Hand Gesture Recognition System Using Image Processing*

Mr. Sagar P.More
M.E. Student Digital Electronics
B.N.College of Engineering,Pusad.
Dist:-Yavatmal, India.
Sagar_pmore@rediffmail.com

Prof. Abdul Sattar
Department of Electronics and Communication Engg.
B.N.College of Engineering,Pusad.
Dist:-Yavatmal, India
shaiksattar@rediffmail.com

Abstract—**A gesture is a form of non verbal communication or non vocal communication in which visible bodily actions communicate particular messages, either in place of, or in conjunction with, speech. Gestures include movement of the hands, face, or other parts of the body. Gestures differ from physical non-verbal communication that does not communicate specific messages, such as purely expressive displays, proxemics, or displays of joint attention. Gestures allow individuals to communicate a variety of feelings and thoughts, from contempt and hostility to approval and affection, often together with body language in addition to words when they speak.**

Gesture processing takes place in areas of the brain such as Broca's and Wernicke's areas, which are used by speech and sign language. In fact, language is thought to have evolved from manual gestures. The theory that language evolved from manual gestures, termed Gestural Theory.

Keywords—Introduction; Literature survey; Pre-Processing and Segmentation; Feature Extraction; PCA method

I. INTRODUCTION

Sign language is widely used by people who can't speak and hear or people who can hear but can't speak. A sign language is composed of various gestures formed by different hand shapes, movements and orientations of hands or body, or facial expressions. There are various sign languages across the world, each with its own vocabulary. These include American Sign Language (ASL) in Northern America, British Sign Language (BSL) in Great Britain, Japanese Sign Language (JSL) in Japan, South African Sign Language (SASL), Indian Sign language (ISL) etc. Gestures are used by the deaf people to express their thoughts. But the use of these gestures are always limited in the deaf-dumb community, normal people never try to learn the sign language. This causes a big gap in communication between the deaf–dumb people and the normal people. Usually deaf people seek the help of sign language interpreters for translating their thoughts to normal people and vice versa. But this system is very costly and does not work throughout the life period of a deaf person. So a system that automatically recognizes the sign language gestures is necessary. Such a system can minimize the gap between deaf people and normal people in the society.

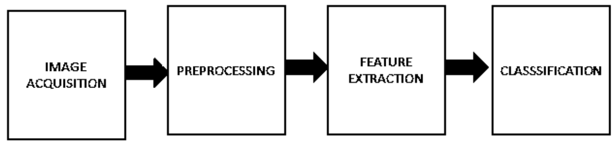


Fig.1.1 Block Diagram of hand gesture recognition system

The general block diagram of hand gesture recognition system shown in Fig 1 as above, it consists of four basic steps such as, Image acquisition, Pre-processing, Feature extraction and classification. Hand Gestures provide a natural and intuitive communication modality for human–computer interaction. Efficient human computer interfaces (HCIs) have to be developed to allow computers to visually recognize in hand gestures. However, vision-based hand tracking and gesture recognition is a challenging problem due to the complexity of hand gestures, which are rich in diversities due to high degrees of freedom (DOF) involved by the human hand. In order to successfully fulfill their role, the hand gesture HCIs have to meet the requirements in terms of real-time performance, recognition accuracy and robustness against transformations and cluttered background.

To facilitate the communication between the Deaf and hearing persons, many gesture recognition systems used the help of color markers or data gloves to make the task easier. However, using of markers and gloves sacrifices the user's convenience. In the dissertation work, focus is given hand gesture recognition without help of any markers and gloves. Detecting and tracking hand gestures in a sequence of images help in extracting hand region. Thus, processing time will be reduced and accuracy will be increased as the features of that region will represent the hand gesture only.

II. LITERATURE SURVERY

A. Related Work

Several methods are available for sign language recognition.

978-1-4673-9939-5/16/$31.00 ©2016 IEEE

N.H. Dardas and N.D. Georganas [1], presented a novel and real-time system for interaction with an application or video game via hand gestures. This system includes detecting and tracking bare hand in cluttered background using skin detection and hand posture contour comparison algorithm after hand subtraction, recognizing hand gestures via bag-of-features and multiclass support vector machine (SVM) and building a grammar that generates gesture commands to control an application.

AEl-Sawah, N. Georganas, et.al [2] ,explained the experience in designing a prototype for 3-D hand tracking and dynamic gesture recognition. Objective is to be able to continuously visually track the hand in a general background and to be able to recognize dynamic gestures in real time.

