ISSN No. 0976-5697

Volume 4, No. 6, May 2013 (Special Issue)

International Journal of Advanced Research in Computer Science

REVIEW ARTICAL

Available Online at www.ijarcs.info

# ASL Based Hand Gesture Recognition System

Mr. Swapnil A. Bobade
Department of Information Technology
Sipna College of Engineering & Technology
Amravati, India
swapnil3105@gmail.com

Prof. Vijaya K. Shandilya
Associate Professor
Computer Sci. & Engineering Department
Sipna College of Engineering & Technology
Amravati, India

***Abstract:*** American Sign Language (ASL) is a well developed and standard way of communication for hearing impaired people living in English speaking communities. Since the advent of modern technology, different intelligent computer-aided application have been developed that can recognize hand gesture and hence translates gestures into understandable forms. In this proposed system, ASL based hand gesture system is presented that uses evolutionary programming technique called SIFT algorithm. Hand gestures images representing different English alphabets are used as an input to the system and then it is tested for a different set of images. The sign recognition accuracy obtained will be up to the mark.

***Keywords:*** ASL character, Hand gestures, SIFT Algorithm, Template matching, Key points.

## I. INTRODUCTION

### *A. American Sign Language:*

Sign languages are the most raw and natural form of languages could be dated back to as early as the advent of the human civilization, when the first theories of sign languages appeared in history. It has started even before the emergence of spoken languages. Since then the sign language has evolved and been adopted as an integral part of our day to day communication process. Now, sign languages are being used extensively in international sign use of deaf and dumb, in the world of sports, for religious practices and also at work places.

The recognition of gestures representing words and sentences as they do in American and Danish sign language undoubtedly represents the most difficult recognition problem of those applications mentioned before. A functioning sign language recognition system could provide an opportunity for the deaf to communicate with non-signing people without the need for an interpreter. It could be used to generate speech or text making the deaf more independent.

In this proposed system my aim is to develop a system which can recognize all static signs of the American Sign Language (ASL), all signs of ASL alphabets using bare hands. The user/signers are not required to wear any gloves or to use any devices to interact with the system. The idea is to make system that understand human sign language and develop a user friendly human computer interfaces (HCI).

## II. LITERATURE REVIEW / RELATED WORKS

### *A. Human Computer Interface System:*

Computer is used by many people either at their work or in their spare-time. Special input and output devices have been designed over the years with the purpose of easing the communication between computers and humans, the two most known are the keyboard and mouse [1]. Every new device can be seen as an attempt to make the computer more intelligent and making humans able to perform more complicated communication with the computer. This has been possible due to the result oriented efforts made by computer professionals for creating successful human computer interfaces [1]. As the complexities of human needs have turned into many folds and continues to grow so, the need for Complex programming ability and intuitiveness are critical attributes of computer programmers to survive in a competitive environment. The computer programmers have been incredibly successful in easing the communication between computers and human. With the emergence of every new product in the market; it attempts to ease the complexity of jobs performed. For instance, it has helped in facilitating tele operating, robotic use, better human control over complex work systems like cars, planes and monitoring systems. Earlier, Computer programmers were avoiding such kind of complex programs as the focus was more on speed than other modifiable features. However, a shift towards a user friendly environment has driven them to revisit the focus area [1].

The idea is to make computers understand human language and develop a user friendly human computer interfaces (HCI). Making a computer understand speech, facial expressions and human gestures are some steps towards it. Gestures are the non-verbally exchanged information. A person can perform innumerable gestures at a time. Since human gestures are perceived through vision, it is a subject of great interest for computer vision researchers. The propose system aims to determine human gestures by creating an HCI. Coding of these gestures into machine language demands a complex programming algorithm. An overview of gesture recognition system is given to gain knowledge [1].

