

# Sign Language Interpreter

Divyashree Shinde<sup>1</sup>, Gaurav Dharra<sup>1</sup>, Parmeet Singh Bathija<sup>1</sup>, Rohan Patil<sup>1</sup>, M. Mani Roja<sup>2</sup>

1. Students, Department of EXTC, ThadomalShahani Engineering College, Mumbai, India

2. Associate Professor, Department of EXTC, ThadomalShahani Engineering College, Mumbai, India

**Abstract** –The sign language interpreter is an interactive system that will recognize different hand gestures of Indian Sign Language and will give the interpretation of the recognized gestures in the form of text messages along with audio interpretation. In a country like India there is a need of automatic sign language recognition system, which can cater the need of hearing and speech impaired people thereby enhancing their communication capabilities, promising improved social opportunities and making them independent. Unfortunately, not much research work on Indian Sign Language recognition is reported till date. The proposed Sign Language Interpreter using Discrete Cosine Transform accepts hand gesture frames as input from image capturing devices such as webcam. This input image is converted to grayscale and resized to 256 x 256 pixels. DCT is then applied to entire image to obtain the DCT coefficients. Recognition is carried out by comparison of DCT coefficient of input image with those of the registered reference images in database. Database image having minimum Euclidean distance with the input image is recognized as the matched image and the sound corresponding to the matched sign is played. Accuracy for different levels of DCT coefficients is obtained and hence, compared. The scope of this paper is extended to recognize words and sentences by taking video as input, extracting frames from the input video and comparing with the stored database images.

**Keywords**–Discrete cosine transform, Euclidean distance, feature extraction, hand gesture recognition.

## I. INTRODUCTION

Sign languages are well structured languages with phonology, morphology, syntax and grammar distinctive from spoken languages [1]. The structure of a spoken language makes use of words linearly, i.e., one after the other, whereas a sign language makes use of several body movements simultaneously in the spatial as well as in temporal space. Sign languages are natural languages that use different means of expression for communication in everyday life. Examples of some sign languages are American Sign Language, British Sign Language, the Indian Sign Language, Japanese Sign Language, and so on. Generally, the semantic meanings of the language components in all these sign languages differ, but there are signs with a universal syntax. For example, a simple gesture with one hand expressing 'hi' or 'goodbye' has the same meaning all over the world and in all forms of sign languages.



Fig.1. Sign language Interpreter

Recognition of a sign language is very important not only from the engineering point of view but also for its impact on the human society [2]. The development of a system for translating Indian sign language into spoken language would be great help for hearing impaired as well as hearing people of our country. The research on sign language interpretation has been limited to small-scale systems capable of recognizing a minimal subset of a full sign language as shown in figure below. The reason for this is the difficulty in recognizing full sign language vocabulary. Recognition of gestures representing words and sentences undoubtedly is the most difficult problem in the context of gesture recognition research.



Fig.2. Different alphabets in Indian sign language

Apart from sign interpretation, there are various other applications of this work in domains including virtual environments, smart surveillance, medical systems etc. It can be used to recognize any sign language [3] in the world just by changing the samples to be trained. One of the effective applications that can utilize hand postures and gestures is robot tele-manipulation [4].

Now the next task is evaluating and analyzing how hand movements have been used in processes mediated through computers. Such an evaluation is of interest since the following problems in the design and implementation of whole hand centered input applications having remained unsolved.

**Tracking needs:** for most applications, it is difficult to decide the needs (e.g. how much accuracy is needed and how much latency is acceptable) when measuring hand motions.

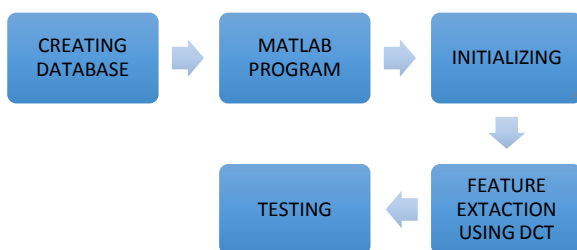
**Occlusion:** since fingers can be easily hidden by other body parts, sensors have to be placed nearby in order to capture their movements accurately.

