Indian Sign Language Recognition Using Combined Feature Extraction

R. Itkarkar Rajeshri, Anil Kumar V. Nandi, and Vaishali B. Mungurwadi

Abstract This research paper aims for the recognition of Indian sign language (ISL). Sign language is a language commonly used by deaf and dumb people to communicate with each other and rest of the world. There is an extensive research carried out for American sign language (ASL), but due to the lack of standard dataset, research for Indian sign language recognition is hampered a lot. This research work focuses on the use of a combined feature extraction technique so as to improve the accuracy and reduce complexity. Histogram of orientation gradient (HOG) and Gabor features are combined and classified using support vector machine (SVM) and K-Nearest neighbor (KNN) with accuracy of 83.92% and 84.92%, respectively.

Keywords Indian sign language (ISL) · HOG · Gabor · SVM

1 Introduction

Vision-based hand gesture recognition is appealing more nowadays as it provides the most natural way to interact for human-machine interaction. The vision base method is widely adopted for research due to low computational complexity. This paper work presents research carried out for the recognition of Indian sign language (ISL). ISL is one of the sign languages which is more complex than American Sign Language as it consists of complex signs (most signs are two hand signs). The typical processes performed for sign language recognition are preprocessing, feature extraction, and classification. Preprocessing consists of converting a color image into gray. Feature extraction is a technique where features such as shape, geometric features, statistical features, texture features, etc. for an image can be extracted. The shape of the hand can be used to identify the gesture termed as shape identification. The contour of the hand identifies the shape. Extracting the contour of hand gives more information in shape detection. Classification techniques such linear classifiers KNN, SVM, and

---

R. Itkarkar Rajeshri (✉) · A. K. V. Nandi
BVB Hubali, Hubli, Karnataka, India

V. B. Mungurwadi
Aavishkar Technologies, Hubli, Karnataka, India

neural network can be applied for recognition. This paper focuses on creation of own database with use of a simple web camera, feature extraction by combining Histogram of orientation gradient (HOG) and Gabor features and classified using support vector machine (SVM) and K-Nearest neighbor (KNN). Gabor filters are used in image processing due to its mathematical and biological properties (Guptaa et al. 2012). The feature dimension generated depends on the selection of parameters for the Gabor filter. The Gabor filter is designed by selecting parameters such as orientation, bandwidth, and frequency. The HOG features give the spatial distribution of local intensity gradients. These features well describe the hand gestures as they describe the edge features. Thus the shape feature of the hand gestures can be extracted by using HOG.

2   Related Work

Hand gesture recognition is one form of interaction between human and computer to achieve typical application. A real-time hand gesture recognition system implemented (Kishore and Rajesh Kumar 2012) with an accuracy of 96% using a combination of color and texture features and fuzzy logic for classification (Nandy et al. 2010). Indian sign language recognition was implemented by evaluating mean feature of histogram gradient and Euclidean distance for recognition and used for controlling a humanoid robot. Gabor is a linear filter that gives best localization characteristics by changing, bandwidth, frequency, and orientation (Huang et al. 2010). In (Zhao et al. 2010) extracted HOG features were converted into low-dimensional subspace using PCA-LDA. It was classified using the nearest neighbor classifier to achieve a recognizing accuracy of 91% in real-time. While in (Teoh and Branunl 2015) the authors used both HOG and Gabor for vehicle detection with three different classifiers SVM, Multi perceptron neural network, and distance classifier. They obtained best performance with HOG and SVM with less processing time. The authors in (Sheenu et al. 2015) have used HOG method followed by sequential minimal optimization with a recognition rate of 93.12%.

3   Methodology

The methodology proposed is a novel method for ISL recognition, where the Gabor features and HOG features combined to form a feature vector. The obtained feature vector is of a higher dimension, and hence PCA is used further to reduce the dimension and then applied to the classifier for recognition. The classifier used is SVM and KNN. Figure 1 shows the proposed methodology for ISL recognition.
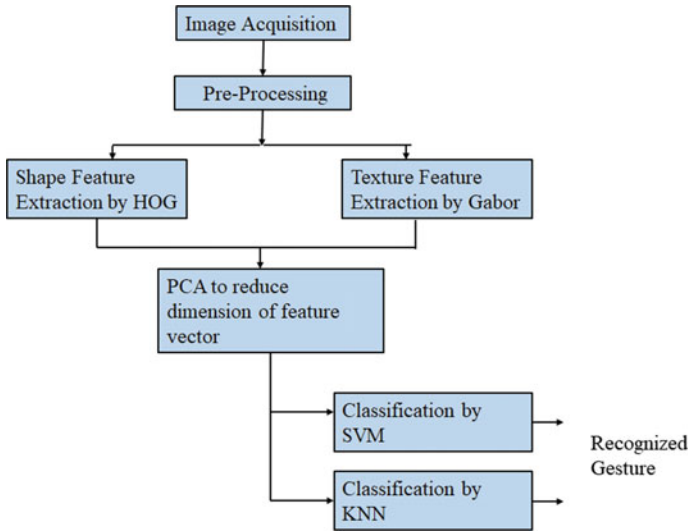
Fig. 1 Proposed methodology for ISL recognition