Dr. S.D.Lokhande and Vaishali S. Kulkarni [2] designed a system visually recognize all static signs of the American Sign Language (ASL), all signs of ASL alphabets using bare hands. But, since different signers vary their hand shape size, body size, operation habit and so on, which bring more difficulties



CONFERENCE PAPER
"A National Level Conference on Recent Trends in Information Technology and Technical Symposium" On 09th March 2013
Organized by
Dept. of IT, Jawaharlal Darda Inst. Of Eng. & Tech., Yavatmal (MS), India

in recognition. Therefore, it realizes the necessity for signer independent sign language recognition to improve the system robustness and practicability in the future. The system gives the comparison of the three feature extraction methods used for ASL recognition and suggest a method based on recognition rate. It relies on presenting the gesture as a feature vector that is translation, rotation and scale invariant. The combination of the feature extraction method with excellent image processing and neural networks capabilities has led to the successful development of ASL recognition system using MATALAB. The system has two phases: the feature extraction phase and the classification. Images were prepared using portable document format (PDF) form so the system will deal with the images that have a uniform background. The feature extraction applied an image processing technique which involves algorithms to detect and isolate various desired portions of the digitized sign. During this phase, each colored image is resized and then converted from RGB to grayscale one. This is followed by an edge detection technique.

The goal of edge detection is to mark the points in an image at which the intensity changes sharply. Sharp changes in image properties usually reflect important invents and changes in world properties. The next important step is the application of proper feature extraction method and the next is the classification stage, a 3-layer, feed-forward back propagation neural network is constructed. The system is proved robust against changes in gesture. Using Histogram technique get the misclassified results. Hence Histogram technique is applicable to only small set of ASL alphabets or gestures which are completely different from each other. It does not work well for the large or all 26 number of set of ASL signs. For more set of sign gestures segmentation method is suggested. The main problem with this technique is how good differentiation one can achieve. This is mainly dependent upon the images but it comes down to the algorithm as well. It may be enhanced using other image processing technique like edge detection.

Md. Hasanuzzaman, S. M. Tareeq, M. A. Bhuiyan, Y. Shirai and H. Ueno [3] describes the Principal component analysis (PCA) method is a standard pattern recognition approach and many researchers use it for face and hand pose classification. The main idea of the principal component analysis (PCA) method is to find the vectors that best account for the distribution of target images within the entire image space. In the general PCA method, eigenvectors are calculated from training images that include all the poses or classes. But for classification a large number of hand poses for a large number of users, need large number of training datasets from which eigenvectors generation is tedious and may not be feasible for a personal computer. Considering these difficulties they have proposed person-specific subspace method that partitions the comparison area. In person-specific subspace method, ASL characters (hand poses) are grouped based on each person and for each person one PCA is used. They have described the algorithm of person-specific subspace method for ASL characters classification, which is very similar to general PCA based algorithm. To utilize person-centric knowledge, this system integrates person identification method with ASL recognition system. From the experimental results have concluded that, the classification accuracy of person-specific subspace method is better than general PCA method and person-specific subspace method is faster than the general PCA method.

Lalit Gupta and Suwei Ma [4] develop a complete system capable of robustly classifying hand gestures for HCI applications. From a visual analysis of hand gestures, it was determined that essential shape information for discriminating gestures was in the boundary of the gestures. Therefore, a contour and vision based classification approach was formulated. The relatively simple system consisted of a video camera, video capturing software, and a personal computer. For flexible operation, no constraint other than holding the gesture approximately parallel to the camera lens was imposed. The processing steps to classify a gesture included gesture acquisition, segmentation, morphological filtering, contour representation, and alignment based classification. Rather than forming an arbitrary set of gestures, the database for off-line evaluation consisted of the gestures for numbers 0 through 9 of the ASL. These gestures were selected because they are typical of the hand gestures that can be used for HCI. Nonlinear alignment is clearly the better choice for classifying hand gestures because of the inherent nonlinear distortions that can be expected in the contours of the gestures. The main goal was to show that robust classification of hand gestures is possible using the system developed. No attempt was made to optimize the code nor was any attempt made to use the fastest available processors. For the system to be successfully applied to the potentially numerous real-time HCI applications, the hand gestures should not only be classified accurately but must also be classified rapidly. It is estimated that if the system is developed using a high-speed PC with high-speed digital signal processing chips, images are read directly from a buffer, and the code for the processing steps is optimized, gestures can be classified fast enough to make real-time applications possible.

