Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

Static Gesture Recognition for Indian Sign Language Alphabets and Numbers using SVM with ORB Keypoints and Image Pixel As Feature

Rakesh Savant
Faculty of Computer Science,
Babu Madhav Institute of Information Technology,
Uka Tarsadia University, Bardoli, India

Jitendra Nasriwala
Faculty of Computer Science,
Babu Madhav Institute of Information Technology,
Uka Tarsadia University, Bardoli, India

Preeti Bhatt
Faculty of Computer Science,
Babu Madhav Institute of Information Technology,
Uka Tarsadia University, Bardoli, India

Abstract **- Sign language is used by deaf-dumb people to communicate with each other. They cannot communicate with normal people as normal people do not understand sign language. To bridge this gap there is a need for a communication medium. The sign language recognition problem belongs to the gesture recognition domain. There are two types of signs, static and dynamic, respectively. The given study covers the methodology for recognition of static signs (35 classes includes 26 alphabets and 9 numbers) of Indian Sign Language (ISL). The work aims to fulfil the fingerspelling application by recognizing alphabets and numerals of ISL. The SVM is used to classify different characters of Indian Sign Language. There are two feature extraction approaches discussed in this paper. In the first approach, an image is used as a feature to feed the classifier, and in the second approach, the ORB keypoint-based feature extraction method is used to train the classifier. The trained models stretch to predict the ISL sign images taken under the real-time(uncontrolled) situation.**

Keywords – Indian Sign Language, Static gesture, SVM, ORB keypoints, image pixel as feature

I. INTRODUCTION

Spoken language is used as a medium of communication between normal human beings. Without spoken language, it would not be possible for a large population to communicate. Even with the existence of spoken language, a part of the population with speech and hearing disabilities cannot communicate with the majority of the people. Sign language comes to the aid of such a community. Deaf-dumb people use sign language as a mode of communication. But they face problems while communicating with normal people because they do not have prior knowledge of sign language. In many cases, a human translator is available who translate the sign language to a normal human-understandable form. Still, the availability of a human translator is not feasible every time. Sign language is well-structured, and every sign gesture has a meaning. Sign language has its alphabets, numbers, and reach set of word dictionaries. Sign language uses hands, movements of hands, and expressions to represent the alphabet, number or dictionary word[1]. Sign language recognition is the problem that belongs to gesture recognition. Broadly the gestures are static and dynamic, respectively[2]. Static gestures are used to present signs with the stable hand/s without any movements. There are non-dictionary words like a person's name, things, etc where alphabets and number recognition play a significant role.

The given study covers recognition of static gestures of Indian Sign Language (ISL), i.e. alphabets and numbers using SVM. The alphabets and numbers use either one hand or two hands to represent particular sign characters. This set of characters belongs to the category of static gestures[2]. The dataset used in the study has different 35 classes of ISL characters (26 alphabets and 9 numbers of ISL). For each class, there are 1200 images captured and stored. There are 42000 images available in the dataset for different 35 classes[3]. All the ISL alphabets and numbers are given in **Figure 1.**

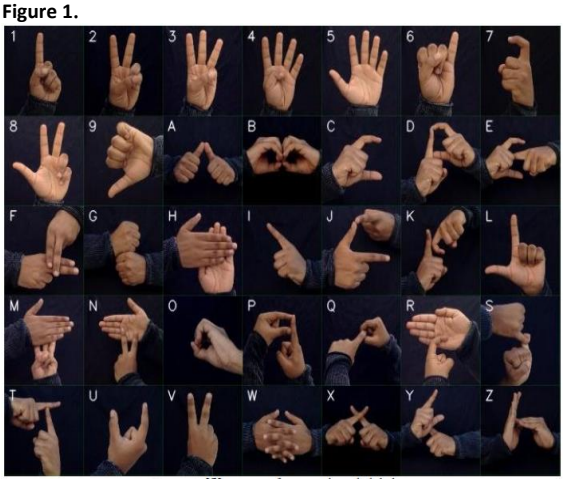


Figure 1 ISL gestures for numbers and alphabets [3]

