is reduced from 60 million to 4 million when compared to AlexNet [21].

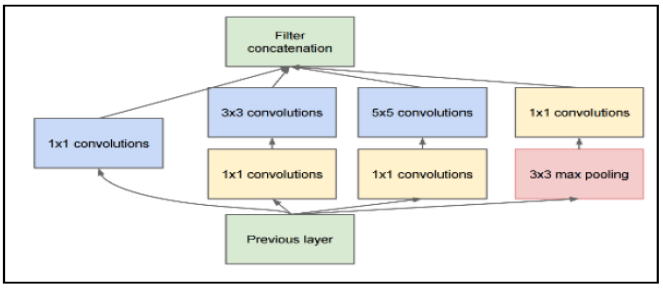

Figure 4. Multiple Convolutions for Dimension Reduction

The basic structure of GoogLeNet can be seen in [10]. It shows the overall network's architecture. Multiple Inception modules combined together form a deeper structure through which high accuracy is achieved.

A. *Pre-trained Networks*

A pre-trained network has pre-trained weights, these weights can be utilized in other similar tasks. A pre-trained network may also perform well in training CNNs where there is a limited sized dataset. In addition to that, training CNN from the beginning is time-consuming as well as computationally expensive. Rather than training the network from the beginning, the best way to retrain a network is to perform fine-tuning through which initial layers of the network are kept frozen and the last set of fully connected layers are replaced with the newly trained layers, so that errors cannot propagate backwards [22]. Stochastic gradient descent (SGD) algorithm was used for the initialization and training of weights in the network.

Each layer in GoogLeNet works as a filter. This formation improves the capability of GoogLeNet in better feature detection in images. Common features like blobs, edges, and colors are detected in the first layer, while high-level features are detected in the last layers. In this research work, the working procedure was divided into 2 phases (Training and Evaluation) is shown in Fig. 5. All of the training images were pre-processed for resizing and data augmentation, where operations like random reflection on the horizontal axis and 30-pixel range translation over the x and y axis to avoid overfitting and memorization of the exact details during the training process [23]. Following are the steps categorized in Training and Evaluation separately:

Training:

1. Pre-Trained Network Loading.
2. Final Layers Replacement.
3. Train Network with new pre-processed images.

Evaluation:

1. Give a pre-processed input image.
2. Extract the features using trained CNN.
3. Classify the image.

Let Z be a set of SL images, j representing the resized image z in set Z, G and G1 representing pre-trained and modified GoogLeNet network and ACC be the element representing accuracy. The methodology is depicted in the Algorithm.

B. *Algorithm*

Input: Input Image.
Output: Accuracy Achieved.

1. ∀ z ∈ Z, ∃ j ∈ Z: j be pre-processed z
2. Let G be a pretrained GoogLeNet network ∈ CNN
3. Modify G to G1 ∈ CNN
4. Train G1 ∀ j ∈ Z.
5. Let ACC be the accuracy
6. ∀ G1, ∃ ACC be the achieved accuracy → CNN (Z)

