IJCRT.ORG ISSN : 2320-2882
INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

SIGN LANGUAGE IDENTIFICATION USING IMAGE PROCESSING

[1]Mrs.P.J.Mercy MCA., M.Phil., [2]S.Amsa Vani
[1]Associate Professor, [2]III MCA Student
Department of Computer Applications and Research Centre,
Sarah Tucker College (Autonomous), Tirunelveli-7

ABSTRACT: **The objective of this research is to develop a real-time system for hand gesture recognition that recognizes hand gestures, features of hands such as peak calculation and angle calculation, and then convert gesture images into text. To implement this system we use a sign language hand gesture dataset..The proposed system has four main modules such as preprocessing, segmentation, feature extraction, and classification. In the preprocessing module, the input image is converted into a grayscale image, noise removal, normalization, and rescale the image. After preprocessing, the image is segmented using an automatic threshold algorithm. Then the feature extraction is calculated using SIFT and SURF descriptors. The collected features are trained using an excel file. Finally, the input is classified using the SVM algorithm.**

1. INTRODUCTION

Communication between deaf-mute people and normal people is more difficult because normal people cannot perceive the speculation and feeling of deaf-mute people. So knowing the various types of sign languages such as sign language for better communication between those people.

There are two ways of approaches for gesture recognition. In the non-vision-based approach, sensors such as flex sensors, pressure sensors are used for sign recognition which cannot require any proper lighting. In a vision-based approach, a Real-Time gesture image made by deaf-mute people is taken as input for sign recognition and this requires proper lighting for accurate results.

Sign language is used by the speech impaired. There are more than 1 million people in the U.S. who are functionally deaf and another 10 million who are hard of hearing, according to a 2005 study. According to the Ministry of National Education number of the hearing impaired in Turkey is 400,000. The main problem that works aims to solve is the communication problem between speech-impaired people and others. As those people cannot express themselves with words, they have many difficulties during their daily life. Some of the existing methods used by them are that they can either have an interpreter that can convey their messages in a word they even use alphabetic charts or short handwritten notes that will convey their message. They prefer writing to be their mode of communication then signs as written notes are more efficient to understand.

People have even developed a speech-to-text converter for the students who are speech and hearing impaired thus helping them understand the speech of the other individual. The work aims to bridge the barrier by building an application that can convert sign language to voice and provide them a medium to communicate. The system fingerspells the letters of the basic English alphabets from A-Z which are in sign language to speech. It will also implement signs that can be converted to speech. Thus the application will solve the problem of communication by letting the speech-impaired people make sing in front of a webcam and thus produces a resultant voice output. Sign recognition is a typical applies of image understanding. It includes two phases: sign detection and sign recognition. Sign detection is to try to find the position of a certain object in a sequence of images. Sign recognition is to try to recognize a certain pattern that differentiates the object from the world.

2. LITERATURE REVIEW

There is much research that has begun carried out in this research area. Oi Mean Foong et al have discussed sign language translation system using Speech and Image processing technique in "V2S: Voice to Sign Language Translation System for Malaysian Deaf People"[1]. The Advantage is Voice (English Language) to sign language of Malaysia. The main disadvantage is System first needs to be trained with speech patterns based on some generic spectral parameter set.

Jonathan Gatti et al have authored "Voice-Controlled Artificial Handspeak System" [2]. A methodology was Prototype. Advantages of this paper robotic hand-designed with openSCAD and manufactured with a low-cost 3D printer used, core automation comprises an Arduino UNO controller by a Raspberry Pi computer and uses open-source speech recognition engine, Julius. The only main disadvantage of this research was the robotic hand has its limitations and possible future developments.

Tirthankar Dasgupta et al have proposed an application that can be used as an educational tool to learn ISL[3]. This research has used a prototype methodology. The system was not only improved information access, but it can also get as main advantage and disadvantages of this research. It was not translated Indian sign language into the text, the system takes only simple English sentences as input, sign synthesis module using an animated avatar has not been developed, and some grammar rules cannot be applied to translate English to ISL, not given clear idea about how the system works properly.

