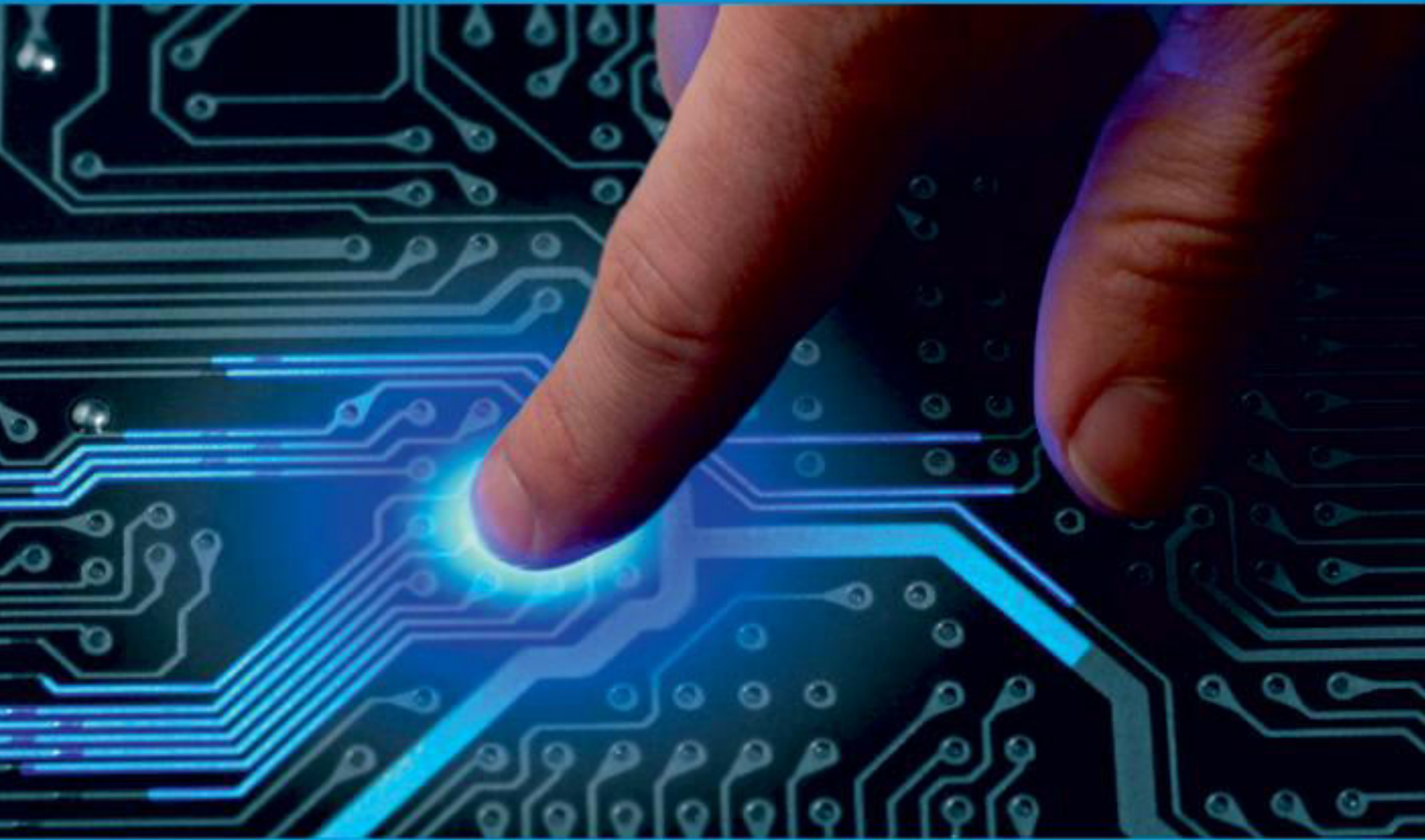
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

9940 572 462 6381 907 438 ijircce@gmail.com www.ijircce.com

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907212 |

Fingerspelling-A Sign Language Learning Tool

Mayank Negi[1],Vinay Madhusudhan[2],Santhosh Kumar[3],Deepak[4],Mr Channappa Gowda Dv[5]

