

# On an algorithm for Vision-based hand gesture recognition

Dipak Kumar Ghosh<sup>1</sup> · Samit Ari<sup>1</sup>

Received: 11 June 2014 / Revised: 5 June 2015 / Accepted: 18 June 2015  
© Springer-Verlag London 2015

**Abstract** A vision-based static hand gesture recognition method which consists of preprocessing, feature extraction, feature selection and classification stages is presented in this work. The preprocessing stage involves image enhancement, segmentation, rotation and filtering. This work proposes an image rotation technique that makes segmented image rotation invariant and explores a combined feature set, using localized contour sequences and block-based features for better representation of static hand gesture. Genetic algorithm is used here to select optimized feature subset from the combined feature set. This work also proposes an improved version of radial basis function (RBF) neural network to classify hand gesture images using selected combined features. In the proposed RBF neural network, the centers are automatically selected using k-means algorithm and estimated weight matrix is recursively updated, utilizing least-mean-square algorithm for better recognition of hand gesture images. The comparative performances are tested on two indigenously developed databases of 24 American sign language hand alphabet.

**Keywords** American sign language (ASL) hand alphabet · Combined feature · Genetic algorithm (GA) · Hand gesture recognition · Least-mean-square (LMS) algorithm · Radial basis function (RBF) neural network

## 1 Introduction

Development of efficient hand gesture recognition system is crucial for successful human–computer interaction (HCI) or human alternative and augmentative communication (HAAC) applications like robotics, assistive systems, sign language communication and virtual reality [1–5]. Gestures are communicative, meaningful body motions connecting physical actions of the fingers, hands, arms, face, head or body with the intent of conveying meaningful information or interacting with the environment [6]. In general, gestures can be classified into static gestures [1–4] and dynamic gestures [7–9]. Static gesture is described in the form of definite hand configuration or poses, while dynamic gesture is a moving gesture, articulated as a sequence of hand movements and arrangements. However, static gestures communicate certain meanings or sometimes act as explicit transition state in dynamic gestures. The sensing techniques which are used in static hand gesture recognition systems include glove-based techniques [10, 11] and vision-based techniques [1, 3, 7]. In glove-based techniques, sensors are utilized to measure the joint angles, positions of the fingers and position of the hand in real time [11]. However, gloves are quite expensive, and the weight of the gloves and associated measuring equipment hamper free movement of the hand. Therefore, user interface is complicated and less natural for glove-based techniques [12]. On the other hand, vision-based techniques use one or more cameras to capture the gesture images. Vision-based system may not be robust enough for few applications like gesture-based remote control [13]. However, vision-based techniques provide a natural interaction between humans and computers without using any extra devices [12]. Various features have been reported in the literature [1–4, 14] to represent static hand gesture including features like statistical moments [1], local-

✉ Dipak Kumar Ghosh  
dipakkumar05.ghosh@gmail.com

Samit Ari  
samit.ari@gmail.com

<sup>1</sup> Department of Electronics and Communication Engineering,  
National Institute of Technology, Rourkela 769008, India

ized contour sequence (LCS)[2], block-based feature [3], Fourier discriminator (FD) [4] and discrete cosine transform (DCT)-based features [14]. Similarly, a variety of methods have been reported in the literature [1–3, 14, 16] to classify hand gesture including methods like multilayer perceptron back-propagation (MLP-BP) neural network [1], minimum distant (MD) classifier [2], fuzzy C-mean (FCM) classifier [3], k-nearest neighbor classifier (KNN) [14] and radial basis function (RBF) neural network [16]. However, accurate feature representation and development of an efficient classifier are still being a challenging task in developing a static hand gesture recognition system for real-time application.

This paper presents a vision-based static hand gesture recognition method which consists of preprocessing, feature extraction, feature subset selection and classification stages. The preprocessing stage involves image enhancement, segmentation, rotation and filtering. This work proposes a rotation-invariant technique by coinciding the first principal component of the segmented hand gestures with vertical axes. In the feature extraction stage, this work constructs a combined feature set by appending LCS feature set with block-based feature set for a better representation of static hand gesture. Genetic algorithm (GA) is used here to select optimized feature subset from combined feature set and improves the recognition performance. This work also proposes an improved version of RBF neural network as a classifier. RBF neural network [15] is used here because: (i) Its architecture is very simple, only one hidden layer is present between input and output layers; (ii) in hidden layer, localized RBF is used for nonlinear transform of feature vector from input space to hidden space; (iii) it has universal approximation and regularization capabilities; (iv) this network is faster and free from local minima problem. In this work, the centers of the proposed RBF classifier are automatically selected through k-means-based center-selection algorithm, and the estimated weight matrix is recursively updated using least-mean-square (LMS) algorithm. For performance evaluation, experiments are conducted on two indigenously developed databases of 24 American sign language (ASL) hand alphabet. The experimental results show that the performance of the static hand gesture recognition technique using GA-based selected combined feature subset is superior compared to the individual feature set as well as combined feature set and the performance of proposed RBF is better than MLP-BP as in [1] and standard RBF as in [16]. Experimental results also show that the proposed hand gesture recognition method provides better performance compared to existing techniques such as block-based features with MLP-BP classifier (Block-MLP-BP), LCS features with MLP-BP classifier (LCS-MLP-BP), combined features with MLP-BP classifier (Combined-MLP-BP), block-based features with standard RBF classifier (Block-SRBF), LCS features with standard RBF (LCS-SRBF), combined features