Miguel Lourenc¸o, Joao P. Barreto, and Francisco Vasconcelos[5] proposes modifications to the broadly used SIFT frame- work that make it resilient to image RD, while preserving the original invariance to scale, rotation, and moderate viewpoint change. The only assumptions are that the camera follows the division model and that the amount of distortion is coarsely known. They ran several experiments, both in synthetic and real frames, that prove the advantages of sRD-SIFT whenever there is significant image distortion. Their method provides significantly more correct point correspondences than the SIFT algorithm run after correcting the distortion via image warping. Comparing with the pSIFT algorithm, the gains in terms of number of matches are marginal, but sRD-SIFT has a higher accuracy in key point localization as proven by the experiments described in Sections V-B–V-D. These benefits are achieved at the expense of a small computational overhead when compared with the standard SIFT implementation. sRD-SIFT can be advantageous in several robot vision tasks, ranging from SfM to visual recognition, as well as in medical applications that rely in endoscopic imagery.

The SIFT keypoints described in this paper are particularly useful due to their distinctive- ness, which enables the correct match for a keypoint to be selected from a large database of other keypoints. This distinctiveness is achieved by assembling a high-dimensional vector representing the image gradients within a local region of the image. The keypoints have been shown to be invariant to image rotation and scale and robust across a substantial range of affine distortion, addition of noise, and change in illumination. Large numbers of keypoints can be extracted from typical images, which leads to robustness in extracting small objects among clutter. The fact that keypoints are detected over a complete range of scales means that small local features are available for matching small and highly occluded objects, while large keypoints perform well for images subject to noise and blur.

Their computation is efficient, so that several thousand keypoints can be extracted from a typical image with near real-time performance on standard PC hardware. This paper has also presented methods for using the keypoints for object recognition. The approach we have described uses approximate nearest-neighbor lookup, a Hough transform for identifying clusters that agree on object pose, least-squares pose determination, and fi- nal verification. Other potential applications include view matching for 3D reconstruction, motion tracking and segmentation, robot localization, image panorama assembly, epipolar calibration, and any others that require identification of matching locations between images. There are many directions for further research in deriving invariant and distinctive image features. Systematic testing is needed on data sets with full 3D viewpoint and illumination changes. The features described in this paper use only a monochrome intensity image, so fur- ther distinctiveness could be derived from including illumination-invariant color descriptors (Funt and Finlayson, 1995; Brown and Lowe, 2002). Similarly, local texture measures appear to play an important role in human vision and could be incorporated into feature descriptors in a more general form than the single spatial frequency used by the current descriptors. An attractive aspect of the invariant local feature approach to matching is that there is no need to select just one feature type, and the best results are likely to be obtained by using many different features, all of which can contribute useful matches and improve overall robustness.

## III. ANALYSIS OF PROBLEM

Automatic detection of text from natural scenes is a very difficult task. The primary challenge lies in variations of text: it can vary in font, size, orientation, and position of text, be blurred from motion, and occluded by other objects. The layout of a sign is usually language dependent. My objective is to automatically detect text from hand gesture for a recognition and translation task. In a sign recognition and translation task, an image of text can be deformed because of an inappropriate angle. Such deformation can be, in fact, recovered by measuring the key points of an image. The basic idea is as follows: I will provide acquired image as an input to the system it may be a captured or static images. I then find the separate key points of an image using SIFT algorithm. Finally, I will match the key points of static image with every image in database and calculate matching score for comparison.

