International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT206346

Hand Gesture Recognition and Voice Conversion for Hearing and Speech Aided Community

Mrs. Divya K V[1], Harish E[2], Nikhil Jain D[3], Nirdesh Reddy B[4]
[1]Sr.Assistant Professor, Department of Information Science, New Horizon College of Engineering, Bangalore, Karnataka, India
[2-4]Student, Department of Information Science, New Horizon College of Engineering, Bangalore, Karnataka, India

ABSTRACT

One of the major drawback of our society is the barrier that is created between disabled or handicapped persons and the normal person. Communication is the only medium by which we can share our thoughts or convey the message but for a person with disability (deaf and dumb) faces difficulty in communication with normal person. For many deaf and dumb people, sign language is the basic means of communication. Sign language recognition (SLR) aims to interpret sign languages automatically by a computer in order to help the deaf communicate with hearing society conveniently. Our aim is to design a system to help the person who trained the hearing impaired to communicate with the rest of the world using sign language or hand gesture recognition techniques. In this system, feature detection and feature extraction of hand gesture is done with the help of Support Vector Machine (SVM), K-Neighbors-Classifier, Logistic-Regression, MLP-Classifier, Naive Bayes, Random-Forest-Classifier algorithms are using image processing.

Keywords : Support Vector Machine, K-Neighbors-Classifier, Logistic-Regression, MLP-Classifier, Naive Bayes, Random-Forest-Classifier, Sign Language, Hand Gesture Recognition, Deaf and Dumb Communication, Image Processing, Machine Learning.

I. INTRODUCTION

Image processing is a rapidly growing area in diverse applications, such as multimedia computing, secured data communication, biomedical, biometrics, remote sensing, texture understanding, pattern recognition, content-based retrieval, compression, and many more. This is all about how a computer can sense pictorial data after processing an image. Among the set of gestures intuitively performed by humans when communicating with each other, pointing gestures are especially interesting for communication and is perhaps the most intuitive interface for selection. They open up the possibility of intuitively indicating objects and locations, e.g., to make a robot change moving direction or simply mark some object. This is particularly useful in combination with speech recognition as pointing gestures can be used to specify parameters of location in verbal statements. This technology can be a boon for disable people who are not able to speak hence can't communicate. Also if the person has different language than receiver, then also, it can be used to as translator. There has been always considered a challenge the development of a natural interaction interface, where people interact with technology as they are used to interact with the real world. A hand free interface, based only on human gestures, where no devices are attached to the user,

will naturally immerse the user from the real world to the virtual environment.

Hands are human organs which are used to manipulate physical objects. For this very reason hands are used most frequently by human beings to communicate and interact with machines. Mouse and Keyboard are the basic input/output to computers and the use of both of these devices require the use of hands. Most important and immediate information exchange between man and machine is through visual and actual aid, but this communication is one sided. Computers of this age provide humans with 1024 * 768 pixels at a rate of 15 frames per second and compared to it a good typist can write 60 words per minute with each word on average containing 6 letters. To help somewhat mouse remedies this problem, but there are limitations in this as well. Although hands are most commonly used for day to day physical manipulation related tasks, but in some cases they are also used for communication. Hand gestures support us in our daily communications to convey out- messages clearly. Hands are most important for mute and deaf people, who depends their hands and gestures to communicate, so hand gestures are vital for communication in sign language. If computer had the ability to translate and understand hand gestures, it would be a leap forward in the field of human computer interaction. The dilemma, faced with this is that the images these days are information tick and in-order to achieve this task extensive processing is required.

II. RELATED WORKS

1) In [1] paper, the author has designed with Indian Sign Language (ISL) is a visual-spatial language which provides linguistic information using hands, arms, facial expressions, and head/body postures. The proposed work aims at recognizing 3D dynamic signs corresponding to ISL words. With the advent of 3D sensors like Microsoft Kinect Cameras, 3D geometric processing of images has received much attention in recent researches. The authors have captured 3D dynamic gestures of ISL words using Kinect camera and has proposed a novel method for feature extraction of dynamic gestures of ISL words.

