

# Computer Vision Based Hand Gesture Recognition For Speech Disabled Persons

Manoj Acharya<sup>1</sup>, Dibakar Raj Pant<sup>2</sup>

*Department of Electronics and Computer Engineering  
Institute of Engineering, Pulchowk, Lalitpur  
corresponding author: manoja328@gmail.com*

---

**Abstract:** This paper proposes a method to recognize static hand gestures in an image or video where a person is performing Nepali Sign Language (NSL) and translate it to words and sentences. The classification is carried out using Neural Network where contour of the hand is used as the feature. The work is verified successfully for NSL recognition using signer dependency analysis.

---

## 1. Introduction

Gestures have always been a part of our communication since eternity. The goal of Human Computer Interaction (HCI) is to bring the performance of human machine interaction closely similar to human-human interaction. It concerns with the design, evaluation and implementation of interactive computing systems for human use [3]. Input devices such as keyboards, mouse etc have been used for human computer interface, but these devices do not provide natural communication with the computer; thus a vision based interface is necessary. A common example in vision based interface is hand gesture recognition. Various algorithms have been applied for hand gesture recognition. For example, Symeonidis [7] proposed orientation histogram technique to recognize static hand gesture. Chen et.al [1] used decisions tree and Marcel [5] used Hidden Markov Models to recognize dynamic aspects of hand gesture. Furthermore, deterministic boosting [4] has been used to recognize single-hand gestures in real time.

Appearance based hand gesture recognition is one of the popular method where 2D image features are used to model the visual appearance of the hand and compare these parameters with the extracted image features from the input image. Hand colors [8], local hand features [6], optical flow [2] are image features used to detect human hands as well as to recognize gestures. The authors propose the use of a neural network base learning for recognizing gestures where a 2 layered NN using back propagation algorithm has been used for training. The contour of the gestures has been used as the feature for training our Neural Network. The input layer consist of 900 neurons, which takes 30\*30 images which are the scaled down version of contours of actual binary hand images. The data is then trained using 25-neurons based hidden layer with gradient descent for optimization.

## 2. Methodology

The hand gesture recognition has 4 stages, namely Background subtraction, Segmentation, Feature extraction and Classification. These stages are briefly explained in the following subsections.

### 2.1 Background Subtraction

The Frame Difference method for background subtraction uses the technique of threshold. If the difference value between two consecutive frames exceeds the threshold, the output value is set to one otherwise it is set to zero. Mathematically, the background subtraction equation for Frame Difference Method is given by equation (1)

$$|I_{(x,y,t)} - I_{(x,y,t-1)}| > Th \quad (1)$$

where,

$I_{(x,y,t)}$  is image at time  $t$ ,

$I_{(x,y,t-1)}$  is image at time  $t-1$ , and

$Th$  is the threshold level.

### 2.2 Segmentation

Segmentation is done by applying the contour extraction method for the hand image. To accomplish this task, the zero order moment of all binary regions of hand image is calculated, which describes the area of the object. The area  $A$  of such region is:

$$A = \sum_x \sum_y b(x, y) \quad (2)$$

where,

$b(x, y)$  is a binary image.

### 2.3 Feature Extraction

The whole contour obtained is scale down to an appropriate version 30\*30 and the image is vectorized after the pixels are normalized. Figure 1, below are few of the images used to generate the training data set.

### 2.4 Classification

The classification is carried out using neural network. Back-Propagation is used as a supervised learning method based on generalized Delta Rule and gradient descent is used as the optimization method where the cost function for logistic regression is:

$$J(\theta) = -\left[\frac{1}{m} \sum_{i=1}^m y^{(i)} \times \log(h_{\theta}(x^{(i)})) + (1 - y^{(i)}) (1 - h_{\theta}(x^{(i)}))\right] + \frac{\lambda}{2m} \sum_{i=1}^n \theta_j^2 \quad (3)$$

where,

$\theta$  = Model parameter of Neural Network

$m$  = number of training examples (observations)

$y$  = output/target variable

$x$  = input variable /feature

$\lambda$  = Regularization Parameter

$$h\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

Training the neural network involves finding the minima of the above cost function.