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

II. RELATED WORK

In this section, various techniques and approaches are discussed which used by researchers to recognize static gestures of ISL. T. Sajanraj et al. [4] present a technique to recognize the numbers of ISL using the Deep Learning technique (CNN). They have achieved 99.56% accuracy for the same subjects and 97.26% accuracy in low lighting conditions. A. Sood et al. [5] works on static gesture recognition where they have used Harris Algorithms for feature extraction. V. Aditya et al. [6] present the work on recognition of 36 different characters 26 alphabets, and 10 numbers for fingerspelling. They have used shape feature derived from the distance transformation of the binary image and used ANN as a classifier. They have achieved an accuracy of 91.11%. R. K. Shangeetha et al. [7] presents a technique to recognize single-handed gestures where the state of a finger is used to identify the sign. They have selected the features like the angle between fingers, number of fingers fully closed, semi-closed or fully opened. S. Chattoraj et al [8] have used a scale-invariant feature transform (SFIT) algorithm for feature extraction. S. Reshna et al. [9] present a technique to recognize 11 static symbols using HOG features with SVM as a classifier. S. Shinde et al. [10] calculate the angles and the peaks between the fingers to extract the features. They have used a 12-bit binary sequence for each hand gesture, classifying the different hand gestures for classification. Shravani K. et al [3] have used SIFT and SURF feature extractions techniques with Bag of Words model. Idea of BoW is adapted from Natural language processing (NLP). In image processing, BoW model concept can be called as "histogram-based representation of independent features". So, an image can be viewed as a document in order to depict any gesture using the BoW model. With their approach all the labels are correctly classified by the SVM except gesture of number 2.

III. PROPOSED SOLUTION

We proposed to develop static Indian Sign Language alphabets and a numbers recognition system. The input to the system is the image of a gesture, and the output will be the text that presents the alphabet or number. The steps include input, pre-processing, feature extraction, classification, as shown in **Figure 2.** The image is used as a feature in a given study. Further, the multiclass Support Vector Machine (SVM) classifies various ISL characters.

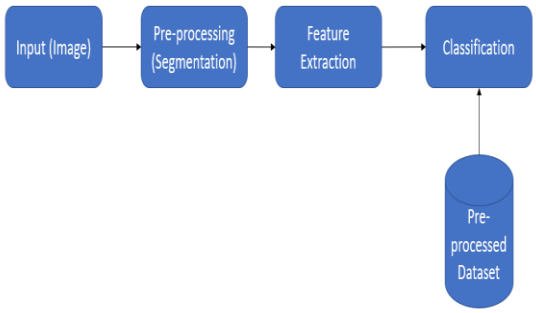


Figure 2 System overview

A. *Dataset Preprocessing*

In this experiment, we have used the ISL alphabets and numbers dataset from Shravani K et al.[3]. The dataset of ISL contains 35 classes, each class with 1200 images. All 35 classes comprise alphabets (A-Z) and numeric (1-9) with a total of 42000 images[3]. The original dataset of different 35 classes is preprocessed by applying neural network-based skin and non-skin segmentation technique, which we proposed. Each image is taken from the original dataset, segmented skin and non-skin pixels in the image using a neural network-based segmentation algorithm, and stored each preprocessed image in appropriate class folders. The skin and non-skin samples dataset is used from UCI Machine Learning Repository[11]. Dataset has a total of 245057 samples. From total samples, 50859(21%) are skin samples and 194198(79%) are non-skin samples. The samples belong to the dataset are Red (R) (0-255), Green (G) (0-255) and Blue (B) (0-255) values of a pixel taken randomly from human face images with labels 1 for skin and 2 for non-skin pixel. The face images of different age groups (young, middle, old), race groups (white, black, Asian), and genders are obtained from FERET and PAL databases[11]. Dataset has 51444 unique samples (28% skin and 72% non-skin), and 11 distinct RGB samples out of 51444 belong to both skin and non-skin classes. Figure 3 shows the sample of each 35 classes before and after preprocessing.

As shown in Figure 3, the images in the preprocessed dataset only contain skin pixels, and other pixels are replaced with black. For batter feature extraction, such a processed dataset gives better features than the original dataset images. Only skin pixels in an image leads to get good classification accuracy.