2) In [2] paper, author has briefly explained about the hand gesture recognition. The main goal of gesture recognition is to create a system which can recognize specific human gestures and use them to convey information or for device control. Hand gestures provide a separate complementary modality to speech for expressing ones ideas. Information associated with hand gestures in a conversation is degree, discourse structure, spatial and temporal structure. The approaches present can be mainly divided into Data-Glove Based and Vision Based approaches. An important face feature point is the nose tip. Since nose is the highest protruding point from the face. Besides that, it is not affected by facial expressions. Another important function of the nose is that it is able to indicate the head pose. Knowledge of the nose will enable us to align an unknown 3D face with those in a face database. Eye detection is divided into eye position detection and eye contour detection. The purpose of this paper is to compare various human Gesture recognition systems for interfacing machines directly to human wits without any corporeal media in an ambient environment.

3) The [3] paper has briefly explained image processing. It usually takes a fusion of image processing and machine learning algorithms in order to build a fully-functioning computer vision system for hand gesture recognition. Fortunately, the complexity of developing such a system could be alleviated by treating the system as a collection of multiple sub-systems working

together, in such a way that they can be dealt with in isolation. Machine learning need to feed on thousands of exemplars (e.g. images, features) to automatically establish some recognizable patterns for all possible classes (e.g. and gestures) that applies to the problem domain. A good number of exemplars helps, but it is also important to note that the efficacy of these exemplars depends on the variability of illumination conditions, hand postures, angles of rotation, scaling and on the number of volunteers from whom the hand gesture images were taken. These exemplars are usually subjected to image processing first, to reduce the presence of noise and extract the important features from the images. These features serve as inputs to the machine learning system. Different sub-systems are integrated together to form a complete computer vision system for gesture recognition. The main contribution of this work is on the production of the exemplars. A minor contribution is given in the form of a specific feature extraction method called moment invariants, for which the computation method and the values are furnished with the dataset.

4) In paper [4], the RGB, HSV and YUV (YCbCr) are standard models used in various color imaging applications, not all of their information are necessary to classify skin color. This paper presents a novel skin color model, RGB-H-CbCr for the detection of human faces. Skin regions are extracted using a set of founding rules based on the skin color distribution obtained from a training set. The segmented face regions are further classified using a parallel combination of simple morphological operations. This model utilizes the additional hue and chrominance information of the image on top of standard RGB properties to improve the discriminality between skin pixels and non-skin pixels. In the proposed approach, skin regions are classified using the RGB boundary rules introduced by Peer et al. and also additional new rules for the H and CbCr subspaces. These rules are constructed based on the skin color distribution obtained from the training images. The classification of the extracted regions is further refined using a parallel combination of morphological operations.

III. PROPOSED SYSTEM

The first step for our proposed system is the capturing of the video using webcam where different alphabets were taken into consideration. Skin Filtering was performed to the input video frames for detection of hand gestures. It was done so that the required hand could be extracted from the background. Skin Filtering is a technique used for separating the skin colored regions from the non-skin colored regions. In our proposed system there are 5 modules: real time Input image from webcam, preprocessing and segmentation, feature extraction, classification and Results analysis (gesture recognition). For gesture recognition is real time recognized in live camera .The proposed system are used in SVM (Support Vector Machine), KNeighbors-Classifier, LogisticRegression, MLPClassifier, Naive Bayes, RandomForestClassifier algorithms. We propose an easy-to-use and inexpensive approach to recognize single handed as well as double handed gestures accurately. This system can definitely help millions of deaf people to communicate with other normal people. A fast, novel and robust system was proposed for recognition of different alphabets of Indian Sign Language for video sequences. The proposed system is a real time video processing that is based on a real time application system.

IV. MODULES AND METHODOLOGIES

The proposed system (Architecture Diagram) consists are five modules of following steps to interpret the gesture from the input image such as:

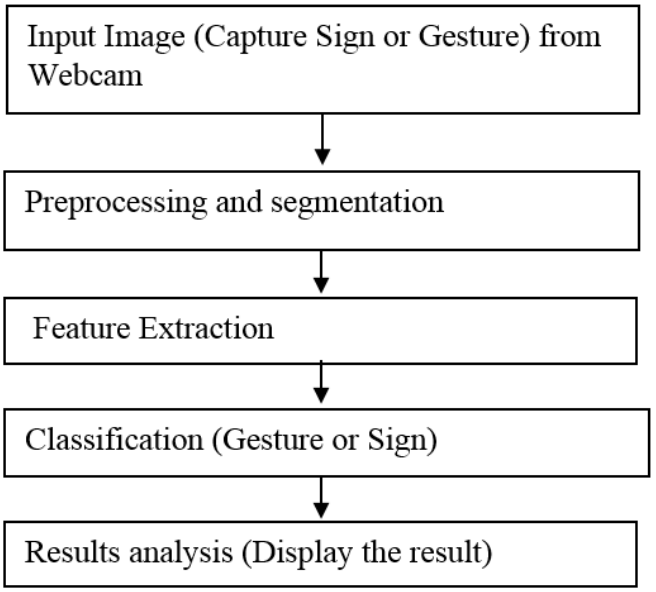

Figure 1: Overview of the Architecture Diagram in five modules

Input Image (Capture sign or Gesture) from Webcam:

The image (gesture or sign) is captured using the laptop camera or the external device webcam to get better image clarity.

Preprocessing and segmentation:

Image processing is necessary for image enhancement. During Preprocessing RGB image to convert into HSV color space. This step was taken because HSV color space was less sensitive to illumination changes compared to RGB. Then it was filtered, smoothened and finally the biggest binary linked object was being considered so as to avoid consideration of skin colored objects other than hand. .To obtains the good result smoothing and filtering is done. Image segmentation is basically performed to locate the hand object in image.

Feature Extraction:

Feature Extraction stage is necessary because certain features has to be extracted so that they are unique for each gesture or sign. After the decision is made that a sign is present, then the last frame is taken into consideration and features. The Feature Extraction is extract the features in all Images (gesture or sign) dataset are store in 'svm.pkl' and finally extract the labels store in 'labels.pkl' best on train data and test data.

- **Classification:**

Classification of hand is done with the help of various features calculated previously. The five bit binary sequence is thus generated to uniquely recognize and utilize these recognized the recognized hand gesture for supporting human computer interaction. By the feature extraction significant peak is encoded as 1 while insignificant peak is encoded as 0 based on intersection to the threshold line.

- **Results analysis:**

Different images were tested and found that the new technique of classification was found to show 97% accuracy. Some images tested with other database images are given in the results analysis. In Results analysis are real time detect the sign language and sign recognize when live camera is start then capture the test images (gesture or sign)that time compare the features 'svm.pkl' and 'labels.pkl' if it is match the dataset after the process in display the result.

V. METHODOLOGIES

A **methodology** is a repeatable process with project-specific methods. This methodology consists of the following phases such as:

phase1: Requirement of the data collection will be performed using two dataset are Train data and Test data.

phase2: During **Technology adaptation**, will be undertaken with three process such as:

a) **Feature Extraction:**

Feature Extraction is extract the all Images data using Hog algorithm can provides all dataset features.

b) Preprocessing:

Preprocessing is processed the dataset apply six best algorithms in preprocessed the SVM, KNeighbors Classifier, Logistic Regression, MLP Classifier, Naïve Bayes, Random Forest Classifier all are create more score such as accuracy score, precision score, f1 score, recall score of the test and train dataset after SVM algorithm create the dataset features store in 'svm.pkl' and finally extract the labels store in 'labels.pkl' best on train data and test data.

Classification:

Classification can classified the both test data and train data predict in sign language using 'svm.pkl' and 'labels.pkl'

phase3: **Results analysis** are provide real time detect the sign and sign recognize when live camera start capture the test images that time compare the features 'svm.pkl' and 'labels.pkl' if it is match the dataset after that display the result.

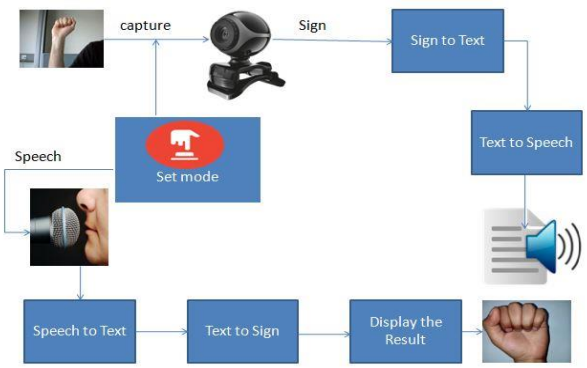

Figure 2: Overview of the High Level Block Diagram.

VI. RESULT AND DISCUSSION

We have proposed a system for recognizing a dynamic hand words gesture of Indian sign language and conversion of recognized gesture into text and speech and vice versa i.e. dual way communication.