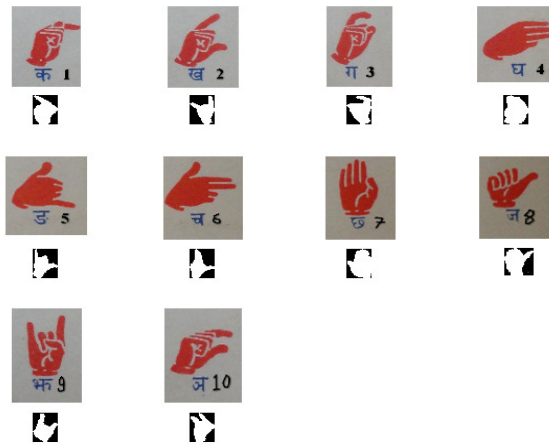


Figure 1: Training set images corresponding to NSL gestures used in training set

### 3. Experimental Results

The applications of the system are documented in the given figures below.



Figure 2: Drawing application using hand gestures

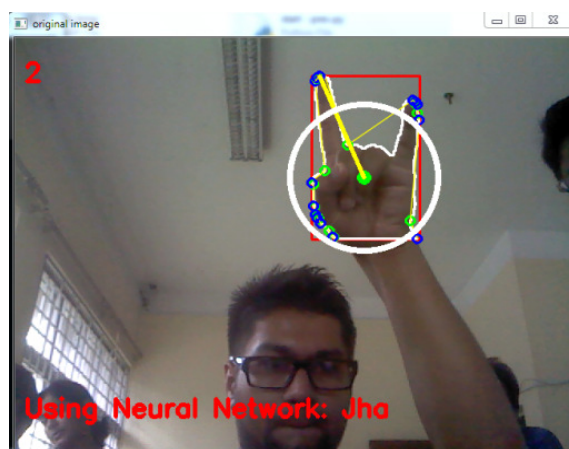


Figure 3: NSL recognition

In the figure 2, the finger tip is used to draw in a drawing application. The process is started by giving a start-draw signal gesture where the points are connected when the hand is moved and the process is stopped using any specific gesture. In the figure 3, the actual recognition of NSL gesture is shown. The identified gesture is for the class "Jha". The recognition is done in real-time and the feedback is given to the user.

### 3.1 Analysis of Signer Dependency Testing

A gesture recognition system should be independent of scale of the signs, although it can be rotation dependent because different orientation of same gesture could mean separate meaning. It must also be invariant to marginal occlusion. To test the independence from the person performing the gesture, hand gestures of 2 different persons is taken and system is trained using the data of one person but tested for the data of another.

Table 1: Number of Training data-set for different class

Gesture Class	First data-set	Second data-set
1(Ka)	28	11
2(Kha)	164	9
3(Ga)	240	13
4(Gha)	147	144
5(Nga)	178	186
6(Cha)	121	127
7(Chha)	194	206
8(Ja)	158	174
9(Jha)	295	336
10(Yan)	168	487

Table 2: Accuracy for different data-set

Gesture Class	Accuracy of first data-set	Accuracy of second data-set
1(Ka)	0.54545	1
2(Kha)	0.44444	0.63415
3(Ga)	1	0.64167
4(Gha)	1	0.71429
5(Nga)	0.875	0.88764
6(Cha)	1	0.66942
7(Chha)	1	0.91237
8(Ja)	0	0
9(Jha)	1	0.81017
10(Yan)	0.92163	0.875

