Research Article Impact Factor: 4.226 ISSN: 2319-507X
Anagha P. Dhote, IJPRET, 2016; Volume 4 (9): 667-679 IJPRET

INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

REAL TIME HAND TRACKING AND GESTURE RECOGNITION SYSTEM

MS. ANAGHA P. DHOTE

Dept. of Computer Science Engineering, Prof. Ram Meghe College of Engineering and Management, Bandera – 444701

Accepted Date: 15/03/2016; Published Date: 01/05/2016

Abstract*:* A gesture is a communicative movement of human body parts having a specific message to be communicated by a receiver. Gesture recognition is an important for developing alternative human-computer interaction. It enables human to interface with machine in more natural way. Gesture recognition helps to understand the meaning of human body movement that is movement of parts, which involves the movement of hand, head, arms, face or body. It presents a method for tracking hand gesture and recognizing hand gestures by extracting unique invariant features from gestures for Marathi Sigh Language (MSL). The extracted feature is used to perform effective matching between different observations of a hand gesture.

Keywords: Indian Sign Language (ISL), Marathi Sign Language (MSL), Keypoints, Scale Invariant Feature Transform (SIFT)

PAPER-QR CODE

Corresponding Author: MS. ANAGHA P. DHOTE

Access Online On:

www.ijpret.com

How to Cite This Article:

Anagha P. Dhote, IJPRET, 2016; Volume 4 (9): 667-679

INTRODUCTION

Hand gesture is an effective and powerful way of communication. Human-Human communication is generally done by using hand gesture. It is most natural thing which is performing automatically when communicates. Lots of research has been on Hand Gesture Recognition Technology. In earlier days people using Keyboard and mouse to interact with the computer. But because of time limitations and increasing speed of life there will be limitations on using these devices so that hand gesture technology is being need for future of computer. So by using this technology a person can operate computer without touching it [1]. It provides a separate balancing modality to speech for expressing one's ideas. By using hand gestures for communication, a more natural interaction between humans and computing devices became so flexible and convenient for human being.

Hand gesture is considered into two types:

1. Static Hand Gesture: Static hand gestures are predefined and fixed gesture. Static gesture deals with one frame at a time. The information used for static gesture may include template or posture. Well known example of these is Indian sign language.

2. Dynamic Hand Gesture: Dynamic hand gestures are the real time gesture. It is also called motion-based gesture as it is sequence of images. This sequence containing large number of frames which is used for extracting motion information.

Gesture recognition is becoming an increasingly important for many applications such as human machine interfaces, security, and communication, multimedia. It provides a platform to express thoughts without speech. The gestures are tracked by using Skin Detection Algorithm and Scale Invariant Feature Transform (SIFT) extract features from the gesture. Using these features it performs gesture recognition. The methodology of SIFT feature detection given by Lowe et.al. [9], which is used for object recognition. As described in paper [9], the invariant features extraction of gesture is done by performing the steady matching between different views of an object. The method has the ability to identify large numbers of features and it is effective for feature extraction.

A different approach for hand gesture recognition is developed, other than discussed earlier. It describes a single-hand gesture recognition system for making meaningful sentences using scale invariant feature transform method.

The main objective is to track the ISL gesture in real time, recognizing ISL tracked gesture and gives the output in audio form.

- **literature review**

Table I. Literature Review in Nutshell for IEEE Papers

Year	Title	Authors	Evaluation Approach	Keywords
2014	Dynamic Hand Gesture Recognition and Detection for Real Time Using Human Computer Interaction	Momin Rijwan Ramjan, Rane Mithilesh Sandip, Phatangare Sushant Uttam, Wani Sumit Srimant	It proposed a system "Eyes On YOU	Blurring Algorithm
2013	Fast SIFT Design for Real-Time Visual Feature Extraction	Liang-Chi Chiu, Tian-Sheuan Chang, Jiun-Yen Chen and Nelson Yen-Chung Chang	It proposes a layer parallel SIFT with integral image, and its parallel hardware design with an on-the fly feature extraction flow for real-time application needs.	SIFT, VLSI design
2012	Real Time Hand Gesture Recognition using SIFT	Pallavi Gurjal, Kiran Kunnur	It decodes a gesture video into the appropriate alphabets.	SIFT
2012	Fast SIFT Algorithm based on Sobel Edge Detector	Yang Li, Lingshan Liu, Lianghao Wang, Dongxiao Li, Ming Zhang.	It proposes a fast SIFT algorithm based on Sobel edge detector.	SIFT, Image Matching
2012	Image Matching	Zheng-Jiandingt,	Image matching	SIFT, Gaussian