VII. CONCLUSION

In this paper, we have proposed a system for recognizing a dynamic hand words gesture of Indian sign language and conversion of recognized gesture into text and speech and vice versa i.e. dual way communication. In this system skin color filtering technique has used for segmentation. Eigen vectors and Eigen values technique has used for feature extraction. For classification, Eigen value weighted Euclidean Distance based classifier has used. Prediction of words sign using one or both hand, working with Indian Sign language dynamic hand gesture words dataset and dual way communication has proposed in this system.

VIII. REFERENCES

[1]. A. Chaudhary, J.L. Raheja, K. Das, S. Raheja, "A Survey on Hand Gesture Recognition in context of Soft Computing", Published as Book Chapter in "Advanced Computing" CCIS, Springer Berlin Heidelberg, Vol. 133, 2011, pp. 46-55

[2]. A. Chaudhary, J.L. Raheja, K. Das, "A Vision based Real Time System to Control Remote Robotic hand Fingers", In proceedings of the IEEE International Conference on Computer Control and Automation, South Korea, 1-3 May, 2011, pp. 118-122

[3]. J.L. Raheja, A. Chaudhary, S. Maheshwari, "Automatic Gesture Pointing Location Detection", Optik: International Journal for Light and Electron Optics, Elsevier, Vol. 125, Issue 3, 2014, pp. 993-996.

[4]. A. Chaudhary, K. Vatwani, T. Agrawal, J.L. Raheja, "A Vision-Based Method to Find Fingertips in a Closed Hand", Journal of Information Processing Systems, Korea, Vol. 8, No. 3, Sep 2012, pp. 399-408.

[5]. A. Chaudhary, J.L. Raheja, K. Das, S. Raheja, "Fingers' Angle Calculation using Level-Set Method", Published as Book Chapter in

"Computer Vision and Image Processing in Intelligent Systems and Multimedia Technologies", IGI, USA, April 2014, pp.191-202

[6]. D. Sturman, D. Zeltzer, "A survey of glove-based input", IEEE Transactions on Computer Graphics and Applications, Vol. 14, No. 1, Jan. 1994, pp. 30-39.

[7]. T.S. Huang, and V.I. Pavlovic, "Hand gesture modeling, analysis and synthesis", Proceedings of international workshop on automatic face and gesture recognition, 1995, pp.73-79.

[8]. O.Y. Cho, and et al. "A hand gestue recognition system for interactive virtual environment", IEEK, 36-s(4), 1999, pp. 70-82.

[9]. M. Sonka, V. Hlavac, R. Boyle, "Image processing, analysis, and machine vision"2014, Cengage Learning.

[10]. A. Chaudhary, M. B. L. Manasa, J.L. Raheja, "Light invariant Neuro-Vision System for Elder/Sick people to express their needs into Lingual Description", Microsoft Research India's annual research symposium TechVista, India, 20 Jan, 2012.

[11]. A. Chaudhary, J.L. Raheja, "ABHIVYAKTI: A Vision Based Intelligent System for Elder and Sick Persons", In proceedings of the 3rd IEEE International Conference on Machine Vision, Hong Kong, 28-30 Dec, 2010, pp. 361-364.

[12]. J L Raheja, Dhiraj, D Gopinath, A. Chaudhary, "GUI System for Elder/Patients in Intensive Care", In proceedings of the 4th IEEE International Conference on International Technology Management Conference, Chicago, IL USA, 12-15 June, 2014, pp.1-5.

[13]. J. Rekha, J. Bhattacharya and S. Majumder, "Hand Gesture Recognition for Sign Language: A New Hybrid Approach", 15th IPCV'11, July, 2011, Nevada, USA.

[14]. A.W. Fitzgibbon, and J. Lockton, "Hand Gesture Recognition Using Computer Vision", BSc. Graduation Project, Oxford University.

[15]. http://en.wikipedia.org/wiki/Principal_component_analysis

Cite this article as :

Mrs. Divya K V, Harish E, Nikhil Jain D, Nirdesh Reddy B, "Hand gesture recognition and Voice conversion for hearing and speech aided community", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 3, pp.223-228, May-June-2020.
Available at doi : https://doi.org/10.32628/CSEIT206346
Journal URL : http://ijsrcseit.com/CSEIT206346