"Sign Speak: American Sign Language Translation through Sensory Glove." authored by JanFizza Bukhara, Maryam Rahman, Samna Ishtar Malik, Awaits M. Kamboh, Ahmad Salman [4]. A methodology was to design a glove that would enable deaf and mute people to communicate by translating their sign language gestures into speech according to American Sign Language, different modules were: glove design, data acquisition system, feature extraction, feature matching, wireless link, and android application. There were main two advantages has this research. Those were focused on the translation of gestures of the alphabets and also the words, gestures were classified very efficiently and accurately. And also there were many disadvantages. Those were required the user to wear clothes with full sleeves to cover arms. Plus, lighting effects could adversely affect this method, no communication happens between two people, more hardware used (flex sensors, Accelerometer, Contact sensors), only supported Android OS.

Dalia Nashat et al have discussed an android application in "An Android Application to Aid Uneducated Deaf-Dumb People"[5]. The main advantages were, support uneducated DeafDumb people who could not read and write Arabic languages to communicate with others, to learn and to entertain, represent quizzes and games for training deaf and dumb people/kids to identify Arabic and English words, introduce Sign language keyboard. The main disadvantage was, only Support Android OS, no face to face communication happens, an educational tool rather than a Real-world tool, use one sign for the one alphabet.

Dr. Sami M.Halawani et al have authored "Arabic Sign Language Translation System on Mobile Device" authored by [6]. Advantages were 3D Animated characters gave more attraction & realism to the system users, the application could work online & offline. There were more disadvantages. The system had not translated sign language into text, no option of face to face communication, need more database space (Arabic Sign, applications for translating text to sign animation, other external systems, or library).

Noor Saliza Mohd Salleh et al have proposed "Sign Language to Voice Recognition: Hand Detection Techniques for Vision-Based Approach" [7]. Advantages were, more flexible and useful than prior approach, fast processing. The main disadvantage was video analysis problems.

"Sign Language to Speech Translation System Using PIC Microcontroller" authored by Gunasekaran and Manikandan. R [8]. Advantages were, the system offered high reliability and fast response, more précised on hand movement, different languages could be installed without altering the code and the main disadvantage was high manufacture cost

Sachin Bhat et al have proposed an application on "Translating Indian Sign Language to text and voice messages using flex sensors [9]. The main advantages were, user-independent, portable system to convert the sign language to text message form which consumes less power because of the low ultra-power AT89S52 microcontroller was designed, used the simple mobile application and also main disadvantage was higher cost.

A.Sujith Kumar et al have discussed mobile chat applications on "Sign Mobiles (An android app for especially able People)" [10]. Advantages are enabled sign language fingerspelling communication, briefly allowed to use mobile communication with face to face chatting, automatic translation, and speech recognition. The main disadvantages were large size database was needed to store video clips, a mismatch of voice with video clips cannot be handle, it took time to compare a video clip and voice rather than matching text with signs.

3. METHODOLOGY

Nowadays many technology devices such as Smartphone-enabled devices prefer speech interfaces over visual ones. The research highlighted that an off-the-shelf speech recognition system cannot be used to detect the speech of deaf or hearing loss people as these systems contain a higher ratio of word error rate. This research recommended using human-based computations to recognize deaf speech and using text-to-speech functionality for speech generation. In this regard, we proposed and developed an Android-based application named vocalize to mute (V2M). The proposed application acts as an interpreter and encourages two-way communication between Deaf-mute and normal people. We refer to a normal person as the one who has no hearing or vocal impairment or disability. The main features of the proposed application are listed below.

- **IMAGE ACQUISITION**
- **PREPROCESSING**
  - **GRAYSCALE**
  - **NOISE FILTER**
  - **BACKGROUND SUBTRACTION**
  - **NORMALIZATION**
  - **SCALE IMAGE**
- **SEGMENTATION**
- **FEATURE EXTRACTION**
  - **SIFT**
  - **SURF**
- **CLASSIFICATION**
  - **SVM**

❖ **IMAGE ACQUISITION:**

The first stage of any vision system is the image acquisition stage. After the certificate image has been obtained, various methods of processing can be applied to the image to perform the many different vision tasks required today. However, if the image has not been acquired satisfactorily then the intended tasks may not be achievable, even with the aid of some form of image enhancement.