B.Tech Final Year, Dept. of ISE, The Oxford College of Engineering, Bengaluru, India[1,2,3,4]

Asst. Professor, Dept. of ISE, The Oxford College of Engineering, Bengaluru, India[5]

ABSTRACT: One of the major drawbacks of our society is the barrier that is created between disabled or handicapped persons and the normal person and communication is the only medium by which we can convey the message. For many deaf and dumb people, sign language is the basic means of communication. Our Project- aims to interpret sign languages automatically by a GUI in order to help the deaf communicate with our society conveniently. It is a gesture based animation system for teaching and learning the Indian Sign Language(ISL). We are using image processing technology and Machine learning techniques to interpret the gestures and bring out the desired output.

KEYWORDS: A Sign Language Learning Tool ,Human Computer Interaction,Hand Gesture, Text or speech ,Python ,K-Neighbour's Algorithm (KNN),Mission Learning (ML).

I. INTRODUCTION

Automated recognition of hand signals has many applications in computer science. It might facilitate the interaction between humans and computers in many situations, especially for people with disabilities. One interesting area of focus is the recognition of sign languages. This way of communication is widely used among Deaf communities.

II. RELATED WORK

Hand pose recognition has already been the subject of much work. We divide this work in three sections: work on hand pose recognition using only regular intensity images, work using only depth images, and work using a mixed intensity-depth system.
a) Intensity Input Only: Ong and Bowden (2004) introduced a tree structure of boosted cascades that can recognize a hand and classify it in different shapes [10]. They achieved an average error of 2.6 % on their test set. However, their method is time-consuming and their classifier is built on automatically chosen shapes. Instead of classifying images of signs which are already categorized in a number of different categories, they use an algorithm to automatically cluster their training set in 300 clusters, each cluster containing similar images.

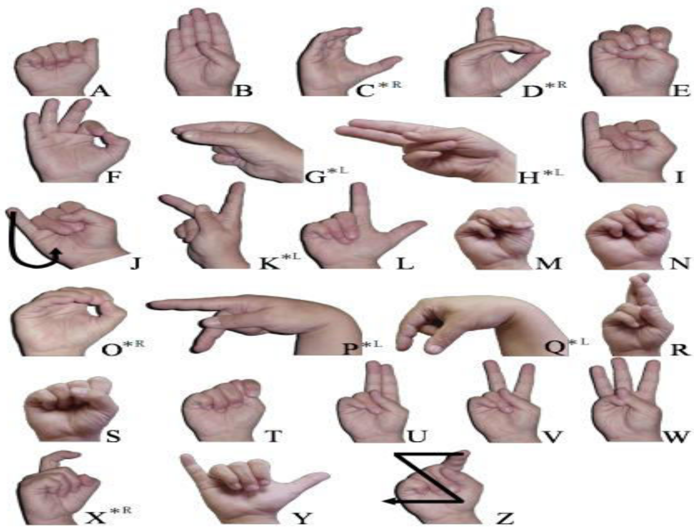


Fig. 1: The American Manual Alphabet. Signs with a * are pictured from the side. Reproduced from [15]

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542
|| Volume 9, Issue 7, July 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0907212 |

Fig. 2: These six gestures only vary by the position of the thumb, making them the most frequently confused.

Fig. 3: Illustration of the difference between the five different users for the same sign (Q).

III. HAND POSE CLASSIFICATION

In order to classify the different hand poses, we use two different input methods: regular intensity images and depth images. Depth images are matrices of the depth of the environment, where the value of each pixel corresponds to the depth of that pixel relative to the camera. Thus, using both intensity and depth, an accurate 3D perspective of an object can be created. Using depth data is becoming more common due to the increased number of 3D sensors available in the market.

For our experiment, we used a Microsoft Kinect sensor. This device has a 640x480 image resolution for both intensity and depth images, with a depth range between two and eight meters.In this article, we will only focus on the task of classifying the hand pose using hand images. Those images were obtained by tracking the hands using readily available functions inthe Open NI+NITE framework [12]. They have a size of about 100x100 pixels and are framed around the hand, with the background still present. From these images, we extract features that are subsequently fed to a Deep Belief Network(DBN), which classifies the hand pose as one of 24 possible letters.