**Size of fingers:** fingers are relatively small, so that sensors, if they are placed nearby, need to be small too, or, if sensors are placed remotely, need a lot of detail.

**Gesture segmentation:** how to detect when a gesture starts/ends, how to detect the difference between a (dynamic) gesture, where the path of each limb is relevant, and a (static) posture, where only one particular position of the limbs is relevant.

**Feature description:** which features optimally distinguish the variety of gestures and postures from each other and make recognition of similar gestures and postures simpler.

The task of this research work is to compare the images taken as input from webcam or test images with images in the training database and interpret the sign shown. This can be done by first applying Discrete Cosine Transform, extracting out its coefficients and then comparing these with the threshold to find the match. A flowchart representing the implementation procedure is as shown in figure below.



Overview of the procedure

Fig.3.

## II. IDEA OF PROPOSED SOLUTION

The focus here has been primarily on detecting hand gestures from images. While this does not ensure real time recognition in all environments, it is a major step forward in Indian Sign Language Recognition System. As the first step of implementation procedure, several Discrete Image Transforms are reviewed. This includes sinusoidal transforms such as Discrete Fourier Transform, Discrete Cosine Transform, Discrete Sine Transform, Hartley Transform and Non sinusoidal transform such as Hadamard Transform, Walsh Hadamard Transform, Haar Transform, Slant Transform and Hotelling Transform, etc. [5].

After studying these transforms, it is realized that the energy compaction is greatly achieved with DCT [6] and Hotelling Transform. Realizing the complexity and image dependency of Hotelling transform, Discrete Cosine Transform is selected for the implementation of proposed work.

A discrete cosine transform [7] expresses a sequence of finitely many data points in terms of a sum of cosine functions oscillating at different frequencies. The DCT converts a signal or image from the spatial domain to the frequency domain.

The general equation for 2D (N by M image) DCT Defined by the following equation,

$$F(u, v) = \sqrt{\frac{2}{N}} \sqrt{\frac{2}{M}} \sum_{i=0}^{N-1} A(i) * \cos\left(\frac{u(2i+1)\pi}{2N}\right) * \sum_{j=0}^{M-1} A(j) * \cos\left(\frac{v(2j+1)\pi}{2M}\right) * f(i, j) \quad (1)$$

Where  $f(i, j)$  is the 2D input sequence.

$$A(i) = A(j) = \frac{1}{\sqrt{2}}, \text{ for } u=0, v=0 \\ = 0 \quad \text{Otherwise}$$

The test image is taken as an input and DCT is applied to the image. Thus for 128 x 128 size image we have 16384 coefficients. After transformation, majority of the signal energy is carried by just a few of the low order DCT coefficients and the higher order DCT coefficients can be discarded. The coefficients are stored in zigzag [9] manner as shown below.

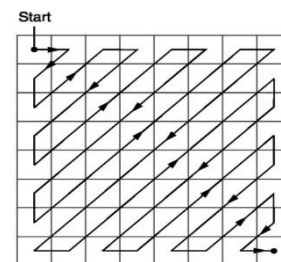


Fig.4. Zigzag Sequence

The coefficients are compared by calculating the Euclidean distance between 'Registered images' and 'Test image'. The test image is compared with the registered images in the database. The formula for this distance between a point  $X(X_1, X_2, \text{etc.})$  and a point  $Y(Y_1, Y_2, \text{etc.})$  is:

$$d = \sqrt{\sum_{i=1}^n (X_i - Y_i)^2} \quad (2)$$

Thus values ' $d_1, d_2, \dots, d_n$ ' are obtained which are sorted and the one with the minimum distance is regarded as the best match.

## III. IMPLEMENTATION STEPS

After being acquainted with the core concepts of image processing and reviewing several discrete image transforms, a road-map of the implementation procedure is to be created. In this process, a large set of images of hand gesture signs of a group of known people (e.g. Classmates) is taken. This is done so that inputting a

known/unknown hand gesture image will help interpreter to recognize the image quickly and to determine whether it matches the image to known Sign or not. The images in fig. 5 and 6, show the database that is used for training and testing.