with standard RBF classifier (Combined-SRBF), Hu moment invariants with MLP-BP classifier (HU-MLP-BP) [1], DCT coefficients with k-nearest neighbors classifier (DCT-KNN) [14] and normalized silhouette distance signal features with k-least absolute deviations classifier (NSDS-KLAD) [17].

The rest of the paper is organized as follows: Sect. 2 briefly describes the gesture database. Section 3 explains the proposed gesture recognition method in detail. Experimental results are discussed in Sect. 4, and finally the conclusions are given in Sect. 5.

## 2 Gesture database

Gesture images from two gesture databases are used in this study, which includes grayscale images for Database I and color images for Database II. Both databases are indigenously developed by capturing static gesture of 24 ASL hand alphabet images from 10 volunteers using VGA Logitech Webcam (C120). Each volunteer provides a guide of postures as appear in the ASL browser developed by Michigan State University and BGU-ASL DB [3]. Database I consists of 2400 hand alphabet grayscale images with spatial resolution of  $320 \times 240$  pixels. Each ASL hand alphabet contains 100 images with 10 samples per volunteer. The ASL hand alphabets are performed by every volunteer against an uniform black background based on different angle, position and distance from camera under different lighting condition. Database II also contains 2400 ASL hand alphabet color images with spatial resolution of  $320 \times 240$  pixels. Each ASL hand alphabet comprises of 100 images; 10 samples are performed by each volunteer against a nonuniform background with variations of angle, position and size changes under different lighting condition. This nonuniform background allows to test robustness of the gesture recognition algorithm.

## 3 Proposed framework

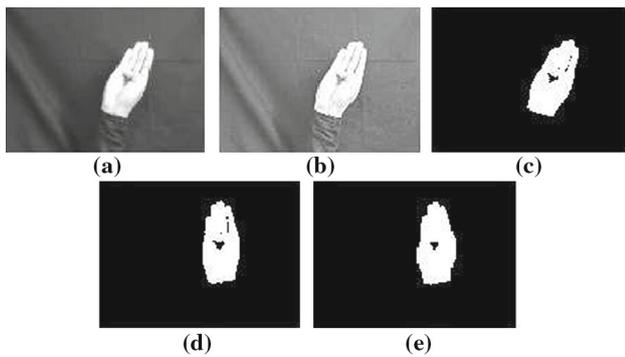
The proposed static hand gesture recognition method consists of following four stages: preprocessing, feature extraction, feature selection and classification.

### 3.1 Preprocessing

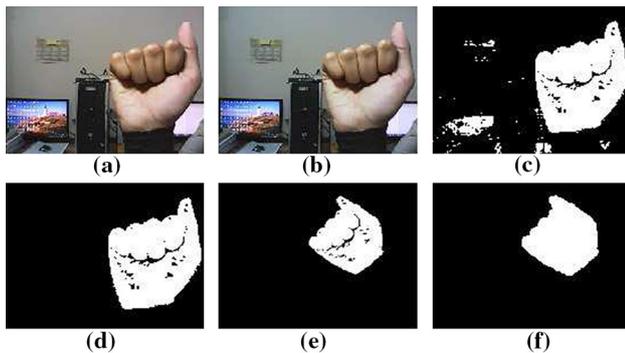
Preprocessing stage of the proposed method comprises of following steps: image enhancement, segmentation, rotation and morphological filtering. Figures 1 and 2 show the results at different steps of preprocessing for a gesture image of Databases I and II, respectively.

#### 3.1.1 Image enhancement

For grayscale images of Database I, this work uses homomorphic filtering [18] technique to enhance hand gesture images



**Fig. 1** Results of preprocessing stage for a grayscale image of Database I. **a** Original gesture image. **b** Enhanced image. **c** Otsu segmented image. **d** Rotated image. **e** Morphological filtered image



**Fig. 2** Results of preprocessing stage for a color image of Database II. **a** Original gesture image. **b** Enhanced image. **c** Detected skin color region. **d** Final segmented image. **e** Rotated image. **f** Morphological filtered image

by normalizing illumination variation within it. For color images of Database II, this work utilizes gray world technique [19] to enhance gesture images by compensating the variation of light. Figures 1b and 2b represent the results of image enhancement done by homomorphic filtering and gray world techniques, respectively.

