

Survey on Hand Gesture Recognition

Padmavati S
Professor, Dept. of CSE,
Amrita School of Engineering
Coimbatore

Kiruthika S
Student, Dept. of CSE,
Amrita School of Engineering
Coimbatore

Navin Kumar P I
Student, Dept. of CSE,
Amrita School of Engineering
Coimbatore

Abstract- The main goal of gesture recognition is to create a system which can recognize specific human gestures and use them to convey information. Types of techniques used in recognizing hand postures and gestures are compared along with advantages and disadvantages of each. Gestures considered as a natural way of communication among human, since it is a physical movement of hands, arms, or body which conveying meaningful information and helps in expressing thoughts and feelings effectively.

I.INTRODUCTION

Gestures are meaningful body motions involving Physical movements of the fingers, hands, arms, head, face, or body with the intent of: 1) conveying meaningful information 2) interacting with the environment. There have been varied approaches to handle gesture recognition [1], ranging from mathematical models based on hidden Markov chains [2] to tools or approaches based on soft computing [3]. Gestures can be or dynamic. Some gestures also have both static and dynamic elements, as in sign languages. Again, the automatic recognition of natural continuous gestures requires their temporal segmentation. Sometimes one needs to specify the start and end points of a gesture in terms of the frames of movement, both in time and in space. 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. Gestures are one of the first forms of communication when a child learns to express its need for food, warmth and comfort. It enhances the emphasis of spoken language. Hand gesture recognition system many challenges are Variation of illumination conditions, Rotation problem, Background problem, Translation problem. The paper is organized as follows : Section (I) Introduces the process of gesture recognition. Section (II) explains Hand Gesture Recognition(HGR) using Scale Invariant Feature Transform(SIFT). Section (III) explains the Neural Networks for Hand Gesture Recognition. Section (IV) gives explanation on Hand Gesture Recognition using Template Matching and the Section(V) explains about Principal Component Analysis(PCA).

II.HGR USING SIFT

Scale invariant feature transform(SIFT) presents a method for detecting distinctive invariant features from images that can be later used to perform reliable matching between different views of an object or scene. Two key concepts are used in this definition: distinctive invariant features and reliable matching. SIFT[5] is broken down into four major computational stages: a) Scale-Space extrema detection b) Keypoint localization c) Orientation assignment d) Keypoint descriptor

In the first phase the key points from various orientations are obtained. The Scale Space Function is available for this phase.

$$(x, y, \sigma) = G(x, y, k\sigma) - G(x, y, \sigma) * I(x, y) \quad \text{Eq(1)}$$

where $I(x, y)$ is the Input Image, $G(x, y, \sigma)$ is the variable Gaussian Scale. Scale Space Extrema (x, y, σ) is found by difference between two images, one with k times the other.

To obtain the local minima and maxima we compare each key point with eight neighbors on the same scale and nine neighbors on scales above and below it. The next phase involves fitting of the key points in the three degree quadratic functions. This function is obtained from the second order Taylor Expansion. The local extrema with lower contrasts will not take into consideration because they are sensitive to noise. The key points below the threshold level are also not taken in account by the system. The next phase involves the orientation assignment where the main orientation is assigned to each feature based on local image gradient. For each pixel around the key point the gradient magnitude and the orientation are computed. The following equations are used to find the magnitude and orientation.

$$(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad \text{Eq. (2)}$$

The above equation is used to calculate the magnitude (x, y) of the detected key points.

$$\Theta(x, y) = \tan^{-1}(L(x, y+1) - L(x, y-1) / L(x+1, y) - L(x-1, y)) \quad \text{Eq. (3)}$$

In the above equation, $\Theta(x, y)$ gives the orientation of the key point. Eq (3) is to calculate the orientation of the detected key points. The orientation and magnitude of the key points are stored and used in the further process of Gesture Classification. The next phase of gesture recognition is the key point descriptor phase. In this the local image gradients are measured at a selected scale around each key point. The region around the keypoint is divided into 4×4 . The gradient magnitude and orientation within each box is computed. For

each box 8 bin orientation histogram is obtained. From the 16 obtained orientation histogram 128 dimensional SIFT descriptor is built. This descriptor is orientation invariant, because it is calculated relative to main orientation. The SIFT descriptor is normalized to unit length to achieve illumination invariance.