Vinod P. R, Usha Gopalkrishnan, et.al [3], proposed a method for the automatic recognition of finger spelling in Indian sign language. The proposed method uses digital image processing techniques and artificial neural network for recognizing different signs.

P. Mekala, Ying Gao, et.al [4], proposed architecture is being using the neural networks identification and tracking to translate the sign language to a voice/text format.

S. Pandita, S. P. Narote, et.al [5] and [6] presented a method for recognizing hand gestures by extracting distinctive invariant features from images that can be used to perform efficient matching between different views of a hand gesture. The features are invariant to image scale and rotation, and provide robust matching across a considerable range of offline distortion, change in 3D view point, addition of noise, and change in illumination.

J.Singha, K. Das [7], proposed a system using Eigen value weighted Euclidean distance as a classification technique for recognition of various Sign Languages of India. The system comprises of four parts are Skin Filtering, Hand Cropping, Feature Extraction and Classification.

D.Y. Huang, W.C. Hub, et.al [8], presented a novel approach for hand gesture recognition based on Gabor filters and support vector machine (SVM) classifiers for environments with varying illumination.

Y. Wu, J. Lin, et.al [9], presented a model-based method to capture hand articulation by learning hand natural constraints.

H. Zhou and T. S. Huang [10], introduced the concept of Eigen-dynamics and propose an Eigen dynamics analysis (EDA) method to learn the dynamics of natural hand motion from labelled sets of motion captured with a data glove.

T. Sterner and A. Pent Land [11], described a real-time HMM-based system for recognizing sentence level American Sign Language (ASL).

L. Bretzner, et.al [12], presented algorithms and a prototype system for hand tracking and hand posture recognition. Hand postures are represented in terms of hierarchies of multi-scale color image features at different scales, with qualitative inter-relations in terms of scale, position and orientation.

A. Argyros and M. Lourakis [13], proposed a vision-based interface for controlling a computer mouse via 2D and 3D hand gestures. The proposed interface builds upon our previous work that permits the detection and tracking of multiple hands that can move freely in the field of view of a potentially moving camera system.

Qing Chen and N.D. Georganas [14], proposed a new approach to solve the problem of real-time vision-based hand gesture recognition with the combination of statistical and syntactic analysis.

G. R. S. Murthy and R. S. Jadon [15] designed method for recognizing specific hand gestures and use them to convey information.

N.A.Ibrahim and R. Z. Khan [16], proposed a method which can recognize specific human gestures and use them to convey information or for device control.

G. Deepa and G. Bhaskar Phani Ram [17], proposed system emotion is derived automatically from hand gestures captured by camera. The data recorded from hand gestures are coupled with musical emotion.

B. Research Gap

To facilitate the communication between the Deaf and hearing persons, highly skilled interpreters have traditionally been used. These interpreters tend to be very costly and it is a great effort to become a good interpreter that can translate between a spoken language and a sign language correctly and efficiently. The use of an interpreter is not always appropriate and they need to be notified in advance when their services are to be required. Another important fact to consider is that there will simply never be enough good trained interpreters that can assist the millions of deaf people.

A Machine Translator (MT) system that can translate between a spoken language, such as English and a sign language, such as ASL and ISL will solve the above problem of insufficient interpreters. To visualize a sign language, an MT system must be generated. Such a system can be used in many different application domains, such as Deaf telephony as well as English and sign language education. This will make it easier for the Deaf to access the various public and information services.

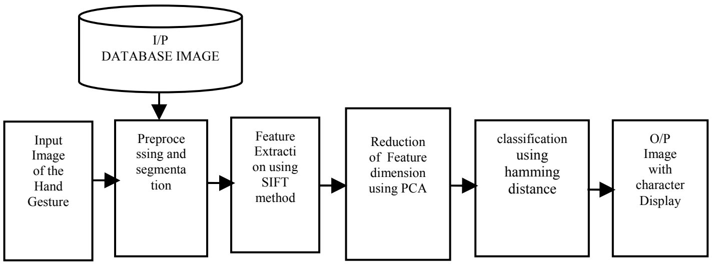
III. MOTIVATION

To facilitate the communication between the Deaf and hearing persons, many gesture recognition systems used the help of color markers or data gloves to make the task easier [11]. However, using of markers and gloves sacrifices the user's convenience. In the dissertation work, focus is given hand gesture recognition without help of any markers and gloves. Detecting and tracking hand gestures in a sequence of images help in extracting hand region. Thus, processing time will be reduced and accuracy will be increased as the features of that region will represent the hand gesture only.