Figure 3 Dataset samples before and after preprocessing

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

B. Pre-processing (Segmentation)

In this stage, all images are pre-processed for feature extraction. The captured gesture image is in RGB format. The image is converted to a grayscale image with a gray level intensity range from 0 to 255. The grayscale image is resized to 128*128 to reduce the processing time and improve the system's accuracy[2]. The next step is segmentation, in which the neural network-based skin pixel segmentation technique is used, as discussed before. After the segmentation, the skin pixel is available in the image, which represents the hand/s of the signer.

C. Feature Extraction

The next step is feature extraction. As discussed in the dataset preprocessing section, the whole dataset is preprocessed for batter feature extraction. In the given study, we have examined two approaches for feature extraction. In the first approach, we have taken the whole image as a feature, and in the second approach, there are ORB keypoints taken as a feature.

In approach – 1 for feature extraction, we have taken the whole image to train the classifier, where the 128*128 RGB image is converted to grayscale and flattened to generate the feature vector. The 16384 sized feature vector is generated from 128*128 image. The feature vectors are generated for all the images available in the preprocessed dataset. The 2-D array of the feature vector is generated where there are 42000 rows in the array. Each row consists of two columns where the first column holds the 16384-pixel values, and the second column stores the class label of that image. Figure 4 describes the approach - 1.

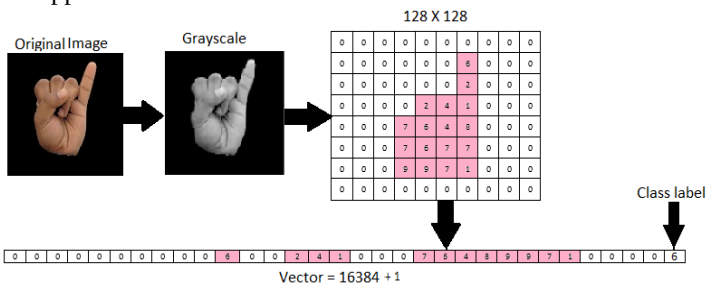


Figure 4 Feature Extraction Approach – 1: Image as feature

The approach - 2 for the feature extraction technique is ORB keypoints-based features. Concerning the research by [12], the ORB is the most efficient feature-detector-descriptor with the least computational cost. Other well-known keypoint detector algorithms like SIFT and SURF are computationally costly, and they are patented algorithms; we cannot use those algorithms for free. The experiments focused on ORB based feature detection technique. The ORB feature descriptor returns the keypoints from the image. We can generate the feature using these key points and use the keypoints vectors to train the classifier. Different 35 classes of images have a different number of keypoints. During the vectorization process, the vector size must be the same to feed the classifier. We found the class with the maximum keypoints to solve the feature vector equalization problem. Next, we created the vector space, which stored the ORB key point descriptor of the maximum number of keypoints and stored the number of total keypoints in the image on the last vector element. Now, to equalize the vector in the case of class with fewer keypoints, we have padded 0 in the vector with the number of keypoints stored at the last vector element. We have generated the feature vectors to feed the classifier by following this method. Table 1 shows the number of keypoints for random images taken from the dataset for different classes. As per analysis, the image from class B gives the highest number of keypoints. To equalize the feature vector, we have to consider the feature vector size of 200 with padding of 0 after the keypoint values, and at the last element of the vector, we have stored the total number of keypoints.

TABLE 1. CLASSWISE NUMBER OF KEYPOINTS (FROM RANDOM DATASET IMAGES)

Class Name	Number of keypoints	Class Name	Number of keypoints
1	151	J	44
2	120	K	119
3	121	L	117
4	108	M	113
5	77	N	127
6	124	O	101
7	59	P	129
8	90	Q	103
9	117	R	154
A	143	S	109
B	174	T	115
C	101	U	79
D	128	V	33
E	135	W	164
F	68	X	127
G	86	Y	118
H	103	Z	87
I	61		

Figure 5 demonstrates the ORB keypoints detection for sample images from the dataset. The corners, edges and blobs in the image are considered as keypoints.