❖ **PREPROCESSING**

- **GRAYSCALE**

A grayscale image is a digital image is an image in which the value of each pixel is a single sample, that is, it carries only intensity information. Images of this sort, also known as black-and-white, are composed exclusively of shades of gray (0-255), varying from black (0) at the weakest intensity to white (255) at the strongest.

Grayscale images are distinct from one-bit black-and-white images, which in the context of computer imaging are images with only two colors, black, and white (also called bi-level or binary images). Grayscale images have many shades of gray in between. Grayscale images are also called monochromatic, denoting the absence of any chromatic variation.

- **NOISE FILTER**

Noise reduction is the process of removing noise from a signal. Noise reduction techniques exist for audio and images. Noise reduction algorithms tend to alter signals to a greater or lesser degree. All signal processing devices, both analog and digital, have traits that make them susceptible to noise.

- **BACKGROUND SUBTRACTION**

Background subtraction is a technique for separating foreground elements from the background and is done by generating a foreground mask. This technique is used for detecting dynamically moving objects from static cameras. Background subtraction technique is important for object tracking.

.

- **NORMALIZATION**

In image processing, normalization is a process that changes the range of pixel intensity values. Applications include photographs with poor contrast due to glare, for example. Normalization is sometimes called contrast stretching or histogram stretching.

- **SCALE IMAGE**

In computer graphics and digital imaging, image scaling refers to the resizing of a digital image. ... When scaling a raster graphics image, a new image with a higher or lower number of pixels must be generated. In the case of decreasing the pixel number (scaling down), this usually results in a visible quality loss.

- **SEGMENTATION**

Segmentation is the technique of dividing or partitioning an image into parts, called segments. It is mostly useful for applications like image compression or object recognition because for these types of applications, it is inefficient to process the whole image.

❖ **THRESHOLD METHOD**

The threshold method proves to be advantageous when the objects in the image in question are assumed to be having more intensity than the background (and unwanted components) of the image.

- **Otsu's Binarization**

According to Otsu binarization, for that image, we can approximately take a value in the middle of those peaks as the threshold value. So in simply put, it automatically calculates a threshold value from an image histogram for a bimodal image. Otsu's Binarization is widely used in document scans, removing unwanted colors from a document, pattern recognition, etc.

❖ **FEATURE EXTRACTION**

- **SIFT**

There are mainly four steps involved in SIFT algorithm. We will see them one-by-one.

1. Scale-space Extrema Detection

We can't use the same window to detect keypoints with a different scale. It is OK with a small corner. But to detect larger corners we need larger windows. For this, scale-space filtering is used. In it, Laplacian of Gaussian is found for the image with various σ values. Log acts as a blob detector that detects blobs in various sizes due to change in σ. In short, σ acts as a scaling parameter. For eg, in the above image, Gaussian kernel with low σ gives high value for a small corner while Gaussian kernel with high σ fits well for a larger corner. So, we can find the local maxima across the scale and space which gives us a list of (x, y, σ) values which means there is a potential keypoint at (x,y) at σ scale.

But this LoG is a little costly, so SIFT algorithm uses the Difference of Gaussians which is an approximation of LoG. A difference of Gaussian is obtained as the difference of Gaussian blurring of an image with two different σ, let it be σ and $k\sigma$. This process is done for different octaves of the image in the Gaussian Pyramid. It is represented in Figure 1:

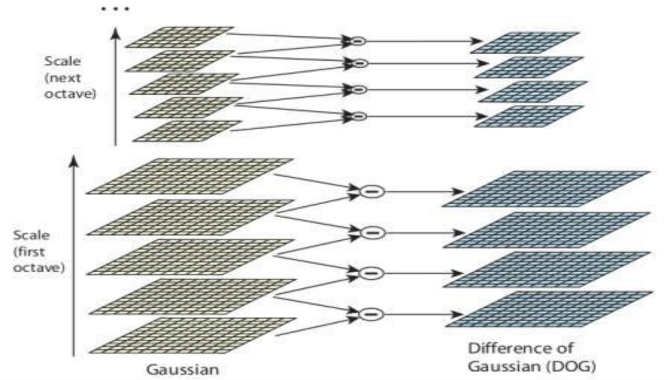


Figure 1:Gaussian pyramid

Once this DoG is found, images are searched for local extrema over scale and space. For eg, one pixel in an image is compared with its 8 neighbors as well as 9 pixels in the next scale and 9 pixels in previous scales. If it is a local extremum, it is a potential keypoint. It means that keypoint is best represented in that scale. It is shown in Figure 2:

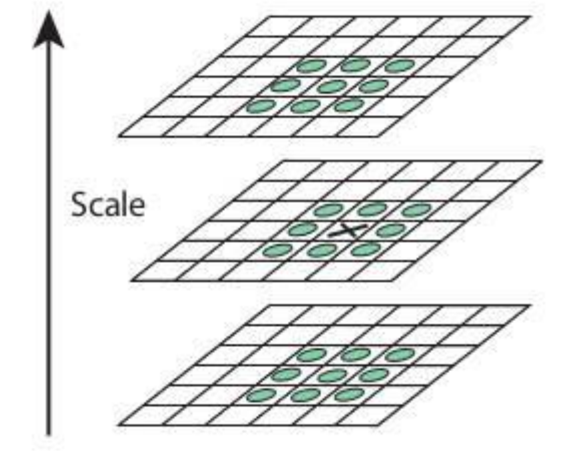


Figure 2:Keypoint Detection in scale-space

Regarding different parameters, the paper gives some empirical data which can be summarized as, number of octaves = 4, number of scale levels = 5, initial $\sigma = 1.6$, $k = \sqrt{2}$ etc as optimal values.

2. Keypoint Localization

Once potential keypoints locations are found, they have to be refined to get more accurate results. They used Taylor series expansion of scale-space to get a more accurate location of extrema, and if the intensity at this extrema is less than a threshold value (0.03 as per the paper), it is rejected. This threshold is called the **contrast Threshold** in OpenCV.

A log has a higher response for edges, so edges also need to be removed. For this, a concept similar to the Harris corner detector is used. They used a 2x2 Hessian matrix (H) to compute the principal curvature. We know from the Harris corner detector that for edges, one eigenvalue is larger than the other. So here they used a simple function,

If this ratio is greater than a threshold, called **e**dge Threshold in OpenCV, that key point is discarded. It is given as 10 in the paper. So it eliminates any low-contrast key points and edges key points and what remains is strong interest points.

3. Orientation Assignment

Now an orientation is assigned to each key point to achieve invariance to image rotation. A neighborhood is taken around the keypoint location depending on the scale, and the gradient magnitude and direction are calculated in that region. An orientation histogram with 36 bins covering 360 degrees is created. (It is weighted by gradient magnitude and Gaussian-weighted circular window with σ equal to 1.5 times the scale of the key point. The highest peak in the histogram is taken and any peak above 80% of it is also considered to calculate the orientation. It creates key points with the same location and scale, but different directions. It contributes to the stability of matching.

4. Keypoint Descriptor

Now keypoint descriptor is created. A 16x16 neighborhood around the key point is taken. It is divided into 16 sub-blocks of 4x4 size. For each sub-block, 8 bin orientation histogram is created. So a total of 128 bin values are available. It is represented as a vector to form a keypoint descriptor. In addition to this, several measures are taken to achieve robustness against illumination changes, rotation, etc.

- **SURF ALGORITHM**

SURF (**S**peeded-**U**p **R**obust **F**eatures) is a feature detection framework introduced by Herbert Bay and his colleagues at ETH Zurich. SURF interest points are in-plane rotation-invariant, robust to noise, and overall, extremely fast to calculate.

The procedure can be divided into three steps:

1. Interest Point Detection
2. Interest Point Description
3. Interest Point Matching

The algorithm works as follow:

1. Find features/key points that are likely to be found in different images of the same object. Those features should be scale and rotation invariant if possible. Corners, blobs, etc are good and most often searched in multiple scales.
2. Find the right "orientation" of that point so that if the image is rotated according to that orientation, both images are aligned regarding that single key point.

3. Computation of a descriptor that has information of how the neighborhood of the key point looks like (after orientation) in the right scale.