3.1 Feature Extraction by HOG

The HOG features are widely used for object detection. The image is divided into small square cells, the histogram of oriented gradients is computed for each cell, normalizes the result using a block-wise pattern, and return descriptors for each cell. Histogram of Oriented Gradient descriptor assumes that the local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions. The implementation of these descriptors can be achieved by dividing the image into small connected regions called cells, and for each cell computing a histogram of gradient directions i.e. edge orientations for the pixels within the cell. The combination of these histograms then represents the descriptor (Savaris and von Wagenheim 2010). The shape features are evaluated by applying color normalization on the input image, then evaluate the horizontal and vertical gradients, next is the formation of spatial blocks and then calculate orientation bin-wise and then forming a feature vector. The gradient magnitude and its orientation are calculated as in Eqs. (1) and (2) respectively, where gx and gy are horizontal and vertical gradients.

$$G = \sqrt{gx^2 + gy^2} \tag{1}$$

$$\theta = \tan^{-1} \frac{gy}{gx} \tag{2}$$

3.2 Feature Extraction by Gabor Filter

Gabor filter is a linear filter used for object edge detection. Gabor transform has strong frequency and orientation selectivity so that the edge features can be extracted. Gabor gives the best resolution in time and frequency domain and hence recognized as a very useful tool in computer vision and image processing (Huang et al. 2010). The parameters such as bandwidth, frequency, and orientation are changed to achieve best local features as Gabor is a linear filter. The features are extracted by convolution of the Gaussian kernel with the input image. A 2-D Gabor filter kernel over the image (x, y) is defined as per Eq. 3

$$G(x, y, \theta, \lambda, \varphi, \sigma, \gamma) = \exp{-\frac{1}{2}\left\{\frac{x'^2}{\sigma x'^2} + \frac{y'^2}{\sigma y'^2}\right\}} \cos\left(\frac{2\pi}{\lambda x'} + \varphi\right) \qquad (3)$$

where $x' = x\sin\theta + y\cos\theta$ and $y' = x\cos\theta - y\sin\theta$, $G(x, y, \theta, \lambda, \varphi, \sigma, \gamma)$ kernel is a function of various parameters $\theta, \lambda, \varphi, \sigma, \gamma$ of the wavelet. θ is the orientation of the Gabor function, varied between 0 and 360. λ is the wavelength of the cosine factor of the Gabor kernel referred to as the wavelength of the filter. φ, it is the phase shift of the Gabor function in degrees which specifies the elasticity of the Gabor function. The features are extracted by convolution of the image with Gabor kernel represented as in Eq. 4

$$G(x, y, \theta, \lambda, \varphi, \sigma, \gamma)(x, y) = I(x, y) * G(x, y, \theta, \lambda, \varphi, \sigma, \gamma) \qquad (4)$$

where $I(x, y)$ is the image.

The HOG features and Gabor features are finally combined together to form feature vector. The features are concatenated and the length of the feature vector.

4 Results

4.1 Results of HOG and Gabor

For implementation, Matlab is used. The final HOG vector Obtained is a vector of 2 $\times$ 2 $\times$ 9 vector i.e. 36 $\times$ 1. Here as the block overlap is of 2 there are 9 matrices of 2 $\times$ 2 size. Thus 8 $\times$ 8 cell finally reduced to 2 $\times$ 2 of 2 block overlap. Therefore 9 bins contain gradients of each cell. The final feature vector evaluated is the combination of both HOG and Gabor. This combination provides a perfect feature matrix which represents signs of Indian sign language. The HOG extraction method applied to a grayscale resized image of 130 $\times$ 130 resolution results in a feature vector of size 2700 $\times$ 1. The cell size selected 8, the block size is 2, and bin size is 3 for HOG extraction. The Gabor coefficients obtained after convolution of the Gaussian kernel
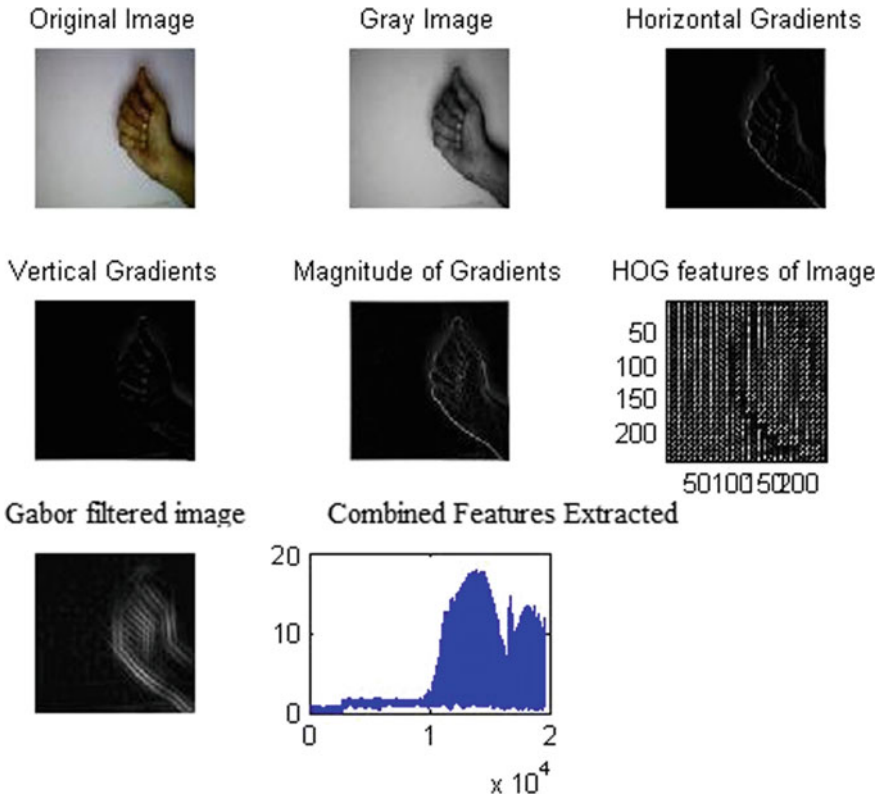
Fig. 2 Results of combined features of HOG and Gabor gestures for sign "0"