Figure 5 ORB Keypoint detection

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

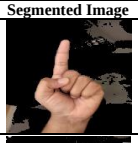
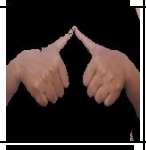
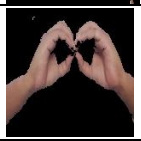
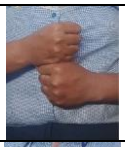
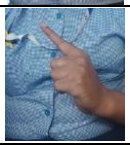
D. Classification

1) Approach 1 – Image as a feature:

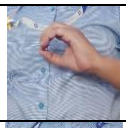
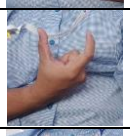
After generating the feature vectors as described in for image as a feature, the next step is classification. It is essential to divide the dataset for training and testing for classification. We have split the dataset 80:20 ratio. Here, 80% of the dataset is used to train, and 20% of the dataset is used to test the classifier. From 42000 images in the dataset, there are 33600 images used for training and 8400 images for testing. We have used Support Vector Machine(SVM) as a classifier. The classifier gives 99.5% accuracy for training and testing the dataset with the image as the feature method. The model gives correct predictions while predicting the class of any image from the dataset. The reference [3] has achieved an accuracy of 99.98%, but their approach does not classify class 2.

For the validity of the classification, we have tested the trained model for the different 30 classes images taken under an uncontrolled environment. The model predicts 50% of the images' classes taken under an uncontrolled environment. Table 2 shows the results for different 30 classes prediction.

TABLE 2 PREDICTION BASED ON THE IMAGES UNDER UNCONTROLLED ENVIRONMENT (APPROACH -1)

Class	Original Image	Segmented Image	Prediction
1			No
2			No
3			No
4			No
5			Yes
6			No
7			No
8			Yes
9			No
A			No
B			No
C			Yes
D			Yes
E			Yes
F			Yes
G			Yes
H			Yes
I			Yes
J			Yes
K			Yes
L			Yes
M			No

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

O			Yes
T			No
U			Yes
V			No
W			No
X			No
Y			Yes
Z			No

2) Approach 2 –ORB keypoints as a feature:

As per the second approach for features, i.e. ORB keypoints, we have used the ORB keypoint vectors to train the SVM classifier. As per the previous approach, we followed an 80:20 ratio for training and testing splits. The classifier gives 99.7% accuracy for training and testing on the dataset with the ORB keypoints feature.

For the validity of the classification, we have tested the trained model for the different 30 classes images taken under an uncontrolled environment. The model predicts 27% of the images' classes taken under an uncontrolled environment.

IV. RESULTS AND DISCUSSION

By experiments with SVM-based classification, while we have selected the image as a feature and fed the classifier, the accuracy given by the model is 99.5%, and the classifier correctly predicts each image of the dataset. Further, we have provided sign images other than dataset images; in that case, the model predicts 5, 8, C, D, E, F, G, H, I, J, K, L, O, U, and Y class correct (as per **Table 2**). Character V and 2 have equal orientation, and the model gives wrong predictions in real-time images. The model predicts the sign of character 2 as V. There are certain limitations of approach – 1 as shown below.

- It is non-invariant to rotation and scaled images. This approach may fail if the hand/s in the image is rotated or scaled.
- The number of features per vector is very high, i.e. 16384. The larger number of features cost the high computation for the classification process.
- This approach is sensitive to noise. If the background of the hand region in the image is the same as the skin tone, then the prediction might be wrong.

ORB keypoint detection is a better feature detection technique invariant to rotation, scale, and non-sensitive noise. Figure 6 demonstrates ORB keypoint detection and matching with scaled and rotated hands in different images. The keypoints are considered a feature, and the number of features is less than the approach – 1. The approach – 2 results under an uncontrolled environment are not impressive. Approach – 2 gives less prediction accuracy than approach - 1, but it has more capabilities to overcome the limitations of the first approach.