III.HGR USING NEURAL NETWORKS

This method has five steps of processing [10]
 1. Cropping input image 2. Resizing image 3.peak and valley detection 4. Dividing image 5. Training

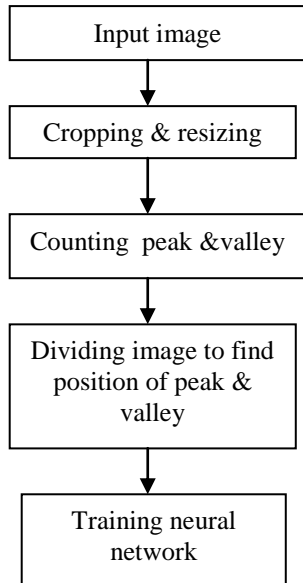


Fig : 1 flow diagram

Filtered image is then resized to 256*256 sized image. Hand portion image is then converted to 256*256 size RGB image, this way hand portion comes at the center of image. Boundary Tracing for Peaks and Valleys Using morphological operations this smoothed image is converted to boundary image. After getting boundary image we first find the boundary tracing point from where to start and where to stop finding peaks and valleys. For this we find maximum value of x where white pixel exists. We call this point as $opti_x$ and then find corresponding value of y. The starting point is on x direction as $0.80*x$. From this x value we find starting y co-ordinate of starting point. This is our starting point to trace boundary and ending point is starting point y position plus one i.e. next row of starting point where white pixel exist. *Condition I:* we start with UP=1, we first travel to top and check whether white pixel exist or not. If exist then continue in same way if not we check it on top left or top right. Again we search on top side and continue until we don't get any pixel on top or top-left or top-right.

0	0	0	0	0	1	0
0	0	0	0	0	1	0
0	0	0	0	1	0	0
0	0	0	0	1	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0
0	0	0	1	0	0	0

Fig:2 condition I

Condition II: If we don't get any pixel it means we have to search on existing pixels right side, if pixel exist we follows the same way until we get no pixel on right side. We again follows as per condition I. if condition I and II not satisfied it means we have to search down, here we mark as peak If condition I and II not satisfied then we search on down side by making DN=1

0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	0	0
0	0	0	1	0	0	0	0	0
0	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0
0	1		0	0	0	0	0	0

Fig:3 condition II

Condition III: we start with DN=1, we first travel to down and check whether white pixel exist or not. If exist then continue in same way if not we check it on down left or downright. Again we search on down side and continue until we don't get any pixel on down or down - left or down -right.

0	0	0	0	0	0
0	0	0	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	0	0	0	0	1
0	0	0	0	0	0

Fig :4 condition III

Condition IV: If we don't get any pixel it means we have to search on existing pixels right side, if pixel exist we follows the same way until we get no pixel on right side and then we follows condition III. If in condition IV there is no pixel on right side we search on existing pixels left side, if pixel exist we follows the same way until we get no pixel on left side and

then we follows condition III. If condition III and IV not satisfied it means we have to search on top side, here we mark as valley.

0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0
0	0	1	0	0	0	1	1
0	0	0	1	1	1	0	0
0	0	0	0	0	0	0	0

Fig 5: condition IV

Feature Extraction Image is then divided in to 16 parts. From the divided image we find other parameters like in which part the highest peak has been detected in an image and which areas have been occupied by peaks and valleys. Using these parameters neural network is trained.