SURF is advertised to perform faster compared to previously proposed schemes like SIFT. This is achieved (as stated by its designers) by:

- Relying on integral images for image convolutions.
- Building on the strengths of the leading existing detectors and descriptors (using a Hessian matrix-based measure for the detector, and a distribution-based descriptor).
- Simplifying these methods to the essential
- This leads to a combination of novel detection, description, and matching steps.

SURF:

The SURF feature detector works by applying an approximate Gaussian second derivative mask to an image at many scales. Because the feature detector applies masks along each axis and at 45 deg to the axis, it is more robust to rotation than the Harris corner. The method is very fast because of the use of an integral image where the value of a pixel (x,y) is the sum of all values in the rectangle defined by the origin and (x,y). In such an image the sum of the pixels within a rectangle of any size in a source image can be found as the result of 4 operations. This allows a rectangular mask of any size to be applied with very little computing time.

To detect features we assemble the Hessian matrix $H = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{bmatrix}$ where L_{xx} is the convolution of the second derivative of a Gaussian with the image at the point. The masks used are a very crude approximation and are shown in Figure 3. The crude approximations are valuable because they can be very quickly run at any scale due to the use of an integral image.

The Hessian determinate values for the same image as Figure 3 are shown in Figure 4 for the range of detector windows that were used in this work. Valid features are found as a local maximum over a 3x3x3 range where the third dimension is detector window size, so a feature must be locally unique over a spatial range and a range of scales. The SURF authors used a fast search algorithm to do non-maximum suppression, we have not implemented this yet.

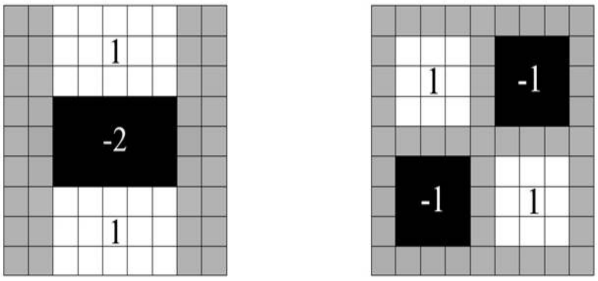


Figure 3:Approximated Gaussian second derivative used for the SURF detector. This detector works well with integral images because only the sum over a rectangle is needed. On the left is the mask used for G_{XX} and G_{YY} and on the right is the mask used for G_{XY}. Taken from the original SURF paper.

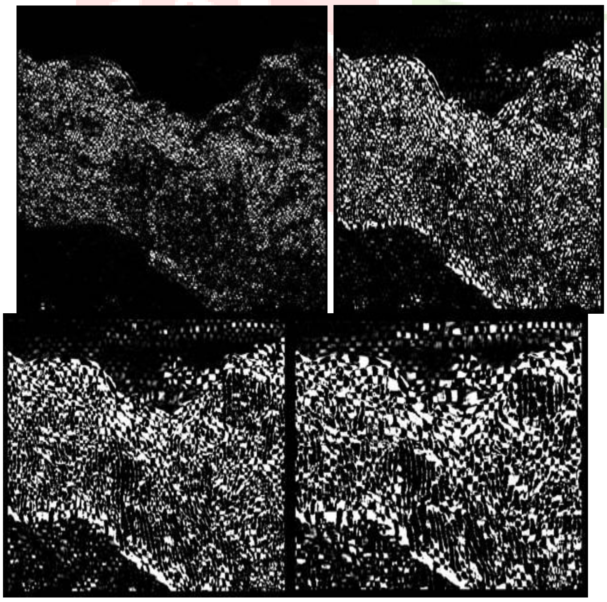


Figure 4: Surf feature values at 4 different detector sizes. Top Left = 9x9, top right = 15x15, bottom left = 21x21, bottom right = 27x27

Feature Descriptors

Having found features in the previous section we must now find some aspects of those features that can be compared among the features. This is the descriptor.

Basic 5x5

We were assigned to implement a basic window descriptor which is the values of the 5x5 window of grayscale values centered on the feature we are describing. This description is invariant to planar motion but fails if there are changes in lighting or rotation.