A. Objective

In Our work, the objective is to create accurate recognition for the detected hand postures using SIFT algorithm. The advantage of using the algorithm is high processing speed which can produce efficient recognition results. The SIFT features in the implementation will reduce the feature dimension vector, to compute at the edges which will be invariant to scaling, rotation, addition of noise. The hand gesture will be used in extracting the key points by scale invariance feature transform (SIFT) algorithm and reduce the dimension of feature vector of an image using Principle Component Analysis (PCA). SIFT features, proposed are features (key points) extracted from images to help in reliable matching between different views of the same object, image classification, and object recognition. The extracted key points are invariant to scale, orientation and partially invariant to illumination changes, and are highly distinctive of the image. Therefore, the modified SIFT is adopted in dissertation work for the bare hand gesture recognition.

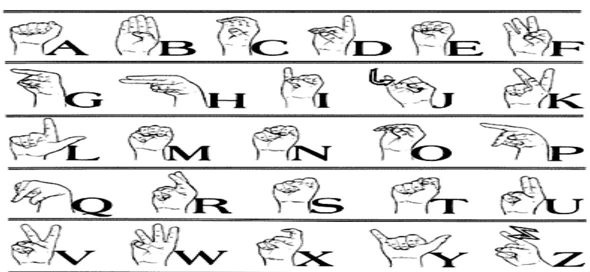
B. Block Diagram



"Fig. 3.1," Project Block Diagram

In dissertation work, hand gesture recognition system consists of the offline testing which is extended by using hand gesture detection. The model is used to recognized hand gesture captured from webcam. Before building the model, images will be captured for each hand gesture, which are the fist, index, palm, and little finger gestures, for different people, scales, and rotations and under different illuminations. The system can also recognize any other gestures, such as two, three, and five. All the training images illustrate the hand gestures without any other objects, and the background has no texture or objects (white wall).

Hand Gesture Detection in a cluttered background will enhance the performance of hand gesture recognition system in terms of accuracy and speed because key points extracted will represent the hand gesture. The Indian and American Sign Language alphabets for A to Z are shown in Figure no 3.2 These alphabets are used for communication of deaf persons. In this dissertation work we have used the Mat lab software for execution of program using digital image processing. The distortion between a input image and database image is extracted using pre-processing and a segmentation.



"Fig.3. 2," Hand gesture Sign Alphabets

Fig 3.3 Hand Gesture of V Image

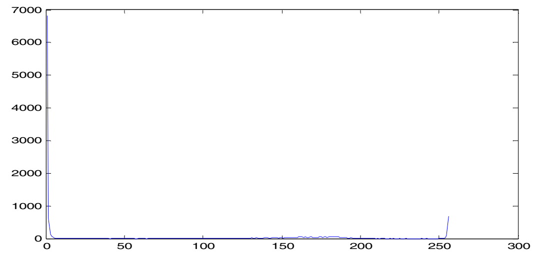


Fig 3.4 Histogram Of Hand Gesture "V"

C. Pre-Processing and segmentation

Local changes due to noise and digitization errors should not radically alter the image scene and information. In order to satisfy the memory requirements and the environmental scene conditions, pre-processing of the raw data is highly important. The objective of gesture segmentation is to extract the gesture region from the background of the image. Hand segmentation is the process of extracting the hand sign from the captured image Efficient hand segmentation has a key role in sign language recognition task. The segmentation process depends on the type of gesture, if it is dynamic gesture then the hand gesture need to be located and tracked, if it is static gesture the input image has to be segmented only.

The result of segmentation produces a binary image with the skin pixels in white color and background in black color. The resulting binary image may contain noise and segmentation error filtering and morphological operations are performed on the input image to decrease noise and segmentation errors if any. Image pre-processing includes the set of operations on images whose goal is the improvement of the image data that suppresses undesired distortions or enhances some image features important for further processing.

D. Interest Point Discriptors

SIFT features extraction and matching

SIFT features are distinctive, invariant features extracted from images that allow for efficient matching with various other view points of the extracted features that may exist in the same or different gestures. The features are invariant to image translation, scaling, and rotation also partially invariant to illumination changes.

The computation of SIFT image features is in four basic steps

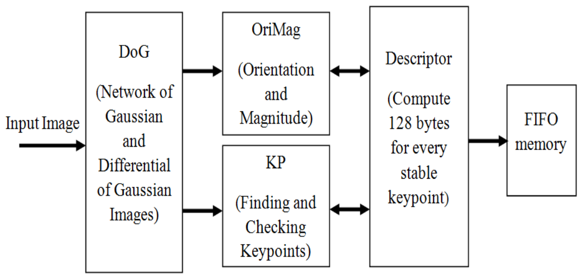
1 Scale-space Local Extreme Detection:

The first step computes the locations of potential interest points in the image by detecting the maxima and minima of a set of difference of Gaussian filters applied at different scales all over the image. The SIFT algorithm was initially presented in a paper by David Lowe in 1999 (Lowe 1999) and then he summarized his algorithm in 2004(Lowe 2004).