## IV. PROPOSED WORK

Vision based analysis, is based on the way human beings perceive information about their surroundings, yet it is probably the most difficult to implement in a satisfactory way. Several different approaches have been tested so far.

a. One is to build a three-dimensional model of the human hand. The model is matched to images of the hand by one or more cameras, and parameters corresponding to palm orientation and joint angles are estimated. These parameters are then used to perform gesture classification.
b. Second one to capture the image using a camera then extract some feature and those features are used as input in a swift algorithm for matching key points of image with the images available in the database.

In this proposed system, I have used second method for modeling the system and taken the static images database from standard hand gesture database which is input to the system.

Following Algorithm shows the overall flow of my proposed system for ASL recognition:

Step 1: Start
Step 2: Image Acquisition (captured or static images).
Step 3: Find the separate key point of static image using SIFT algorithm.
Step 4: Matching the key points of static image with every image in database and calculate matching score i.e. Threshold value.
Step 5: Matching score will be declared whether pattern is found or not?
Step 6: If image found with threshold value > 75%
Step 7: It will be declared as proper output character
Step 8: It may happen two or more images have threshold value > 75%
Step 9: Compare all those images, Maximum value will consider as a result
Step 10: Stop.

### *A. SIFT Algorithm:*

For any object in an image, interesting points on the object can be extracted to provide a "feature description" of the object. This description, extracted from a training image, can then be used to identify the object when attempting to locate the object in a test image containing many other objects. To perform reliable recognition, it is important that the features extracted from the training image be detectable even under changes in image scale, noise and illumination. Such points usually lie on high-contrast regions of the image, such as object edges.

Another important characteristic of these features is that the relative positions between them in the original scene shouldn't change from one image to another. For example, if only the four corners of a door were used as features, they would work regardless of the door's position; but if points in the frame were also used, the recognition would fail if the door is opened or closed. Similarly, features located in articulated or flexible objects would typically not work if any change in their internal geometry happens between two images in the set being processed. However, in practice SIFT detects and uses a much larger number of features from the images, which reduces the contribution of the errors caused by these local variations in the average error of all feature matching errors [7].

Here's an outline of what happens in SIFT.

***a. Constructing a scale space:***

This is the initial preparation. You create internal representations of the original image to ensure scale invariance. This is done by generating a "scale space".

***b. LoG Approximation:***

The Laplacian of Gaussian is great for finding interesting points (or key points) in an image. But it's computationally expensive. So we cheat and approximate it using the representation created earlier.

***c. Finding keypoints:***

With the super fast approximation, we now try to find key points. These are maxima and minima in the Difference of Gaussian image we calculate in step 2

***d. Get rid of bad key points***

Edges and low contrast regions are bad keypoints. Eliminating these makes the algorithm efficient and robust. A technique similar to the Harris Corner Detector is used here.

***e. Assigning an orientation to the keypoints:***

An orientation is calculated for each key point. Any further calculations are done relative to this orientation. This effectively cancels out the effect of orientation, making it rotation invariant.

***f. Generate SIFT features:***

Finally, with scale and rotation invariance in place, one more representation is generated. This helps uniquely identify features. Lets say you have 50,000 features. With this representation, you can easily identify the feature you're looking for (say, a particular eye, or a sign board) [7].

## V. APPLICATION

a. Deaf or dumb people can able to interact with the computer system through sign language.
b. User friendly human computer interfaces (HCI).

## VI. DATABASE DESCRIPTION

The American Sign Language (ASL) Alphabet and a description of (ASL) in Text. All letters are signed using only the right hand which is raise with the palm facing the viewer so a straight finger is will normally point upwards. When fingers are folded they point down across the palm. When the thumb is folded it crosses the palm towards the little finger. In these descriptions left and right are from the position of the viewer. In the case where the hand is turned or tilted the positions of the fingers is described first for an upright hand and the turn or tilt is added.