Fig.5. Training Images

The basic implementation procedure for sign language interpreter using the discussed algorithm is shown in the fig.7.

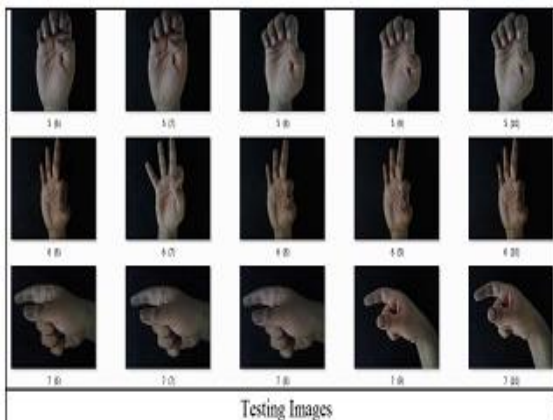


Fig.6. Testing Images

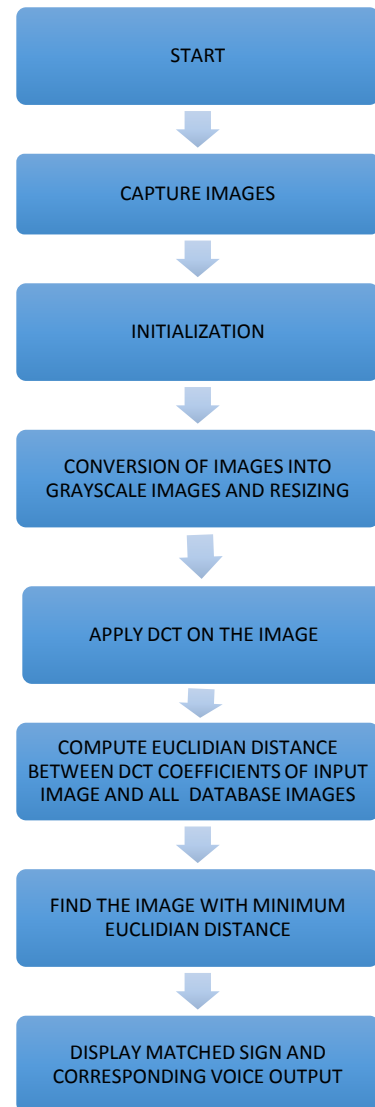


Fig.7. Basic implementation steps

### 3.1 Testing of hand gestures from test images

In this case, user can select any sign to test from test images. To obtain matched sign, following steps are followed.

- Extract the training images from the database. Convert the image into grey scale and then resize to 256 x 256.
- Calculate the DCT and extract DCT coefficients for all database images.
- Select the sign to be tested and calculate its DCT coefficients.
- Now calculate the Euclidian distance between DCT coefficients of input image and database images.
- Obtain the minimum Euclidian Distance and find out to which image in database it corresponds to. Display the matched image and input image selected by user. Play the sound corresponding to matched image.

### 3.2 Static input module

The above algorithm is modified to take image as input from webcam and then matched with the corresponding images in database.

### 3.3 Dynamic input module

The proposed algorithm is modified to take video (4 frames) as input for J and Z, various words for example hi, how, you, good, etc. and sentences (6 frames) for example “Hi!!! How are you?” at frame rate of 1/3 fps. The input frames are matched with corresponding frames in training database. If match is found, the video for input and matched sign along with the corresponding audio clip is played.

## IV. RESULTS

After the execution of several well thought cases, the results have been analyzed based on several criterions. To remove any possible subjective human bias in evaluating the results, any given case has been evaluated by a minimum of three human observers.

### Case 1: Static Input Module (Alphabets A-Y excluding J)

There are five signs placed in training and testing database corresponding to each alphabet. The signs from testing database or input from webcam are matched with the signs in the training database using the discussed algorithm. Result of this module is as shown below.