	of Gaussian Blurred Image Based on SIFT Algorithm.	Yang Zhangt, A-Qing Yangt, Dai-Li2.	process of smoothing is done in which the edge points and the pixels whose gray value changed largely are also smoothed, then leading to the number of feature points reduced.	Blurring
2011	An Improved SIFT Algorithm for Image Feature-Matching	Wenyu Chen, Yanli Zhao, Wenzhi Xie, Nan Sang	It solve the disadvantages of large computation reduces the dimension of the feature vector and reduces the mismatching of SIFT algorithm	SIFT, Real time dimension
2008	Vedio Object Matching Based on SIFT Algorithm	Xuelong H, Yingcheng Tang, Zhenghua Zhang	SIFT is used to solve visual tracking problem,	Feature matching, Video object retrieval, Video object tracking, SIFT
2005	Object Recognition from Local Scale-Invariant Features.	David G. Lowe	It shows that robust object recognition can be achieved in cluttered partially-occluded images with a computation time of	SIFT

			under 2 seconds.	
2004	Distinctive Image Features from Scale-Invariant Keypoints	David G. Lowe	It presents a method for extracting distinctive invariant features from images that can be used to perform reliable matching between different views of an object or scene.	SIFT

- Proposed Method

The paper proposed the work on hand tracking and gesture recognition for ISL sentences. From web cam video frames in real time the gestures are tracked using Skin Detection Algorithm and derived Scale Invariant Feature Transform (SIFT) features Further, the features are used for gesture recognition. The default threshold value for the algorithm is 0.035. But the algorithm works better in 0.025 threshold value. The block diagram for proposed work is shown in Figure 1.

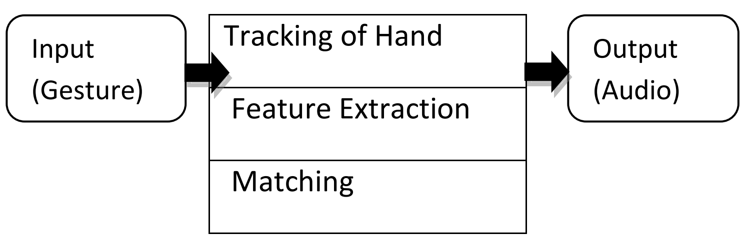

Figure 1. Block Diagram of System

The individual block functions as follows:

i. Input: The input gesture is captured by camera. The camera used is Logitech HD Webcam C270. This webcam displays in 1280x720 max resolutions and 30fps (generally 3.0 megapixels). The higher the mega-pixel, the better the picture the higher the resolution it can capture.

ii. Tracking of Hand: In this step, the input gesture is tracked by skin detection algorithm.

iii. Feature Extraction: The SIFT algorithm is used for feature extraction. It extract the keypoints of the gesture. In SIFT algorithm, four steps are given for extracting the keypoints of gesture which is explain later.

iv.Matching: The extracted keypoints from the gesture are used for matching. If the stored keypoints of gesture are matched with the real time gesture, it recognized the gesture.

v.Output : Finally, it gives the output in audio form.

- *Skin Detection*

Skin detection is the process of finding skin-colored pixels and regions in an image or a video. This process is typically used as a preprocessing step to find regions that potentially have human gesture in images. Several computer vision approaches have been developed for skin detection. A skin detector typically transforms a given pixel into an appropriate color space and then uses a skin classifier to label the pixel whether it is a skin or a non-skin pixel.