A. Feature Extraction

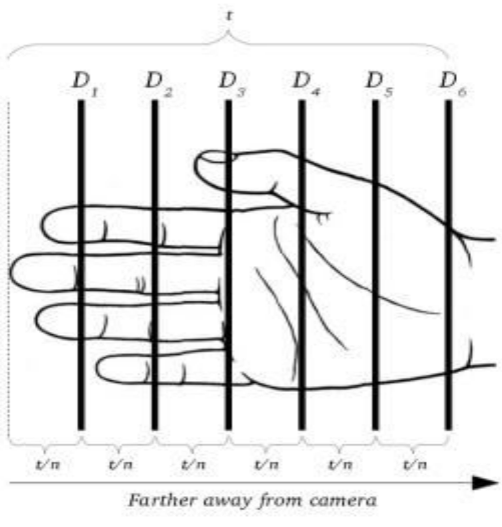


Fig. 4: Drawing of the layers used to decompose the hand. Each binary image D_l represents the parts of the hand to the left of the layer.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542
|| Volume 9, Issue 7, July 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0907212 |

$$D_l(i,l) = \begin{cases} 1 & \text{if } D(i) \leq ((l-1) \times \frac{t}{n}) + 1 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Finally, each layer is centered using the bounding box technique. Fig. 5 illustrates this process. In our experiments, we found that using $n = 6$ binary images and a maximum hand depth of $t = 12$ cm gave the best features, so that the layers Dl are separated by 2 cm. This distance is consistent with the approximate thickness ofa finger, and being able to differentiate finger position is keyin distinguishing one sign from another.

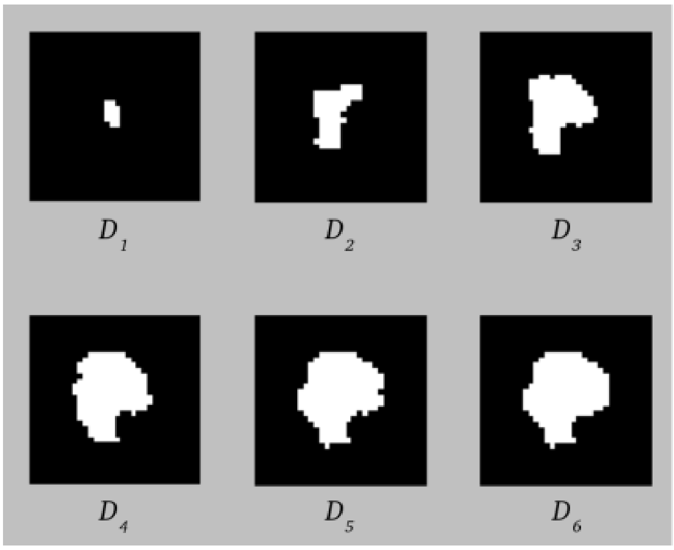


Fig. 5: Successive binary depth images for a hand doing the "G" sign. D_1 is the tip of the two forward-pointing fingers, D_2 and D_3 add more fingers, D_4 includes the hand except for a part of the palm, and more parts of the palm can be seen on D_5 and D_6. Here, we used $n = 6$ layers.

Also, the fingers are always located on the front part of the hand, well behind our12 cm margin. To generate the depth feature vector *fdepth*, each of the *n*binary depth images is first resized to 32x32, unrolled into a1x1024 vector and all of them are concatenated in a 1x6144vector. Finally, the intensity image features *fintensity* and depth image features *fdepth* are concatenated together into afeature vector *fcombined* of size 1x10240.