a. A closed fist, all finger folded against the palm, thumb is straight, alongside the index finger.
b. All fingers are straight. Thumb is folded across palm.
c. All fingers partially folded. Thumb is partially folded. Hand is turn slightly to the left so viewer can see backward "C" shape formed by thumb and index finger.
d. Middle, ring and little fingers are partially folded. Tip of thumb is touching tip of middle finger. Index finger is straight. Hand is turned slightly so viewer can see "d" shape formed by thumb, middle and index fingers.
e. Thumb is folded across in front of palm but not touching it. All fingers are partially folded with the tips of index, middle and ring fingers touching the thumb between the knuckle and the tip.
f. Tip of index finger is touching tip of thumb. Middle, ring and little fingers are straight and slightly spread.
g. Middle, ring and little fingers are folded down across palm. Thumb is straight but pulled in so that it is in front of the index finger. The index finger is straight and pointing forwards slightly so that it is parallel to the thumb, The thumb and index finger are not touching. The whole hand is turned towards the left and tilted slightly so the thumb and index finger are towards the viewer and pointing up at about 45 degrees.
h. Ring and little finger are folded down. Thumb is folded over ring and little finger. Index finger and middle finger are straight and together. The hand is tilted over so that the fingers are horizontal and pointing to the left.
i. Index, middle and ring fingers are folded down. Thumb is folded across index middle and ring fingers. Little finger is straight.
j. Index, middle and ring fingers are folded down. Thumb is folded across index middle and ring fingers. Little finger is straight. The hand is moved so that little finger draws a "J" shape. Motion is a curve moving forward and then right. The hand turns to the right.
k. Ring and little fingers are folded down. Index and middle finger are straight and slightly spread. Thumb is straight and pointing up to the middle finger. (This is very similar to V the only difference is the position of the thumb.
l. Middle, ring and little finger are folded down over palm. Index finger and thumb are straight. Thumb is sticking out sideways at 90 degrees to index finger to form "L" shape.
m. Little finger is folded. Thumb is folded across to touch little finger. Index, middle and ring fingers are folded down over thumb.
n. Little and ring finger are folded. Thumb is folded across ring and little finger. Index finger and middle finger are folded down over thumb.
o. All fingers are partially folded. Thumb is partially folded and tip of thumb is touching tip of index finger. Hand is turned slightly so viewer can see "O" shape

formed by thumb and index finger.

p. Ring and little finger are folded down. Index finger is straight. Middle finger is straight but pointing forward so that is at 90 degrees to index finger. Tip of thumb is touching middle of middle finger. Hand is turned to the left and twisted over so that index finger is horizontal and middle finger is pointing down. Viewer can (sort of) see a "P" shape formed by middle finger and thumb.

q. Ring and little fingers are folded down across palm. Thumb is straight but pulled in so that it is in front of the index finger. The index finger is straight and pointing forwards slightly so that it is parallel to the thumb. The index finger and thumb are not touching. The Middle finger is bent down and across to the right of the thumb (this hurts !). The whole hand is turned towards the left and tilted so the thumb and index finger are towards the viewer and pointing almost straight down.

r. Ring and little finger are folded against the palm, held down by thumb, index and middle finger are straight and crossed, index finger in front.

s. Clenched fist. All fingers folded tightly into palm. Thumb is across index and middle fingers.

t. Middle, ring and little fingers are fold down across palm. Thumb is folded across middle finger. Index finger is folded over thumb.

u. Ring and little finger are folded against the palm, held down by thumb, index and middle finger are straight and together.

v. Ring and little finger fold against the palm, held down by thumb, index and middle finger are straight and spread to form a "V" shape.

w. Tip of little finger is touching tip of thumb. Index, middle and ring fingers are straight and slightly spread.

x. Middle, ring and little fingers are folded down. Index finger is bent at both joints. Thumb is pulled in and slightly bent at the joint. The hand is turned to the left so view can see thumb and index finger.

y. Index, middle and ring ringers folded against palm. Little finger and thumb are straight and spread wide.

z. Middle, ring and little fingers are folded. Thumb is folded across middle and ring fingers. Index finger is straight. The hand is moved so that the tip of index finger draws out a "Z" shape. The motion is (1) from right to left. (2) from left to right and forward. (3) from right to left [8].