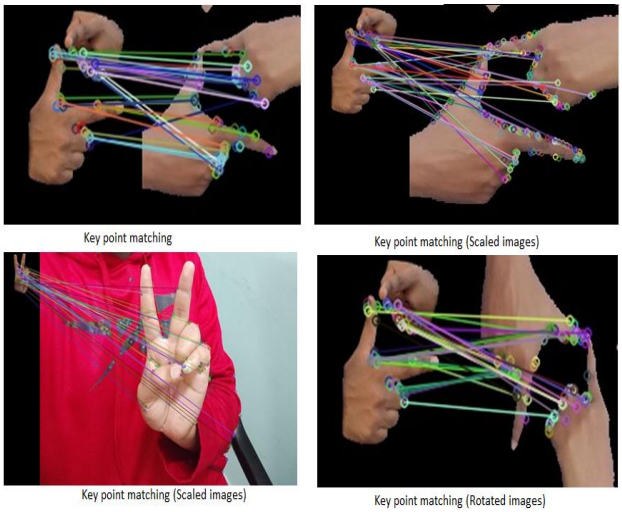


Figure 6 ORB keypoint detection for scaled and rotated hand images

V. CONCLUSION

Indian Sign language recognition is essential to research for deaf-dumb communication in India. The given research covers the recognition of static signs of Indian Sign language. The objective of the work is to recognize ISL alphabets and numbers for fingerspelling applications. The different 35 classes, including 26 ISL alphabets and 9 digits, are experimented. We have performed classification using SVM with two feature extraction techniques and compared and analyzed both approaches with their strength and limitations. Few experiments are performed on the sign images taken under a real-time(uncontrolled) environment for both approaches. The ORB keypoint detection is invariant to scale and rotation and overcomes the limitations of approach - 1. ORB does not perform better for classification on real-time images than approach -1. Still, we can improve ORB keypoint-based classification accuracy by applying good feature engineering techniques.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 01, January-2022

REFERENCES

[1] Das, Aditya, et al. "Sign language recognition using deep learning on custom processed static gesture images." 2018 International Conference on Smart City and Emerging Technology (ICSCET). IEEE, 2018.

[2] Nagarajan, Sathish, and T. S. Subashini. "Static hand gesture recognition for sign language alphabets using edge oriented histogram and multi class SVM." International Journal of Computer Applications 82.4 (2013).

[3] K. Shravani, A. Lakshmi, Sree, M. Sri Geethika, and K. Dr.Sapna B, "Indian Sign Language Character Recognition," *IOSR J. Comput. Eng.*, vol. 22, no. 3, pp. 14–19, 2020.

[4] Sajanraj, T. D., and M. V. Beena. "Indian sign language numeral recognition using region of interest convolutional neural network." 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE, 2018.

[5] Sood, Anchal, and Anju Mishra. "AAWAAZ: A communication system for deaf and dumb." 2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO). IEEE, 2016.

[6] Adithya, Vinod, P. R. Vinod, and Usha Gopalakrishnan. "Artificial neural network based method for Indian sign language recognition." 2013 IEEE Conference on Information & Communication Technologies. Ieee, 2013.

[7] Shangeetha, R. K., V. Valliammai, and S. Padmavathi. "Computer vision based approach for Indian Sign Language character recognition." 2012 International Conference on Machine Vision and Image Processing (MVIP). IEEE, 2012.

[8] Chattoraj, Subhankar, Karan Vishwakarma, and Tanmay Paul. "Assistive system for physically disabled people using gesture recognition." 2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP). IEEE, 2017.

[9] Reshna, S., and M. Jayaraju. "Spotting and recognition of hand gesture for Indian sign language recognition system with skin segmentation and SVM." 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). IEEE, 2017.

[10] SHINDE, SHWETA SONAJIRAO, and RM AUTEE. "Real TIME Hand Gesture Recognition and Voice Conversion System for Deaf and Dumb Person Based on Image Processing." JournalNX 2.9 (2016): 39-43.

[11] A. D. Rajen Bhatt, "Skin Segmentation Data Set," *UCI Machine Learning Repository*. https://archive.ics.uci.edu/ml/datasets/Skin+Segmentation.

[12] Tareen, Shaharyar Ahmed Khan, and Zahra Saleem. "A comparative analysis of sift, surf, kaze, akaze, orb, and brisk." 2018 International conference on computing, mathematics and engineering technologies (iCoMET). IEEE, 2018.