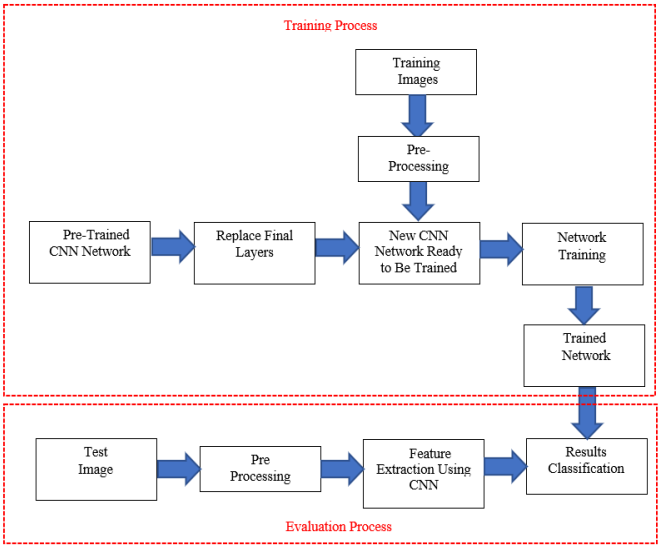

Figure 5. Working Procedure of The Model

IV. RESULTS

This section describes the results for the evaluation of 50 classes of HSL images. The Schematic diagram of the overall system is shown in Fig. 6. The performance of GoogLeNet is evaluated using the accuracy achieved. The accuracy was achieved by finding the number of correctly classified images.

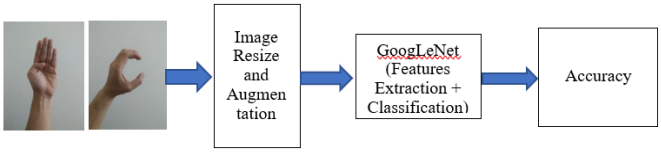

Figure 6. Schematic Diagram of The Overall System

To evaluate the most appropriate settings the experiments were performed for 3 different learning rates i.e., [0.01, 0.001, 0.0001], with the size of mini-batch and number of epochs set to 128 and 10 respectively.

The simulations were ran using MATLAB 2019a. The network was trained on an individual system with the specification of an Intel Core Processor i5-8250U with 4 cores at 1.6 GHz and 16 GB RAM.

The images were collected on our own. 16,500 images were captured using a mobile camera, including 330 instances of each alphabet. Images were divided into training and testing sets. 70% i.e. 11,550 images were used for the training process

and the remaining 30% i.e. 4,950 images for validation. The original images were of resolution 1080 × 1920 × 3 (1080 width, 1920 height, RGB). The images had to be resized to 224 × 224 × 3 pixels to fit in the GoogLeNet. It categorizes the input image in one of the 50 classes of HSL. Table 1 summarizes the evaluations of the network with different learning rates.

TABLE I. GoogLeNet Settings With 3 Different Learning Rates

Network Configuration and Accuracy Achieved	Learning Rates		
	0.01	**0.001**	**0.0001**
Mini-Batch Size	128	128	128
Epochs	10	10	10
Iterations	900	900	900
Iterations per Epoch	90	90	90
Validation Frequency	90	90	90
Time Consumption	240 minutes	245 minutes	250 minutes
Accuracy	99.98%	99.92%	99.78%

GoogLeNet trained with a learning rate of 0.01 achieved an accuracy of 99.98%. The value to be chosen for the learning rate requires some testing. We had to test and tune with each model before we knew exactly where we wanted to set it. The idea is to set the value somewhere from 0.01 to 0.0001. Fig. 7, 8, 9 shows the training processes of GoogLeNet with 3 different settings as shown in table 1. It can also be noticed that with decreasing the learning rate the accuracy has reduced and the time consumption has increased because with lower learning rate the system learns slowly and takes more time to complete.