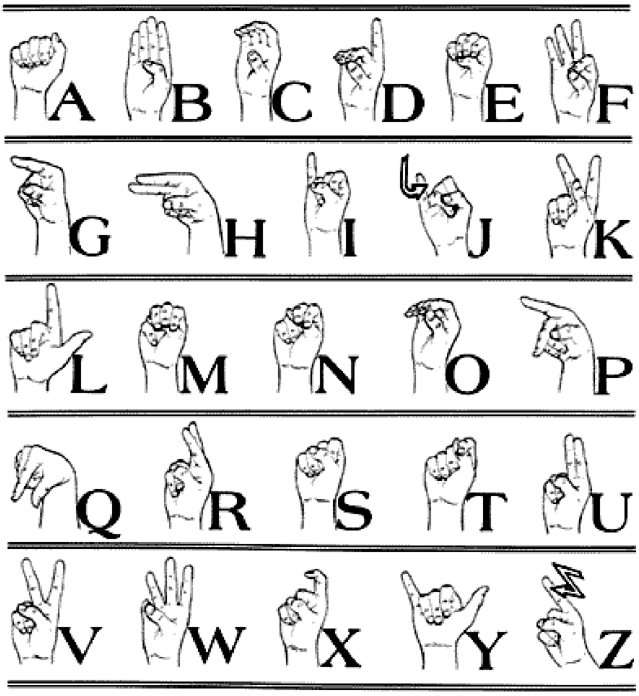


Figure:1: American Sign Language

## VII. CONCLUSION

In this proposed system my aim is to develop a system which can recognize all static signs of the American Sign Language (ASL), all signs of ASL alphabets using bare hands. The user/signers are not required to wear any gloves or to use any devices to interact with the system. The idea is to make system that understand human sign language and develop a user friendly human computer interfaces (HCI). This paper has also presented methods for using the keypoints for object recognition. The SIFT keypoints described in this paper are particularly useful due to their distinctiveness, which enables the correct match for a keypoint to be selected from a large database of other keypoints. This distinctiveness is achieved by assembling a high-dimensional vector representing the image gradients within a local region of the image. Large numbers of keypoints can be extracted from typical images, which leads to robustness in extracting small objects among clutter. Our hope is to make this sign language recognition system more robust and capable to recognize dynamic gestures for interaction with different intelligent machine.

## VIII. REFERENCES

[1]. A Thesis by Sanjay Meena "A Study on Hand Gesture Recognition Technique", National Institute of Technology, Rourkela Orissa, INDIA 2011

[2]. Vaishal.S.Kulkarni, “Appearance Based Recognition of American Sign Language Using Gesture Segmentation” et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 03, 2010, 560-565

[3]. Md. Hasanuzzaman, S. M. Tareeq, M. A. Bhuiyan, Y. Shirai and H. Ueno “ASL Characters Recognition Using Person

 

Specific Subspace Method" Daffodil International journal of Science and Technology, Vol. 4, Issue2, July 2009

[4]. Lalit Gupta and Suwei Ma IEEE "Gesture-Based Interaction and Communication: Automated Classification of Hand Gesture Contours Transactions on Systems", MAN and Cybernetics—PART C: Applications and Reviews, VOL. 31, No. 1, FEB 2001

[5]. Miguel Lourenc¸o, Joao P. Barreto, and Francisco Vasconcelos "sRD-SIFT: Keypoint Detection and Matching in Images with Radial Distortion" IEEE TRANSACTIONS ON ROBOTICS, VOL. 28, NO. 3, JUNE 2012

[6]. David G. Lowe "Distinctive Image Features from Scale-Invariant Keypoints" Computer Science Department University of British Columbia Vancouver, B.C., Canada lowe@cs.ubc.ca January 5, 2004

[7]. www.cs.st-andrews.ac.uk/~yumeng/yumeng-SIFTreport-5.18_bpt.pdf

[8]. sandlersignlab.haifa.ac.il/wendy.html