There are two types of Skin Detection method.

1) Pixel Based Method: Classify each pixel as skin or non-skin individually, independently from its neighbors.

2) Region Based Method: It tries to take the spatial arrangement of skin pixels into account during the detection stage to enhance the methods performance.

1. Pixel reading

In this, the data send by camera are interpreted by pixel. Each pixel contains the R, G, B value. The R, G, B components are separate out using following formula:

$$r = \frac{R}{R+G+B} \qquad (1)$$

$$g = \frac{G}{R + G + B} \quad (2)$$

$$b = \frac{B}{R + G + B} \quad (3)$$

Figure 2. Original Image

Figure 3. RGB Image

2. RGB –YCgCr Conversion

In this Operation, the Red, Green, Blue pixel representation of the RGB scheme are change into the corresponding YCgCr representation using following equations and shown in Figure 4.

$$Y = 16 + 65.481 * R + 128.553 * G + 24.966 * B \quad (4)$$

$$C_g = 128 - 81.085 * R + 112 * G - 30.195 * B \quad (5)$$

$$C_r = 128 + 112 * R - 93.784 * G - 18.214 * B \quad (6)$$

Figure 4. YCgCr Image

3. Image Segmentation

There are many methods available for skin segmentation like HSV model, HSI model, YCbCr model, YCbCr model. Here, we refer YCgCr model because of following reason:

- It is robust for varying light condition.
- It has sharp color map.
- Just changing the value of Y component, we can compensate light effect.

Using YCgCr model, the image is get converted into binary representation as shown in Figure 5. If value=1, the skin color becomes white and if value=0, rest of images become black. The threshold value used for segmentation is:

F(x,y)= 1, [Y(x,y)>60 &&100<Cg(x,y) <130 &&135<Cr(x,y)<175]

0, Otherwise

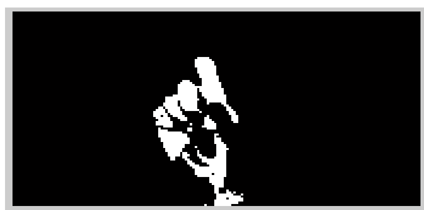

Figure 5. Skin Segmented Image

4. Image Filtering and Reduction

After segmentation, the noise still present in the image is filtered with dilation and erosion method. Then it gives the cropped image as shown in Figure 6.

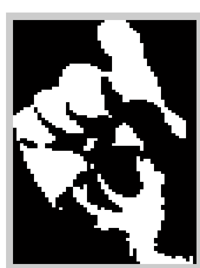

Figure 6. Cropped Image

B. Gesture Keypoint Recognition

David Lowe developed the Scale Invariant Feature Transform (SIFT) algorithm [8]. SIFT algorithm is used to detect distinct features of gesture. Computation of SIFT gesture features is performed through the four steps which are briefly described below:

Step 1: Scale-Space Local Extrema Detection

In this stage, to identify those locations and scales that are detectable from different views of the same image, filtering is done [5]. It can be achieved using a "scale space" function and it is based on the Gaussian function as below:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \qquad (7)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2}\sigma^2 \qquad (8)$$

Where,

$*$: the convolution operator.

$G(x, y, \sigma)$: a variable-scale Gaussian and

$I(x, y)$: the input image.

Difference of Gaussians technique is used to detect keypoint location in the scale space, which locate the scale-space extrema of image, $D(x, y, \sigma)$ and it is computing by the difference between two images, one with scale k times the other [6]. $D(x, y, \sigma)$ is then given by:

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \qquad (9)$$

Where,

$D(x, y, \sigma)$:scale-space extrema.

$L(x, y, k\sigma)$: k times scale.

$L(x, y, \sigma)$: scale.

The position of the local maxima and minima of $D(x, y, \sigma)$ is detected by comparing each point of scale with its 8 neighbors at the same scale, and its 9 neighbors up and down one scale. If the value is the minimum or maximum of all these points then this point is an extrema [2]

Step 2: Keypoint Localization

A detailed approximation of data for scale, location and ratio of principle curvature is performed to locate the points. The points that are poorly localized along an edge or sensitive to noise are rejected. For this, Taylor series expansion is used for 3D fitting curve which gives scale space function $D(x, y, \sigma)$ [8].