IV.HGR USING TEMPLATE MATCHING

Template matching a fundamental pattern recognition technique has been utilized in the gesture recognition. Template matching is performed by the pixel-by-pixel comparison. Because of the pixel-by-pixel image comparison template matching is not invariant to scaling and rotation. It is sensitive to noise and occlusion. We first compute the difference between the image and the previous one, and convert it to grayscale. The grayscale map is obtained by adding the contribution of each channel of the RGB original frame to the difference image. The map is obtained by differencing the grayscale image and the reference throughout the grayscale image[16]. The Sum of Absolute Difference (SAD) described in Eq. (4) is used to create the correlation map.

$$Dist_{Img \rightarrow Ref}(x,y) = \sum_{x'=0}^m \sum_{y'=0}^n (Img(x+x', y+y') - Ref(x'+y'))^2 \quad Eq(4)$$

$$0 \leq x \leq M-m-1$$

$$0 \leq y \leq N-n-1$$

Where Img is the residual image resulting from the difference of two consecutive frames, Ref is the reference, $(x; y)$ are the coordinates of the destination and $(x_0; y_0)$ the coordinates of the reference, M and N are the width and height of Img and m and n are the width and height of Ref . We obtain a residual map of size $(M - m - 1) \times (N - n - 1)$. We select the maximum correlation and compare it to a threshold. If the threshold is reached, we can accurately predict that the hand is at the estimated position. The threshold is determined using the following procedure. The distance between a reference and itself is computed using Eq. (1) where $Img = Ref$. We obtain a single value since $M = m$ and $N = n$. For each the threshold is set as a fixed percentage of this value in Eq. (5).

$$Threshold = \alpha \times Dist_{ref \leftrightarrow ref}(0,0) \quad Eq(5)$$

A good value for α as been determined as $25\% \leq \alpha \leq 35\%$. Higher threshold would create too many false negatives

whereas a lower threshold would create too many false positives. The major drawback of this method is inherent in all pattern matching methods is the processing time is very long for one pattern.

V.HGR USING PCA

Principal Component Analysis(PCA) is basically dimension reduction method in which the significant eigen vectors based on eigen values are used to project the data[20]. This approach captures the significant variability in the data and thus can be used to identify the gestures and postures in the vision based system. Though this method can be exploited for glove based approaches also, but till that time only vision based techniques have been exploited. This method suffers from a drawback that there should be variance in the at least one direction. If variance is uniformly distributed in the data, it will not yield the relevant Principle vectors also, if there is noise, PCA would consider it as a significant bias too. Besides this method suffers from scaling in hand size and position, which can be taken care by normalization. Even then, this method is user dependent. PCA methods require an initial training stage, in which a set of images of similar content is processed. Typically, the intensity values of each image are considered as values of a 1D vector, whose dimensionality is equal to the number of pixels in the image; it is assumed, that all images are of equal size. For each such set, some basis vectors are constructed that can be used to approximate any of the training images in the set. In the case of gesture recognition, the training set contains images of hands in certain postures. The above process is performed for each posture which the system should later be able to recognize. In PCA-based gesture recognition, the matching combination of principal components indicates the matching gesture as well. This is because the matching combination is one of the representatives of the set of gestures that were clustered together in training, as expressions of the same gesture. A problem of eigen space reconstruction methods is that they are not invariant to image transformations such as translation, scaling, and rotation.

VI.CONCLUSION

In SIFT feature which are computed are invariant to scaling and rotation. This features enable correct match for key points between different hand gestures. By using neural networks for recognition we have extracted simple features form images and network is trained using Support Vector Machine. The accuracy obtained in this work is 100 % as only few signs have been considered. In Template matching the drawback is it finds the local maximum on measure on match surface sometimes. This technique requires a separate template for each scale and orientation. Template matching become thus too expensive, especially for large templates. PCA analysis has given us the original images in terms of the differences and similarities between them. The PCA analysis has identified the statistical patterns in the data. Since all the vectors are N^2 dimensional, we will get N^2 eigenvectors. In practice, we are able to leave out some of the less significant eigenvectors, and the recognition still performs well.