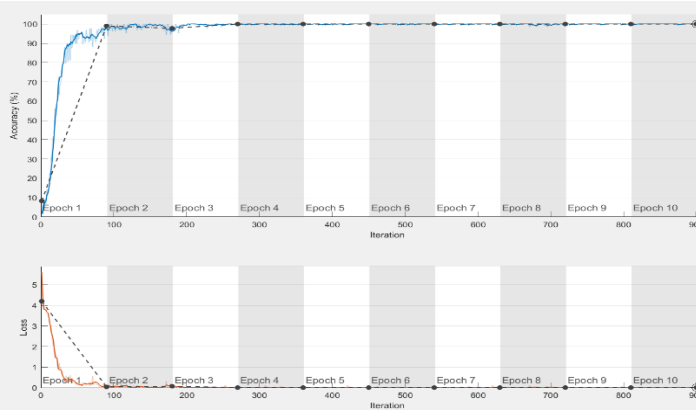

Figure 7. 99.98 % Accuracy Is Achieved With 0.001 Learning Rate

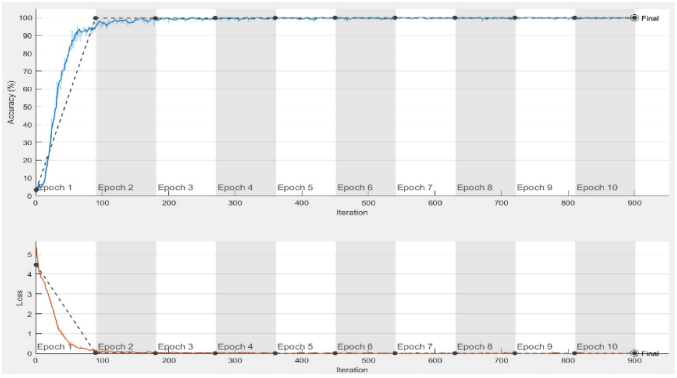

Figure 8. 98.92 % Accuracy Is Achieved With 0.001 Learning Rate

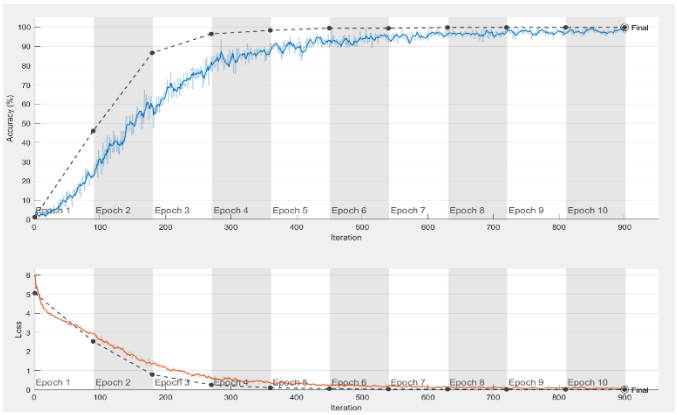

Figure 9. 99.78 % Accuracy Is Achieved With 0.001 Learning Rate

V. DISCUSSION

Based on the results, we came to the conclusion that for the method presented in this research paper, a high learning rate of 0.01 provides the best result. This is the best result achieved in SLR in comparison with other state-of-the-art techniques for classification of SLs', as shown in Figure 11. Detection of ASL signs Using the Leap Motion Controller with Machine Learning Approach achieved an accuracy of 93.81% [17]. Another methodology shows a simple algorithm, which was used for feature extraction to recognize ASL alphabets using hand gestures and then on the basis of the extracted features, ANN achieved 95% classification accuracy [18]. A pre-trained CNN using AlexNet architecture and MCSVM achieved an accuracy of 94.5% [12]. To detect PSL gestures, images were captured through a webcam, after that the images were segmented to separate hand from the background PCA and then finally classified the gesture feature by utilizing KNN achieved an accuracy of 85% [16]. A system using a DWT for feature extraction and ANN with a back-propagation learning algorithm is employed to recognize PSL alphabets achieved an accuracy of 86.40% [14]. In an automatic recognition system of PSL alphabets using MCSVMs' validated over the data set of 3414 PSL signs achieved 77.18% accuracy [19].

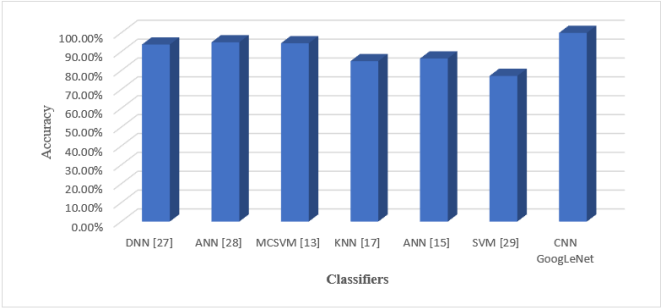

Figure 10. Comparison of State-Of-The-Art Techniques for Classification of SL's

The comparison shown in Fig. 10 indicates that the methodology implemented outperforms other state-of-the-art methods for sign language detection. The classification accuracy obtained satisfactorily, which is 99.98%. Fig. 11 shows the sign wise classification accuracy of HSL alphabets.