Scale Invariant Feature Transform is used to extract the features of image by finding and Describing keypoints.

The SIFT algorithm for image recognition is partitioned into four different stages as shown in Figure 1. Data of an image is entered in the circuit and in the first block, differential of Gaussian images are computed. The results of DoG block are prepared for two next blocks. OriMag calculates orientation and gradient magnitude for every pixel. Finding and checking the stability of keypoints are done in KP block. Finally, the descriptor block computes a 128 bytes definition for every stable keypoint. There is a FIFO memory in output to save the results of the plan.

2. Keypoint Localization: The keypoint localization of the set of interest points found in Step 1, a stable set is selected and further localized in image space.



"Fig. 3.5," Block Diagram of SIFT

3. Orientation Assignment: For all stable interest points selected, an orientation is then assigned to each key point based on local image features.

4. Key point descriptor: Finally, a local feature descriptor is computed at each key point. This descriptor is based on the local image gradient, transformed according to the orientation of the key point to provide orientation invariance.

IV. DIMENSIONALITY REDUCTION TECHNIQUE

As previously mentioned, our goal is to reduce the dimension of the feature vectors. There are several ways to achieving such goal, two of them will be: discovering more distinctive features and applying dimensionality reduction to the existing descriptors feature vectors. Principal Components Analysis (PCA) is a technique recommended when there is a large amount of numeric variables (observed variables) and it is desired to find a lower number of artificial variables, or principal components that will be responsible for the higher variance in the observed variables. Then, these principal components can be used as predictor variables in subsequent analysis. It converts an original random vector, whose components have correlations, to a new random vector, whose components have no correlations, by an orthogonal transformation. Its description procedure can be divided into two sub-steps, projection matrix generating and descriptor establishing. It makes a new vector significantly smaller than a standard SIFT vector.

A. Introduction To PCA

Principal component analysis (PCA) is probably the most widely used subspace projection technique for hand gesture recognition also known by Eigen hand gesture based approach. Principal Component Analysis is a universally used dimensionality reduction technique for image-processing applications. PCA is an optimal feature extraction and dimensionality reduction technique, from the Information theoretic point of view. The idea is to find the components or the dimensions along which the collection of all possible images is expected to have its energy distributed. Then only those dimensions are retained, and the rest of the images are discarded for the future stages of processing.

Eigen gestures suggest the use of information of coding and decoding of hand gesture image, emphasizing the significant local and global features. Such features may or may not be directly related to our intuitive notation of hand features such as the eye, nose, lips and hair. The relevant information of hand image is extracted, encoded as efficiently as portable and then compared with a database of models encoded. Similarly a simple approach to extracting the information contained in an image of a hand is to somehow capture the variation in a collection of hand images, independent of any judgment of features, and use this information to encode and compare individual hand images.

In mathematical terms, the principal components is the distribution of hands or the Eigen vectors of the covariance matrix of the set of hand images treating an image as point in a very high dimensional space is sought. The Eigen vectors are ordered, each one accounting for a different amount of the variation among the hand images. These eigenvectors are the set of features that together characterize the variation between hand images. Each image location contributes more or less to each eigenvector, so that it is possible to display these Eigen vector as a sort of ghostly hand gesture image, which is called as "Eigen hands". Each Eigen hand deviates from uniform gray where some facial feature differs among the set of training hands. Eigen hand gesture can be viewed as sort of map of the variation between hands. Therefore, each individual is characterized by a small set of features or Eigen picture weights needed to describe and reconstruct them.

B. Classification Method

The Hamming or Euclidean distance between two feature vectors to determine whether the two vectors belong to the same key point in different images. Threshold this distance generates a binary decision, and adjusting this threshold enables one to select the appropriate trade-off between false positives and false negatives and last stage of proposed block diagram we get a output image of hand gesture with character display.

V. Applications, Limitaioins and Advantages.

A. Applications

1. The system can be used at public places like, Airports, Railway Stations, Counters of Banks, Hotels, etc.
2. It can be used in intelligent classrooms and intelligent environments for real-time translation.
3. It can be also used in Virtual reality and computer games. As a translating device for Mute people.