with the sign image are of size $16,900 \times 1$. Figures 2 and 3 shows the results for sign "0" and "A".

The performance of the classifier algorithm is stated by evaluating the accuracy from confusion matrix. The confusion matrix is as shown in Table 1 is for SVM and Table 2 is for KNN. The average accuracy obtained with SVM 83.92% and with KNN is 84.92% for $K = 3$. As there is very less research carried on ISL and no method used based on combined recognition comparison with existing work cannot be obtained.

5 Conclusion

The combined feature extraction by HOG and Gabor technique is obtained to increase the accuracy and reduce complexity of the system, though average accuracy obtained is just 83.92 and 84.92%. Gabor though is a robust technique accuracy decreases as filter output depends on many parameters. The recognition for ISL is a challenging

Fig. 3 Results of combined features of HOG and Gabor gestures for Sign "*A*"

Table 1 Confusion matrix by combined hog and Gabor features with SVM

Sign	*A*	*B*	*C*	*D*	1	2	3	4
A	**71**	0	0	0	0	0	0	0
B	0	**50**	0	0	0	0	0	0
C	0	0	**80**	0	0	5	0	0
D	0	0	0	**75**	0	0	0	25
1	0	0	0	0	**100**	0	0	0
2	0	0	0	0	0	**100**	0	0
3	0	0	0	0	0	0	**100**	0
4	0	0	0	0	0	0	0	**95**

Table 2 Confusion matrix by combined hog and Gabor features with KNN

Sign	A	B	C	D	1	2	3	4
A	**71**	0	0	0	0	0	0	0
B	0	**75**	5	0	0	0	20	0
C	0	0	**65**	0	15	5	0	0
D	0	0	0	**75**	0	0	0	25
1	0	0	0	0	**100**	0	0	0
2	0	0	0	0	0	**100**	0	0
3	0	0	0	0	0	2	**98**	0
4	0	0	0	0	0	0	0	**95**

task as the signs in ISL are complex. Further, the technique can be extended to recognize sentences and generate audio output for the recognized gestures.

References

Chen Q, et al (2008) Hand gesture recognition using Haar-like features and a stochastic context-free grammar. IEEE Trans Instrum Measure 57:9

Geetha M, Manjusha UC (2013) A vision based recognition of Indian sign language alphabets and numerals using B-spline approximation. Int J Comput Sci Eng (IJCSE)

Guptaa S, Jaafar J, Ahmad WFW (2012) Static hand gesture recognition using local gabor filter. In: International symposium on robotics and intelligent sensors

Huang Z, Jiang D, Zhao W (2010) Study of sign language recognition based on gabor wavelet transforms. In: International conference on computer design and applications (ICCDA 2010). IEEE

Kishore PVV, Rajesh Kumar P (2012) A model for real time sign language recognition system. Int J Adv Res Comput Sci Softw Eng 2(6):30–35

Nandy A, Mondal S, Prasad JS, Chakraborty P, Nandi GC (2010) Recognizing and interpreting Indian sign language gesture for human robot interaction. In: International conference on computer and communication technology, ICCCT'10, pp 712–717

Savaris A, von Wagenheim A (2010) Comparative evaluation of static gesture recognition techniques based on nearest neighbor, neural networks and support vector machines. J Braz Comput Soc 16:147–162

Sheenu, Joshi G, Vig R (2015) A multi-class hand gesture recognition in complex background using sequential minimal optimization. In: International conference on signal processing, computing and control

Teoh SS, Branunl T (2015) Performance evaluation of HOG and Gabor features for vision based vehicle detection. In: IEEE international conference on control system, computing and Engineering 27–29 Nov 2015

Zhao Y, Wang W, Wang Y (2011) A real-time hand gesture recognition method. 978-1-4577-0321-8/11 ©2011 IEEE