ت 99.9%	پھ 100%	پ 100%	پ 100%	ب 100%	ا 99.9%	آ 99.9%
ج 100%	ش 100%	ٹھ 100%	ٹ 100%	ٹ 99.9%	تھ 100%	ث 100%
ڈ 99.9%	د 100%	خ 100%	ح 100%	چھ 100%	چ 99.9%	چ 100%
ش 100%	س 100%	ژ 99.9%	ز 99.9%	ڑ 100%	ر 100%	ذ 100%
ف 100%	غ 98.9%	ع 99.7%	ظ 99.9%	ط 100%	ض 100%	ص 100%
م 99.9%	ل 100%	گ 100%	کھ 100%	ک 99.9%	ک 100%	ق 100%
ی 100%	ء 99.9%	ہ 100%	و 99.9%	ں 99.9%	ٹ 100%	ن 99.9%
						ے 100%

Figure 11. Classification Accuracy of HSL Alphabets

CONCLUSION

In recent times, deep learning has been frequently used in research which has yielded a significant improvement in automated analysis and recognition in comparison with traditional machine learning algorithms. The innovativeness of this research work lies in illustrating the use of pre-trained CNN network for recognition of SL images. A pre-trained GoogLeNet network model was modified according to the requirements of this particular work and achieved high accuracy ad excellent results. The exceptional detection rates in this research work are expected to reinforce its use in Sign Language Detection. Furthermore, for the first time, a pre-trained CNN is used for Hindko Sign Language recognition, which has the possibility to lay the foundation for employing pre-trained CNNs within a CAD system for accurate detection. In this paper static images have been used to recognize Hindko sign language. Hindko has 13 more alphabets than Urdu, and for these 13 alphabets new gestures have been developed, which will prove to be a valuable addition for further research. Since, this work is limited to static signs, in future this work can be given a more practical form by installing it into an embedded system with the webcam.

REFERENCES

[1] Verma, S., Omanwar, R., Sreejith V., and Meera, G. S., "A Smartphone Based Indoor Navigation System", 28th IEEE International Conference on Microelectronics, pp. 345-348, 2016.

[2] Varga, R., and Prekopcsák, Z., "Creating a Database for Objective Comparison of Gesture Recognition System", 15th International Student Conference on Electrical Engineering, pp. 1-6, 2011.

[3] Sigalas, M., Haris, B., and Panos, T., "Gesture Recognition Based on Arm Tracking for Human-Robot Interaction", IEEE International Conference on Intelligent Robots and Systems, pp. 5424-5429, 2010.

[4] Lewis, M. P., Simons, G. F., and Fennig, C. D., "Ethnologue: Languages of The World", Texas: SIL International, 2015. [Online]. Available: https://www.ethnologue.com/sites/default/files/Ethnologue-18-Honduras.pdf

[5] Sulman, D. N., and Zuberi, S., "Pakistan Sign Language–A Synopsis", Pakistan, June, 2000. [Online]. Available: https://www.academia.edu/2708088/Pakistan_Sign_Language_-_A_Synopsis

[6] Organization, W. H., "10 Facts About Deafness", Posje´ceno, Vol. 14, pp. 2017, 2017.

[7] Toker, H., "A Practical Guide to Hindko Grammar", Trafford Publishing, 2014. [Online]. Available: https://www.scribd.com/book/387784340/A-Practical-Guide-to-Hindko-Grammar

[8] Han, J., Zhang, D., Cheng, G., Liu, N. and Xu, D., "Advanced Deep-Learning Techniques for Salient and Category-Specific Object Detection: A Survey", IEEE Signal Process. Mag., Vol. 35, No. 1, pp. 84–100, 2018.