Surf Descriptor

The SURF descriptor is designed to be scale-invariant and rotationally invariant. To ignore scale the descriptor is sampled over a window that is proportional to the window size with which it was detected, that way if a scaled version of the feature is in another image the descriptor for that feature will be sampled over the same relative area.

Rotation is handled by finding the dominant direction of the feature and rotating the sampling window to align with that angle. Once the rotated neighborhood is obtained it is divided up into 16 subs (A_i) squares, each sub-square is again divided into 4 squares. Derivatives in the x and y directions are taken in these final squares. The descriptor for the sub-square (A_i) is the sum of the x derivatives over its four quadrants, the sum of the absolute values of the x derivatives, and similarly for y. The total descriptor is 4 values per (A_i) for a total of 64 values. This vector is normalized to length 1 and is the feature descriptor. The process is summarized in Figure 5.

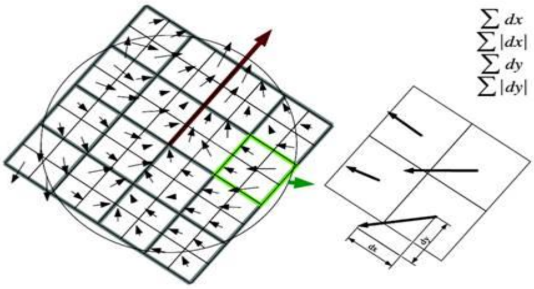


Figure 5: A graphical representation of the SURF descriptor.

There are a few aspects of the SURF detector and descriptor that we have not yet had time to implement. First creating the features is a very slow process, no optimization has been done yet on this step. The SURF authors include a Gaussian weighting over the descriptor and a more efficient matching system we have not implemented these yet. Currently, our best performance is with the sum-squared-distance matching system but it ignores a lot of information that is available. An additional problem is shown in Figure 6, we haven't created a good way of dealing with borders yet so we avoid them. This leads to ignoring potential features near the edges of the image.

❖ **CLASSIFICATION**

- **SVM**

SVM stands for **Support Vector Machine**. It is a **machine learning** approach used for classification and regression analysis. It depends on **supervised learning models** and is trained by learning algorithms. They analyze the large amount of data to identify patterns from them. An SVM generates parallel partitions by generating two parallel lines. Each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible. Do this partitioning by a plane called *a* hyper plane.

An SVM creates hyperplanes that have the largest margin in high-dimensional space to separate given data into classes. The margin between the 2 classes represents the longest distance between the closest data points of those classes.

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other, making it a no probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. The main advantage of the SVM network used as a classifier is its very good generalization ability and extremely powerful learning procedure, leading to the global minimum of the defined error function.

Inputs:

The complete MRI database consists of 30 patients: 10 Normal, 10 MCI, and 10 AD. In this project SVM using the nonlinear algorithm for classification.

Applications:

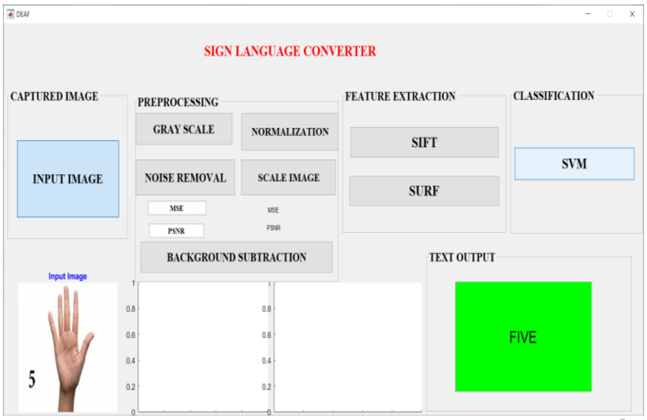
SVMs can be used to solve various real-world problems:

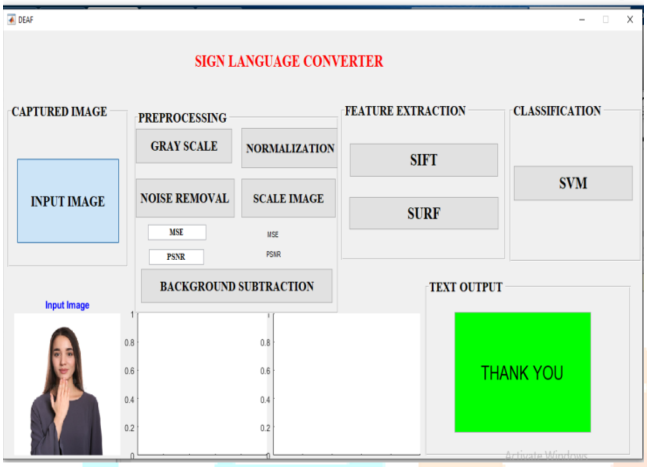
❖ Classification of images can also be performed using SVMs.
❖ SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly.
❖ Hand-written characters can be recognized using SVM.

Advantages:

❖ Effective in high dimensional spaces.
❖ Still effective in cases where several dimensions are greater than the number of samples.
❖ Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
❖ Versatile: different Kernel functions can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels

4. RESULT





4. CONCLUSION

. The main inspiration behind the proposed application is to remove the communication barrier for Deaf-mutes especially children. This app uses the hand gesture images of deaf-mute persons and translates them into sign language as the text input. It provides a speech recognition system for the distorted speech of Deaf-mutes. The speech recognition system uses SIFT and SURF feature extraction techniques to extract the image input from speech samples. The application can recognize Deaf-mute speech samples of English digits (0 to 9) and 15 common sentences used in daily routine life, i.e., good morning, hello, good luck, thank you, etc. It provides message service for both Deaf-mutes and normal people. Deaf-mutes can use a customized sign language keyboard for composing the message. The app also can convert the received sign language message to text for a normal person. The proposed application was also tested on 15 children aged between 7 and 13 years. The accuracy of the proposed application is 97.9%. The qualitative feedback of children also highlighted that it is easy for Deaf-mutes to adapt the mobile technology and mobile app can be used to convey their message to a normal person.

5. REFERENCE

1. W. Tin, Z. Lin, -. Swe, and N. K. Mya, "Deaf-mute or Deaf," *Asian Journal of Medical and Biological Research*, vol. 3, no. 1, pp. 10–19, 2017.View at: Publisher Site | Google Scholar
2. W. Leigh and J. F. Andrews, *Deaf People and Society: Psychological, Sociological and Educational Perspectives*, Psychology Press, 2016.
3. WHO, "Deafness and Hearing Loss," *Fact Sheet*, vol. 300, 2017.View at: Google Scholar
4. J. G. Clark, "Uses and Abuses of hearing loss classification," *Asha*, vol. 23, no. 7, pp. 493–500, 1981.View at: Google Scholar
5. B. H. Jacobson, A. Johnson, C. Grywalski, et al., "The Voice Handicap Index (VHI): development and validation," *American Journal of Speech-Language Pathology*, vol. 6, no. 3, pp. 66–70, 1997.View at: Google Scholar
6. P. Vostanis, M. Hayes, M. Du Feu, and J. Warren, "Detection of behavioral and emotional problems in deaf children and adolescents: Comparison of two rating scales," *Child: Care, Health and Development*, vol. 23, no. 3, pp. 233–246, 1997.View at: Publisher Site | Google Scholar
7. F. Soltani, F. Eskandari, and S. Golestan, "Developing a gesture-based game for the deaf/mute people Using Microsoft Kinect," in *Proceedings of the 2012 6th International Conference on Complex, Intelligent, and Software Intensive Systems, CISIS 2012*, pp. 491–495, July 2012.View at Publisher Site | Google Scholar
8. D. G. Blazer, S.Domnitz, C. T. Liverman, C. on Accessible, A. H. H. C. for Adults, E. National Academies of Sciences, and Medicine, Hearing Loss: Extent, Impact, and Research Needs. 2016.
9. WHO, "Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report, 2017, in Global costs of unaddressed hearing loss and cost-effectiveness of interventions: a WHO report," Tech. Rep., 2017.View at: Google Scholar
10. T. Humphries, "Of Deaf-mutes, the Strange," in *Deaf World: A Historical Reader and Primary Sourcebook*, 2001.