	Translation Invariance	Rotation Invariance	Scale Invariance
SIFT	Yes	Yes	Yes
Neural network	No	No	No
Template Matching	Yes	No	No
PCA	No	No	No

Table I: Properties of various Feature Extraction Techniques

REFERENCES

- C. L. Lisetti and D. J. Schiano, "Automatic classification of single facial images," *Pragmatics Cogn.*, vol. 8,2000.
- V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human computer interaction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, Jul. 1997.
- L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, Feb. 1989.
- S. Mitra and T. Acharya, *Data Mining: Multimedia, Soft Computing, and Bioinformatics*. New York: Wiley, 2003.
- S.Pandita,S.P Narote "Hand Gesture recognition using SIFT "International Journal of Engineering Research & Technology,vol.2,Issue 1, Jan 2013
- David G.Lowe, "Object recognition from local scale invariant feature" International conference on computer vision" Sep 1999.
- Sven Siggelkow,"Feature Histogram for Content-Based Image Retrieval".PhD Thesis,Albert-Ludwigs UniversityFrieiburg,Dec 2002.
- David G. Lowe "Distinctive Image Features from Scale-Invariant Keypoints". International Journal of Computer Vision, 60,2,(2004),pp.91-110.
- Pallavi Gurjal, Kiran Kunnur, " Real Time hand gesture Recognition using SIFT", International Journal of Electronics and Electrical Engineering, Volume 2 Issue 3 ,March 2012.
- Rajesh Mapari,Dr. Govind Kharat"Hand gesture recognition using neural network" International Journal of Computer science and Networks, vol 1, issue 6, Dec 2012.
- Er. Aditi Kalsh, Dr. N.S. Garewal, " Sign Language Recognition System", International Journal of Computational Engineering Research,Vol, 03,Issue, 6,June 2013.
- Arnaud J. Bernard," Human Computer Interface Based On Hand Gesture Recognition" A Thesis for the Degree Master of Science in the School of Electrical and Computer Engineering, Georgia Institute of Technology December 2.
- Jun-Ki Min, Bongwhan Choe, and Sung-Bae Cho, "A Selective Template Matching Algorithm for Short and Intuitive Gesture UI of Accelerometer-Built-in Mobile Phones", Proceedings of the World Congress on Nature and Biologically Inspired Computing (NaBIC2010).
- Chai, X., Fang, Y., and Wang, K., "Robust hand gesture analysis and application in gallery browsing," in *Multimedia and Expo, 2009. ICME 2009*. IEEE International Conference on, pp. 938 -941, June 2009.
- Freeman, W. and Weissman, C., "Television control by hand gestures," in *Proc.Int'l Workshop on Automatic Face and Gesture Recognition*, pp. 179-183, June 1995.
- G.S.Cox,"Template matching and measure of match in Image Processing",University of Cape town, July 12,1995.
- R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*. J. Wiley & Sons,2009.
- R. Brunelli and T. Poggio, 1997, *Template matching: Matched spatial filters and beyond*. *Pattern Recognition* 30, 751-768.
- Lars Bretzner, Ivan Laptev, Toney Lindberg, *On Hand gesture recognition using multiscale colour features hierarchiel models and partial filtering*, CVAP Laboratory, Department of numerical analysis and Computer Science, Sweden,2002.
- Vaishali S Kulkarani, ME Digital Systems, S. D. Lokhande, "Appearance based segmentation of sign language using gesture segmentation", Sinhgad College of Engineering, 2010.
- Jashan M, Madhusoodhan H V, "Hand Gesture Recognition System Using Principal Component Analysis and its Applications" *Electronics and Communication, SIT, Tumkur, Karnataka, India.*
- Nasser H. Dardas and Emil M, "Hand Gesture Detection and Recognition Using Principal Component Analysis", *Petriu School of Information Technology and Engineering, Discover Lab. University of Ottawa Ottawa, Canada.*
- Sauvik Das Gupta, Souvik Kundu, Rick Pandey, " Hand Gesture recognition and classification by Discriminant and Principal Component Analysis using Machine Learning techniques", *International Journal of Advanced Research in Artificial Intelligence*, Vol. 1, No. 9, 201.
- Raheja J.L., Shyam R., Kumar U., Prasad P.B., "Real-Time Robotic Hand Control using Hand Gesture", 2nd international conference on Machine Learning and Computing, 9-11 Feb, 2010, Bangalore, India, pp. 12-16 .
- Jolliffe, I, *Principal component analysis*, Springer Verlag, New York,1986.
- Jolliffe, I. T. & Uddin, M.'A modified principal component technique based on the lasso',*Journal of Computational and Graphical Statistics* ,2003,12, 531-547.