IV. EXPERIMENTS

Before experimenting with our new feature extraction technique, we looked at other ways to extract features from the depth and intensity images. As a baseline experiment, we trained a DBN on the raw intensity and depth images. For this experiment, the images were pre-processed as explained in Sec. III-A. We kept the original size of the pictures (128x128),then unrolled them into two1x16384 vectors with which the DBN was trained. A second experiment was done using Gabor filters on the processed images, using four different scales and orientations. The parameters were fine-tuned by hand in order to extract the key points of the images that would be useful in the classification step.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542
|| Volume 9, Issue 7, July 2021 ||
| DOI: 10.15680/IJIRCCE.2021.0907212 |

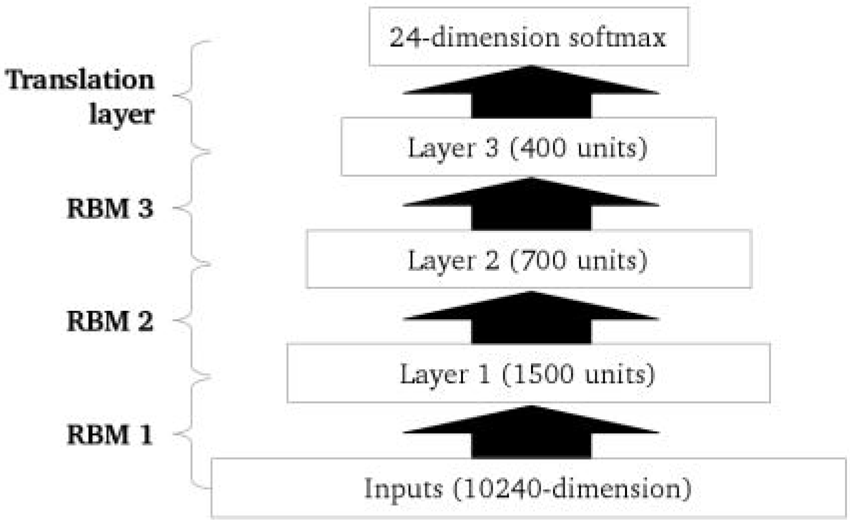


Fig. 6: Representation of the DBN we used. The feature vector $f_{combined}$ is the input, and the output indicates which of the 24 static letters has been identified.

V. RESULTS

We evaluated our results with precision and recall measures, computed for each letter. We compared both testing approaches with the results of Pugeault & Bowden [12] as theyuse the same data set. A detailed comparison of our results is presented in Fig. 8 for the precision measure and

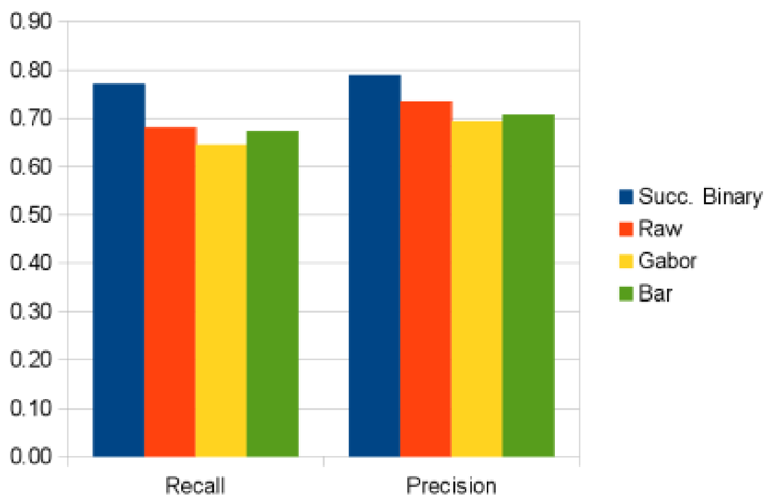


Fig. 7: Comparison of the recall and precision for the different types of features used in the T_{unseen} method.