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x \qquad (10)$$

where,

D and its derivatives: evaluated at the sample point.

x : (x, y, ∂^T) offset from the point.

By taking the derivative of $\hat{x}$ function with respect to x and equating it to zero, it gives location of extreme $\hat{x}$.

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x} \qquad (11)$$

Step 3: Orientation Assignment

In this step, to each feature, orientation is assigned based on local image gradients [4]. Using pixel differences, the gradient magnitude, $m(x, y)$, and orientation, $\theta(x, y)$, is calculated for each image $L(x, y)$ [7]. The gradient magnitude and orientation are computed for each Pixel of the region around the feature location respectively as:

$$m(x, y) = \sqrt{(\mathrm{L}(x+1, y) - \mathrm{L}(x-1, y))^2 + (\mathrm{L}(x, y+1) - \mathrm{L}(x, y-1))^2} \qquad (12)$$

$$\theta(x, y) = tan^{-1}((L(x, y+1) - L(x, y-1))/(L(x+1, y) - L(x-1, y))) \qquad (13)$$

Where,

$L(x, y)$: image sample.

$m(x, y)$: image sample gradient magnitude.

$\theta(x, y)$: image sample orientation.

Step 4: Keypoint Descriptor

In this step, a keypoint descriptor is created. The keypoint descriptors are biased by a Gaussian window. Here, the images are accumulated into orientation histograms which summarized the contents over 4x4 sub-regions [9].

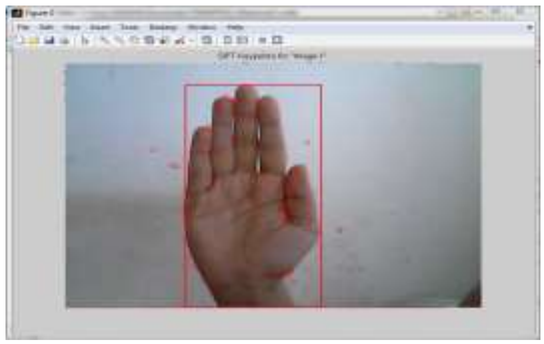

Figure 7. The real time SIFT feature of hand gesture

C. Gesture Recognition:

In order to get a constant recognition, the features extracted from the gesture are noticeable under the changes in image scale, noise and illumination [3]. Keypoints are generally lying on high-contrast regions of the gesture, for example object edges. By matching each keypoint of real time gesture individually to the stored keypoints extracted from gesture, gesture recognition is performed.

- Results

Figure 8 shows result of static gesture recognition using GUI. The experimental setup used for this is MATLABR2012.The gesture is selected from database and pre-processing, feature extraction steps are carried out and finally a recorded sound is generated for that particular gesture. The results for MSL gesture is shown in the figure.