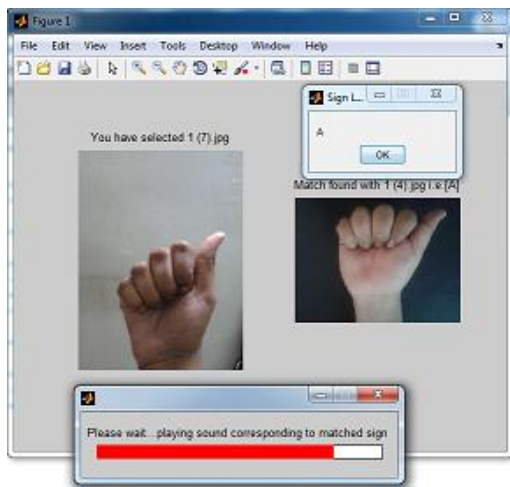


Fig.8. Match found for alphabet “A”

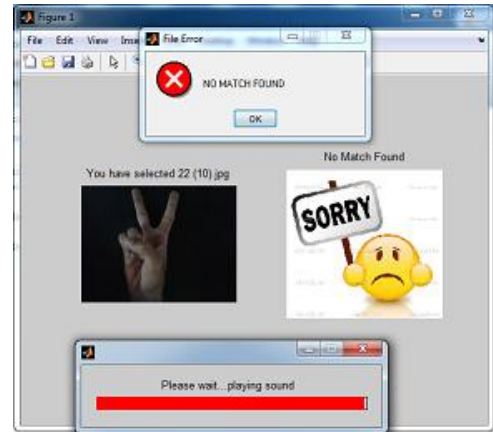


Fig.9. No match found

Table 1: Comparison of Accuracy rate for different No. of DCT coefficients

No. of DCT coefficients	No. of signs accepted	No. of signs not accepted	% Accuracy
8	97	23	80.83
16	108	12	90.00
32	112	8	93.33
64	115	5	95.83
128	114	6	95
256	113	7	94.16

$$\% \text{ Accuracy} = \frac{y_{cnt}}{y_{cnt} + n_{cnt}} \times 100 \quad (3)$$

Where  $y_{cnt}$  = No. of accepted signs  
 $n_{cnt}$  = No. of not accepted signs

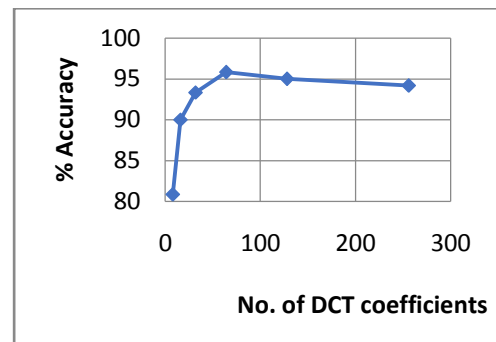


Fig.10. Plot of Accuracy for different No. of DCT coefficients

The above table and graph represent accuracy rate of testing images for different DCT coefficients.

### Case 2: Dynamic Input Module (Alphabets J and Z)

There are four frames per sign and five such signs placed in training and testing database corresponding to alphabet J and Z each. The signs from testing database or input from webcam are matched with the signs in the training database using the discussed algorithm. The graph below represents

the plot of percent error versus samples (A-Z) along with an image showing output of this module.

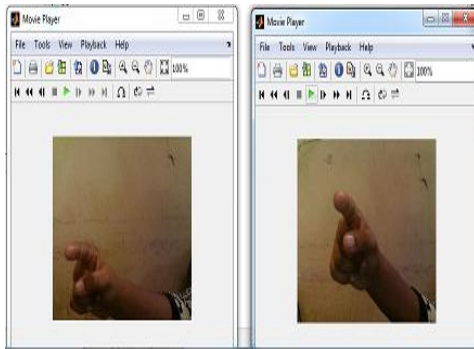


Fig.11. Match found for alphabet “Z”

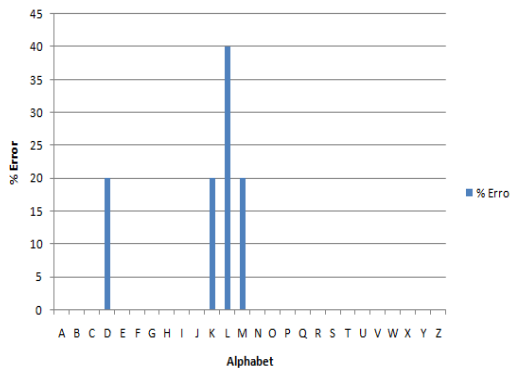


Fig.12. Plot of Percent Error versus samples

Case 3: Dynamic Input Module – Real time (Words and Sentences)

In this module, video input from webcam is taken in real time and matched with the frames in trained database. Result of this module is as shown below.