We first evaluated our three initial attempts at feature extraction using the *Tunseen* method. For the raw intensity and depth vectors, we obtained 68 % recall and 73 % precision. Forthe Gabor filters, we obtained 64 % recall and 69 % precision, and for the bar filters, 67 % recall and 71 % precision.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907212 |

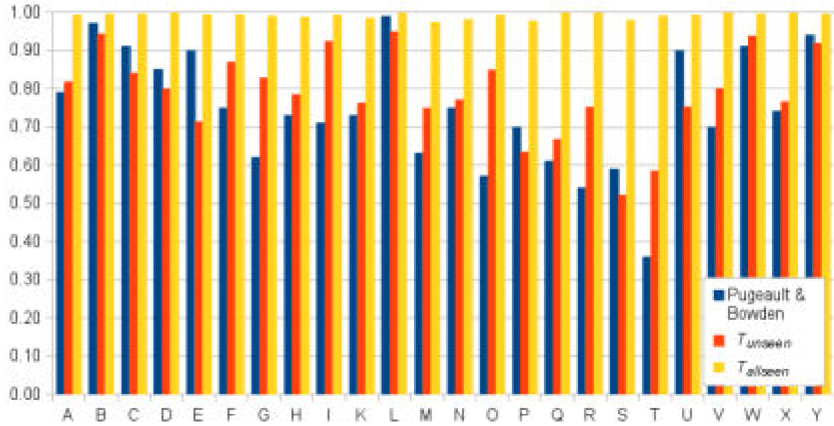


Fig. 8: Comparison of the precision for all letters between our two testing methods and [12]. Pugeault & Bowden use $T_{allseen}$.

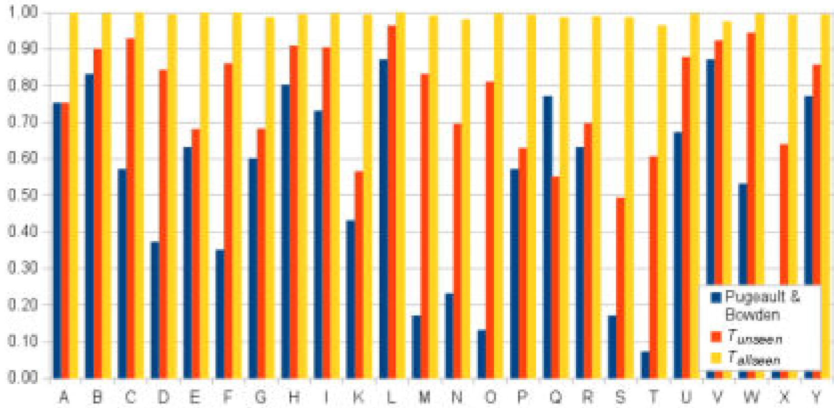


Fig. 9: Comparison of the recall for all letters between our two testing methods and [12]. Pugeault & Bowden use $T_{allseen}$.

However, our system was able to distinguish with good accuracy the letters M and N (83 % and 69 % recall respectively), with little confusion (under 10 % between them).Those signs are very hard to distinguish from each other using only the intensity data (see Fig. 2), but this task becomes easier using the depth data. This shows that the layer-based feature extraction technique provides non-negligible information for classifying hand signs in 3D.

VI. CONCLUSION AND FUTURE WORK

We have presented a novel way to extract features from depth images along with a way to segment the hand on intensity images using depth images. We have combined this technique with a Deep Belief Network for the task of ASL fingers pelling classification. We demonstrated the value of these new features by comparing them against more classical features (raw images, Gabor and bar-like). We obtained excellent results on two different scenarios, including one with anew unseen user. However, some restrictions remain. Our system is only capable of classifying static gestures. However, ASL finger spellingis not only static, but also includes two dynamic gestures.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907212 |

REFERENCES

[1] O. Aran. *Vision based sign language recognition: modeling andrecognizing isolated signs with manual and non-manual components*.PhD thesis, Bogazic¸i University, 2008.
[2] I. B. V. da Silva and P. J. Adeodato. Pca and gaussian noise inmlp neural network training improve generalization in problems withsmall and unbalanced data sets. In *Neural networks (IJCNN), the 2011international joint conference on*, pages 2664–2669. IEEE, 2011.
[3] P. Doliotis, A. Stefan, C. McMurrough, D. Eckhard, and V. Athitsos. 1Comparing gesture recognition accuracy using color and depth information. In *Proceedings of the 4th International Conference on Pervasive Technologies Related to Assistive Environments*,

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details