Rabea Ahmed

1711894042

Week 3 - Paper Review

CSE499B

**Bangla Sign Language Recognition with Skin Segmentation and Binary Masking**

November 2019

**Abstract**

The barrier between the hearing and speech im-

paired and society has been bridged through sign language for

ages. Hereby, we present a system that recognizes Bangla Sign

Language (BdSL) through the segmentation of hand gestures,

binary masking and SVM classiﬁcation. In the system, the

features of different notations have been carefully distinguished

and extracted to improve accuracy compared to the existing

work. Through conversion into YCbCr color space and Fuzzy C-

means clustering, a binary image is produced in the system which

is later undergone binary masks designed for particular skin

segmentations. Since YCbCr color space provides an effective

use of chrominance information for modeling the human skin

color, this color space has been used to pre-process the images.

The dataset used in this system has been created and curated to

perform under the manual feature extraction process. Extracting

about 18 different features of the hand gestures of 10 alphabets,

they are compared, classiﬁed and trained using a multiclass

SVM classiﬁer. Multiclass SVM classiﬁer is extensively used for

classiﬁcation of high dimensional spaces and also effective for a

smaller dataset. Using this methodology, the system can achieve

an accuracy of 99.8% for static hand poses

The hearing and speech impaired society has been bridged through sign language. Therefore, a system that recognizes Bangla Sign Language (BdSL) through the segmentation of hand gestures is important, on which this project is based on. In the system, the features of different notations have been carefully distinguished and extracted to improve accuracy compared to the existing work. In this project, about 18 different features of the hand gestures of 10 alphabets are extracted, compared, classiﬁed and trained using a multiclass SVM classiﬁer. Using this methodology, the system can achieve an accuracy of 99.8% for static hand poses

**Introduction**

Visual and manual modes are used in Sign Language in order to communicate. It is a complex mechanism, which is why with the progress in the field of computer vision, a lot of research has been done to detect Sign Language. However, the amount is rather small in case of detection of BdSl. In BdSL there are 38 symbols of which 9 are vowels and 27 are consonants... Machine Learning classifier, Support Vector Machine (SVM) was used by a lot of researchers to detect each alphabet from the hand gestures. But image classifiers, like CNN, are now popular among scientists to detect Sign Languages from images. In this paper, the researchers proposed a novel model which is a combination of Image Network (CNN) and Pose Estimation Network.

**Literature Review**

Through conversion into YCbCr color space and Fuzzy C-means clustering, a binary image is produced in the system which is later undergone binary masks designed for particular skin segmentations. Since YCbCr color space provides an effective use of chrominance information for modeling the human skin color, this color space has been used to pre-process the images. The dataset used in this system has been created and curated to perform under the manual feature extraction process. Extracting about 18 different features of the hand gestures of 10 alphabets, they are compared, classiﬁed and trained using a multiclass SVM classiﬁer. Multiclass SVM classiﬁer is extensively used for classiﬁcation of high dimensional spaces and also effective for a smaller dataset. Using this methodology, the system can achieve an accuracy of 99.8% for static hand poses

**Methods**

First, a database was created with hand gestures of 10 Bangla alphabets. 40different images of different skin color, hand measurements and rotational difference for each picture was captured and processed to curate a suitable dataset for the system. The ROI (Region of Interest) has been identiﬁed and each image has been resized to perform all the operations. The pixel ratio, ﬁxed background, pixel sizes, and ﬁxed hand gestures are necessary conditions for preparing the data for the operations.

The binary image is segmented into eight parts and masked using a vertical and a horizontal mask to extract distinguished characteristics. Since each hand gesture is unique, every segment contains different ratios of pixels. Through binary masking, it gets easy and effortless to differentiate and compare the result vs pixel plot for each segment of different alphabets. A 32x3 matrix was used as the ﬁrst mask.

A different mask is used next which has dimension 3x32.Every image is segmented vertically into 8 blocks and the mask is slid over the blocks vertically. After performing AND operation and FFT, 8 more features are extracted for each image. Similar to the horizontal segmentation, the ratio of pixels for each segment is non identical for every alphabetical notation which produces unique peak amplitude for every division of every binary image.

The area of each segment of the binary image is measured and added to calculate the surface area of the hand used for each gesture. Since our input is a bi-level image, the area is nothing but the number of pixels in the image. That is also why the ratio of black and white has not been used as a feature, as the ratio can be zero or unidentiﬁed at certain blocks with no black background.

The output image after implementing Canny edge detection is used as the input for this particular feature. The number of on pixels along the edge

**Results**

After implementing the proposed operations and classiﬁcation algorithm, satisfactory result has been achieved. For comparing the usability of every feature, different combinations of the features have been examined as well. After the application of Cubic SVM, Quadric SVM, Cubic KNN, and LDA to classify and examine the result, Cubic SVM is chosen to be appropriate for the system. Furthermore, the relevance of the features is also examined by using multiple combinations of them. It is discovered that all of the features combined give the most accuracy. Hence, a combination of all the features extracted has been used to reach the ﬁnal decision. Test accuracy is achieved after randomly training 80% data as a training dataset and testing 20% data of the remaining dataset. Our model recognizes 9 out of the 10 alphabets without any errors. The confusion matrix shows that after 5 fold validation, the system recognizes all the alphabets of the existing dataset perfectly except .The 0.2% inaccuracy occurs due to the alphabet. Overall, we achieved a test accuracy of 99.8%

**Discussion**

The original system uses ideal samples as both train and test subjects. To examine the cogency of the algorithm, further operation is conducted on the ideal dataset. At this point, the contrast of the sample images is modiﬁed to acquire slightly lower quality. These low contrast samples are then used as the test dataset while training the system with the ideal one. Here the accuracy achieved by the system drops to 86%. To examine the rotational invariance, the samples are rotated 5◦ anti-clockwise and then used as the test dataset. In this case, the accuracy does not vary much and the ﬁnal accuracy achieved is 98%. With both the rotation and low contrast effect on action, the accuracy drops to 92%.Therefore, the algorithm may be rotation invariant to quite some extent, but to achieve a higher accuracy the dataset must be curated with good quality images

**Conclusion**

The proposed system provides an accurate result in recognizing hand poses for 10 alphabets of Bangla Sign Language (BdSL) using multiclass SVM. The system has an ideal dataset and has explored different color spaces and utilizing YCbCr space, and converted binary image features extraction. Most importantly, it has the most suitable classiﬁer detection. About 18 distinguished features have been extracted and a combination of them has been used to deliver the ultimate solution. The individual features are concrete and thus gives a quite high accuracy even when using separately. Different types of classiﬁers have been explored as well. Almost all of the classiﬁers give a pretty similar result which proves the features to be deﬁnitive. The system presented in this paper is not only exclusive for BdSL translation. If the dataset is available, Sign Language from other regions can also be trained and recognized using this system.

**Reference**

[(PDF) Bangla Sign Language Recognition with Skin Segmentation and Binary Masking (researchgate.net)](https://www.researchgate.net/publication/339652786_Bangla_Sign_Language_Recognition_with_Skin_Segmentation_and_Binary_Masking)