Fig.13. Match found for real time input for word “Hi”

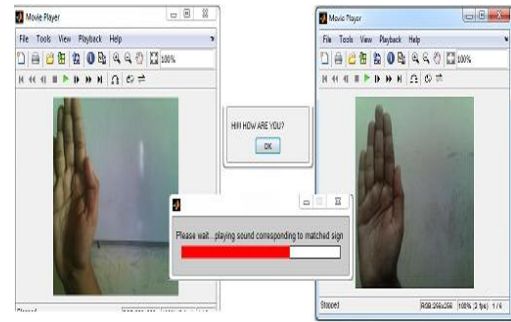


Fig.14. Match found for real time input for sentence “Hi!!! How Are You?”

V. CONCLUSION

The first and foremost requirement to start off with the proposed work was to create an efficient database. This was created using webcam in proper brightness conditions. To extract feature vectors of different signs from database, we selected the DCT – Euclidian Distance Algorithm. This algorithm involves accepting hand gesture frames as input from image capturing devices such as webcam. This input image is converted to grayscale and resized to 256 x 256 pixels. DCT is then applied to entire image to obtain the DCT coefficients.

Recognition is carried out by comparison of DCT coefficients of input image with those of the registered reference images in database. Euclidean Distance forms the basis of comparison. Database image having minimum Euclidean distance with the input image is recognized as the matched image and the sound corresponding to the matched sign is played. DCT approach gave fairly accurate results. Accuracy for different levels of DCT coefficients is obtained and tabulated as well as represented graphically. Entire work was implemented using MATLABR2013a [10] in Intel core i3 2.53 GHz processor with 1.3 Megapixel webcam.

Thus till now, this work has succeeded in recognizing static hand gestures completely and dynamic hand gestures including words and sentences to an appreciable level. The recognition of an alphabet in real time takes approximately 15 seconds and recognition of sentences consisting of three words on an average takes approximately 30 seconds. An important future scope in this project would be developing a real time application that can identify different hand gestures in Indian Sign Language in any environment. To overcome this constraint, analysis may be carried out using Discrete Wavelet Transform.

## REFERENCES

- [1] Automatic Sign Language Detection website, <https://sites.google.com/site/autosignlan>, October, 2013.
- [2] Sign Language Recognition System website, <http://www.iitg.ernet.in/isl/abtsign.html>, October, 2013.
- [3] Murthy, G. R. S., & Jadon, R. S., "A review of vision based hand gestures recognition", International Journal of Information Technology and Knowledge Management, 2(2), 405-410, 2009.
- [4] Joseph J. & LaViola J., "A survey of hand posture and gesture recognition techniques and technology", Master Thesis, NSF Science and Technology Center for Computer Graphics and Scientific Visualization, USA, 1999.
- [5] DhananjayTheckedath, "Digital Image Processing using Matlab codes", Fourth edition, Nandu Printers and Publishers Private Limited, 2009
- [6] Gonzalez, Woods, and Eddins, "Digital Image Processing Using MATLAB" second edition, 2009.
- [7] Aamer. S. S. Mohamed, Ying Weng, Jianmin Jiang and Stan Ipson; "An Efficient Face Image Retrieval through DCT Features", School of Informatics, University of Bradford, UK.
- [8] H. B. Kekre, KavitaSonawane, "Retrieval of Images Using DCT and DCT Wavelet over Image Blocks", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.
- [9] A.L.N.Benhadj-Djilali, "Another HCI challenge: Facial Feature Extraction made easy", Master's Thesis at Cognitive Science, Lund University, Sweden, 2004, pp.1-59.
- [10] Mathworks, "MATLAB Website", [www.mathworks.com](http://www.mathworks.com), November, 2013

IJERT