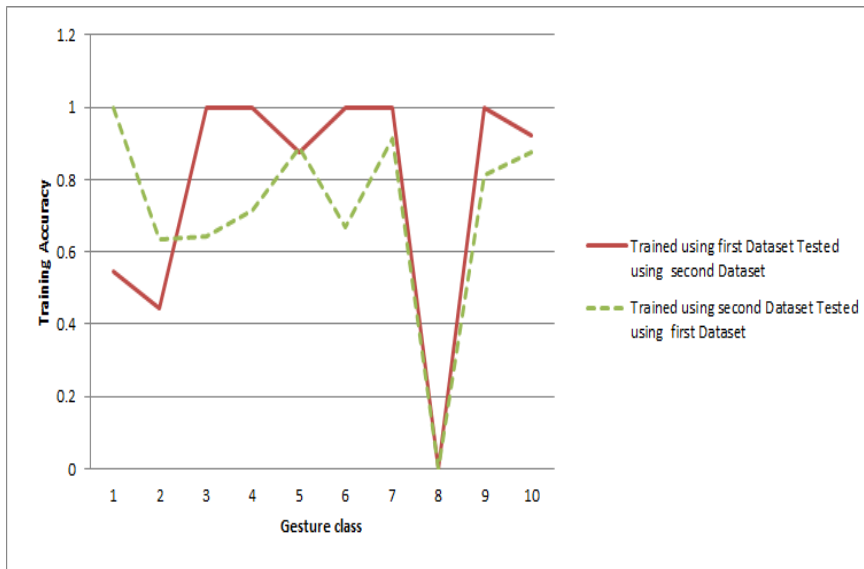


Figure 4: Graph of accuracy for different Data-set

Figure 4 displays the plot for accuracy of the data shown in Table II. The overall recognition accuracy is about 97%. But there is an anomaly in the 8th class. A closer observation reveals that although they are same gesture they are rotated by 180 degree as shown in figure 5. Thus, resulting the zero percent accuracy and this is indeed what is wanted for the system.

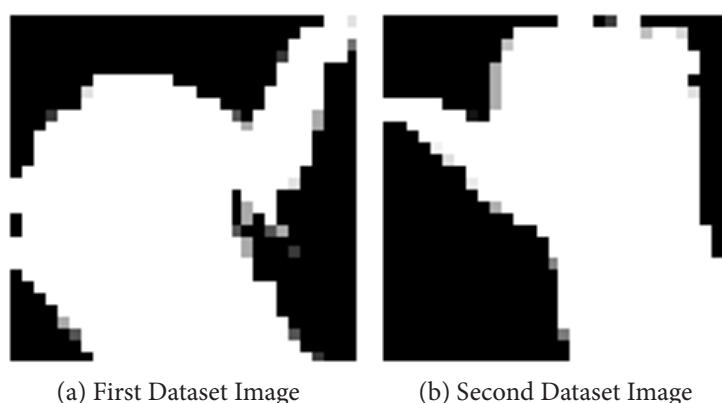


Figure 5: Samples for eighth class

#### 4. Conclusion

Neural network based recognition method using contour can produce a robust system for Nepali Sign Language Recognition which is verified using the signer dependency analysis. This analysis confirms that the trained sign gestures are independent to scale of the sign gestures inputted and also the robustness of Neural Network ensures a correct recognition even in cases of minor occlusion or noise. Since, the system operates in real time; a real time communication can also be managed by the proposed system.

#### References

- [1] Chen Y, Gao W, MA J, “Hand Gesture Recognition Based on Decision Tree”, Institute of Computing Technology, Chinese Academy of Sciences, Beijing.
- [2] Cutler R and Turk M (1998), View-based interpretation of real-time optical flow for gesture recognition. *Proc. Third IEEE Conference on Face and Gesture Recognition*.
- [3] Dix A, Finlay J, Abowd GD and Beale R (2003), *Human-Computer Interaction. 3rd ed. Prentice Hall*.
- [4] Lockton R and Fitzgibbon AW (2002), “Realtime gesture recognition using deterministic boosting”, Department of Engineering Science, University of Oxford, BMVC.
- [5] Marcel S, Bernier O, Viallet JE and Collobert D, “Hand Gesture Recognition using Input–Output Hidden Markov Models”, *France Telecom CNET, 2 avenue Pierre Marzin, 22307 Lannion, FRANCE*.
- [6] Oka K, Sato Y, and Koike H (2002), Real-time fingertip tracking and gesture recognition. *Proc. IEEE Computer Graphics and Applications* **22(6)**, 64-71.
- [7] Symeonidis K (2000), Hand Gesture Recognition Using Neural Networks. Degree of Master of Science in Multimedia Signal Processing communications, School of Electronic and Electrical Engineering.
- [8] Wu Y and Huang TS (2002), Non-stationary color tracking for vision-based human computer interaction. *IEEE Trans. on Neural Networks*, **13(4)**, 948- 960.