B. Limitations

1. More light intensity exposure on hand gesture images affects the result accuracy.
2. Proper hand gesture image should be taken.
3. In hand gesture, image resolution, pixel density should be moderate.

C. Advantages

1. Efficient way for mute communication.
2. Less time delays.
3. Quick Response Time.
4. Fully automate system robust system, and low power requirement.

VI. Conclusion

The various approaches of hand gesture recognition have been discussed. The proposed method has been presented. The sign for all alphabets A to Z will be recognized using SIFT. The advantage of using the algorithm is high processing speed which can produce better results. The SIFT features in our implementation will compute at the edges which will be invariant to scaling, rotation, addition of noise. These features will be useful due to their distinctiveness, which enables the correct match for key points between different hand gestures. Also computation time will be less. This work can further be extended to text to speech conversion.

Acknowledgment (Heading 5)

I would like to express my gratitude to my supervisor Prof. Abdul Sattar for his guidance, advice and constant support throughout my thesis work. I would like to thank him for being my advisor here at B.N. College of Engineering, Pusad.

References

[1] Nasser H. Dardas and Nicolas D. Georganas, "Real-Time Hand Gesture Detection and Recognition Using Bag-of-Features and Support Vector Machine Techniques," *IEEE Transaction on Instrumentation and Measurement, volume 60, no. 11, pp.132-137 November 2011.*

[2] A. El-Sawah, et.al "A prototype for 3-D hand tracking and gesture estimation," IEEE Transaction Instrumentation Measurement. volume 57,no. 8, pp. 1627–1636, Augest 2008.

[3] Priyanka Mekala, et.al , Asad Davari"Real-time Sign Language Recognition based on Neural Network Architecture" *978-1-4244-9592-4/11/ ©2011IEEE.*

[4] S. Pandita, and S. P. Narote "Hand Gesture Recognition using SIFT" International Journal of Engineering Research and Technology (IJERT) Volume 2, no.1, pp1760-1764, January 2013.

[5] Usha Gopalkrishanan , Vinod P. R, et.al " Artificial Neural Network Based Method for Indian Sign Language Recognition" *IEEE Conference on Information and Communication Technologies (ICT 2013).*

[6] D. G. Lowe, "Distinctive image features from scale-invariant key points, "International Journal of Computer Science, volume 60, no. 2, pp. 91–110, November 2004.

[7] J. Singha and K. Das, "Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique*",(IJACSA) International Journal of Advanced Computer Science and Applications, Volume 4, no. 2 , pp. 188-195,July 2013.*

[8] D.Y. Huang, W.C Hub, and et.al, "Gabor filter-based hand-pose angle estimation for hand gesture recognition under varying illumination*", Expert Systems with Applications 38,June 2011.*

[9] Y. Wu, J. Lin, and et.al, "Capturing Natural Hand Articulation". *In IEEE International Conference on Computer Vision II, pp.426–432, September 2001.*

[10] H. Zhou and T. S. Huang, "Tracking articulated hand motion with Eigen dynamics analysis," *in Processing Of International Conference on Computer Vision, volume 2, pp. 1102–1109, November 2003.*

[11] T. Sterner and A. Pent Land, "Real-time American sign language recognition from video using hidden markov models", *Technical Report, M.I.T Media Laboratory Perceptual Computing Section, Technical Report No. 375, 1995.*

[12] L. Bretzner, et.al "Hand gesture recognition uses multi-scale color features, hierarchical models and particle _ltering*", in Processing of International Conference. Au-tom. hand Gesture Recognition, Washington, DC, May 2002.*

[13] A. Argyros and M. Lourakis, "Vision-based interpretation of hand gestures for remote control of a computer mouse", *in Processing of Workshop Computer Human Interaction, pp.4051, July 2006.*

[14] Qing Chen, N. D. Georganas, "Hand Gesture Recognition Using Haar-Like Featuresand a Stochastic Context-Free Grammar*", IEEE transactions on instrumentation and measurement, volume 57, no. 8, pp.113-117, August 2008.*

[15] G. R. S. Murthy and R. S. Jadon, "Hand Gesture Recognition using Neural Networks", *IEEE Transactions, Volume 6, pp.427-431, May 2010.*

[16] N.A.Ibrahim and R.Z. Khan, "Survey on Various Gesture Recognition Technologies and Techniques", *International Journal of Computer Applications, Volume 50, No. 7, pp1230-1235, May2014.*

[17] G. Deepa, et.al 'Emotion based Music Retrieval for Gesture Recognition' *International Journal of Computer Applications, Volume 78, no.14, pp.0975 – 8887, September 2013.*