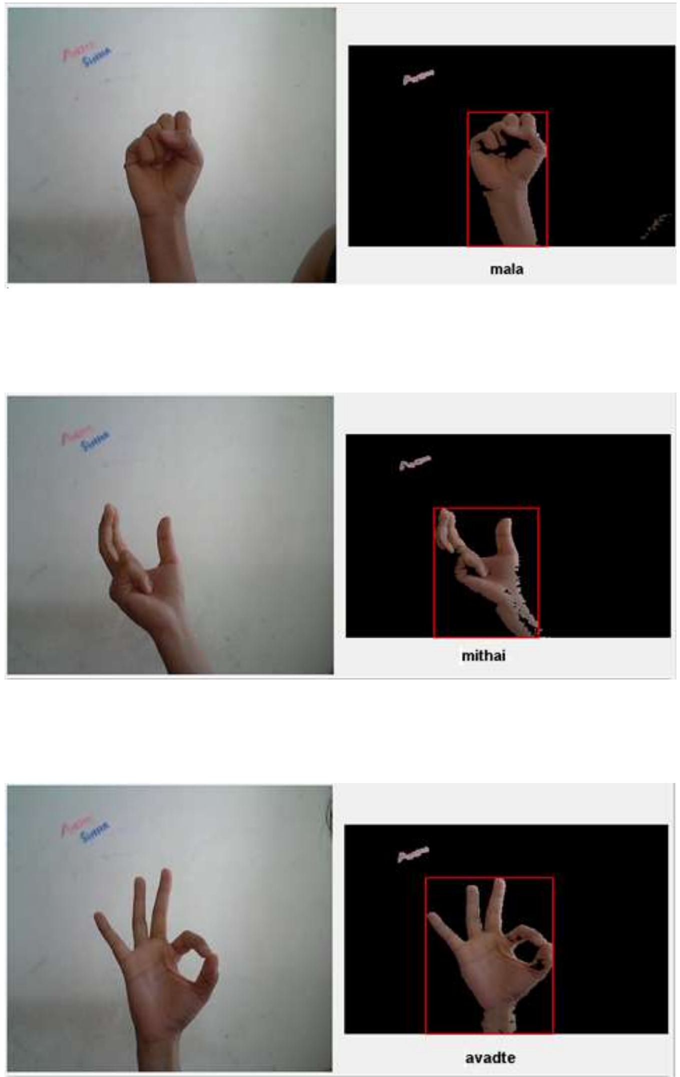

Figure 8. Gesture recognition in real time ‘mala’, ‘mithai’, ‘avadte’

- **CONCLUSION**

The hand tracking algorithm is capable to track and recognize MSL single hand gestures using features. We have developed a system for the purpose of the recognition of sign language. The system have three stages: tracking of gesture stage, feature extraction stage and the gesture recognition stage. The number of keypoints detected by the proposed algorithm can decrease the runtime of the implementation while the matching rate maintains at a high level. The importance of this work is to help primarily for deaf to solve their communication problem.

REFERENCES

1. Momin Rijwan Ramjan, Rane Mithilesh Sandip, Phatangare Sushant Uttam, Wani Sumit Srimant, Dynamic Hand Gesture Recognition and Detection for Real Time Using Human Computer Interaction, International Journal of Advanced Research in Computer Science and Management Studies, Vol. 2, Issue 3, March 2014.

2. Liang-Chi Chiu, Tian-Sheuan Chang, Jiun-Yen Chen, and Nelson Yen-Chung Chang, Fast SIFT Design for Real-Time Visual Feature Extraction, IEEE Transactions on Image Processing, VOL. 22, NO. 8, pp. 3158-3159, August 2013.

3. Pallavi Gurjal, Kiran Kunnur, Real Time Hand Gesture Recognition Using SIFT, International Journal of Electronics and Electrical Engineering,ISSN : 2277-7040 Volume 2 Issue 3,pp. 22-25, March 2012.

4. Yang Li, Lingshan Liu, Lianghao Wang, Dongxiao Li, Ming Zhang, Fast SIFT Algorithm based on Sobel Edge Detector, IEEE, pp. 1820-1821, 2012.

5. Zheng-Jianding, Yang Zhang, A-Qing Yang, Dai-Li, Image Matching of Gaussian Blurred Image Based on SIFT Algorithm, IEEE, pp. 121-122, 2012.

6. Wenyu Chen, Yanli Zhao, Wenzhi Xie, Nan Sang, An Improved SIFT Algorithm for Image Feature- Matching, IEEE, pp. 197-198, 2011.

7. Xuelong Hu, Yingcheng Tang, Zhenghua Zhang, Vedio Object Mathing Based on SIFT Algorithm, IEEE Int. Conference Neural Networks & Signal Processing Zhenjiang, China, pp. 412-414, June 8~10, 2008.

8. David G. Lowe, Object Recognition from Local Scale-Invariant Features.

9. David G. Lowe, Distinctive Image Features from Scale-Invariant Keypoints, International Journal of Computer Vision, 2004.