[9] Lee, A., "Comparing Deep Neural Networks and Traditional Vision Algorithms in Mobile Robotics," Traditional Vision Algorithms in Mobile Robotics", 2016.

[10] Krizhevsky, A., Sutskever, I., and Hinton, G., "Image Net Classification with Deep Convolutional Neural Networks", In Proceedings of The Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, pp. 1097–1105, 2012.

[11] Chaman, S., D'souza, D., D'mello, B., Bhavsar, K., and D'souza, T., "Real-Time Hand Gesture Communication System in Hindi for Speech and Hearing Impaired", IEEE International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1954-1958, 2018.

[12] Islam, M. R., Mitu, U.K., Bhuiyan R. A., and Shin, J., "Hand Gesture Feature Extraction Using Deep Convolutional Neural Network for Recognizing American Sign Language", IEEE 4th International Conference on Frontiers of Signal Processing (ICFSP), pp. 115-119, 2018.

[13] Sun, J. H., Ji, T. T., Zhang, S. B., Yang, J. K., and Ji, G. R., "Research on the Hand Gesture Recognition Based on Deep Learning", 12th IEEE International Symposium on Antennas, Propagation and EM Theory (ISAPE), pp. 1-4, December, 2018.

[14] Khan, N., Shahzada, A., Ata, S., Abid, A., Khan Y., and Farooq, M.S., "A Vision-Based Approach for Pakistan Sign Language Alphabets Recognition", Pensee, Vol. 76, No. 3, pp. 274-285, 2014.

[15] Dadiz, B.G., Abrasia, J. M. B., and Jimenez, J. L., "Go-Mo (Go-Motion): An Android Mobile Application Detecting Motion Gestures for Generating Basic Mobile Phone Commands Utilizing KLT Algorithm", IEEE 2nd International Conference on Signal and Image Processing (ICSIP), pp. 30-34, 2017.

[16] Malik, M. S. A., Kousar, N., Abdullah, T., Ahmed, M., Rasheed, F., and Awais, M., "Pakistan Sign Language Detection using PCA and KNN", International Journal of Advanced Computer Science and Applications, Vol. 9, No. 54, pp. 78-81, 2018.

[17] Chong T.W., and Lee, B.G., "American Sign Language Recognition Using Leap Motion Controller with Machine Learning Approach", Sensors, Vol. 18, No. 10, pp. 3554, 2018.

[18] Thongtawee, A., Onamon, P., and Yuttana, K., "A Novel Feature Extraction for American Sign Language Recognition Using Webcam", 11th IEEE Biomedical Engineering International Conference (BMEiCON), pp. 1-5, 2018.

[19] Shah, S. M. S., Naqvi, H. A., Khan, J. I., Ramzan, M., and Khan, H. U., "Shape Based Pakistan Sign Language Categorization Using Statistical

Features and Support Vector Machines", IEEE Access, Vol. 6, pp. 59242-59252, 2018.

[20] Goodfellow, I., Bengio Y., and Courville, A., “Deep Learning”, MIT Press: Cambridge, MA, USA, 2016. [Online]. Available: www.deeplearningbook.org

[21] Krizhevsky, A., Ilya, S., and Geoffrey, H. E., "ImageNet Classification with Deep Convolutional Neural Networks", Advances in neural information processing systems, pp. 1097-1105, 2012.

[22] Rumelhart, D., Hinton, G., and Williams, R., “Learning Representations by Back-Propagating Errors,” Nature, Vol. 323, pp. 533–536, 1986. [Online]. Available: https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf

[23] Mathworks, “Transfer Learning Using GoogLeNet - MATLAB Simulink - MathWorks United Kingdom”, 2018. [Online]. Available: https://uk.mathworks.com/help/nnet/examples/transfer-learning-using-googlenet.html