### 3.1.2 Segmentation

The objective of segmentation is to extract the hand region from the background of the gesture image. For grayscale images of Database I, the hand region of the enhanced gesture image is segmented using Otsu segmentation method [20]. The result of Otsu segmentation method is shown in Fig. 1c. For color image of Database II, the hand is segmented from background of the enhanced gesture image using skin color detection method in YCbCr color space [21]. In this work, a pixel is considered as skin pixel if  $Th < Y < 255$ ,  $85 < Cb < 128$  and  $129 < Cr < 185$ , where  $Th$  is  $(1/3)$  of average  $Y$  value of all pixels. The skin color detection result for hand gesture image is shown in Fig. 2c. From the detection result, it is observed that the skin color detection image also contains other skin color objects not belonging

to hand region. Therefore, the hand region is assumed to be largest connected skin color object, and other detected skin color objects are filtered out by comparing their area. The final segmented output is shown in Fig. 2d.

### 3.1.3 Rotation

In this work, an image rotation technique is proposed to make segmented image rotation invariant using the direction of the first principal component. Stepwise description of the proposed technique is as follows.

1. First, represent all nonzero pixel position of segmented gesture as  $H = [X_1, X_2, X_3, \dots, X_M]$ , where  $H$  is the position vector,  $X_i = [x_i \ y_i]^T$  is the position of  $i$ th nonzero pixel with respect to the input coordinate system, and  $M$  is total number of nonzero pixels in segmented gesture.
2. Compute covariance matrix  $C = (H - \bar{X})(H - \bar{X})^T$ , where  $\bar{X} = [\bar{x}_1 \ \bar{y}_1]^T$  is the centroid of the segmented gesture.
3. Compute eigenvalues and eigenvectors of  $C$ , and find the first principal component. Let, eigenvalues of  $C$  are  $\lambda_1$  and  $\lambda_2$  and corresponding eigenvectors are  $e_1 = [e_{11} \ e_{21}]^T$  and  $e_2 = [e_{21} \ e_{22}]^T$ . Since  $\lambda_1 > \lambda_2$ , therefore  $e_1 = [e_{11} \ e_{21}]^T$  is the first principal component of segmented gesture.
4. Compute the direction of the first principal component as direction =  $\tan^{-1}(e_{21}/e_{11})$ .
5. Find the rotation angle between the first principal axes of the segmented hand gesture and vertical axes.
6. Finally, rotate the segmented hand gesture, so that the first principal axes of the segmented hand gesture coincide with vertical axes.

The results of proposed image rotation technique for a image of Databases I and II are shown in Figs. 1d and 2e, respectively.

### 3.1.4 Morphological filtering

A morphological filtering [18] technique is applied here to reduce object noise and to obtain a well-defined smooth, closed and complete segmented hand gesture. The outputs of morphological filtering for Databases I and II are shown in Figs. 1e and 2f, respectively.

## 3.2 Feature extraction

### 3.2.1 Localized contour sequence (LCS)

The LCS [2] is selected here as a feature set to represent hand gesture because it has following important properties:

(i) It is not bounded by shape complexity of gesture; (ii) it efficiently represents the gesture contours; and (iii) this representation is quite robust with respect to contour noise. This work uses the canny edge detection algorithm [22] to detect the edge of the preprocessed hand gesture and a contour tracking algorithm to track the contour of the edge detected gesture in the clockwise direction starting from the topmost left contour pixel. If  $h_i = (x_i, y_i)$ ,  $i = 1, 2, \dots, N$  is the  $i$ th contour pixel in the sequence of  $N$  ordered contour pixels of a gesture, the  $i$ th sample  $h(i)$  of the LCS is obtained by computing the perpendicular Euclidean distance between  $h_i$  and the chord connecting the endpoints  $h_{[i-(w-1)/2]}$  and  $h_{[i+(w-1)/2]}$  of a window of size  $w$  boundary pixels centered on  $h_i$  [2]. The duration and amplitude of the LCS are varied from gesture to gesture. Therefore, these are normalized by setting its duration as the average LCS duration of training dataset and standard deviation as unity. In this work, the LCS durations are normalized as 300 and 400 for Databases I and II, respectively. Note that normalized LCS feature set is position and size invariant of gesture images.

### 3.2.2 Block-based features

A bounding box is constructed around hand region of the preprocessed gesture image [3]. The box is then cropped and partitioned into blocks. The block-based feature vector of length  $V = 1 + B_r \times B_c$  is denoted as  $f_b = (f_1, \dots, f_i, \dots, f_V)$ , where  $B_r$  and  $B_c$  represent the number of rows and columns of the block partition. The first feature of  $f_b$  represents the aspect ratio of the bounding box, and the remaining features represent block averages indexed row-wise from left to right. The bounding box and a restriction on the height of the static gesture make the block-based feature vector position and size invariant. The parameters  $B_r$  and  $B_c$  are chosen through heuristic approach [3]. In this work, the static hand gestures are represented by feature vectors of length 13, i.e.,  $(1 + 4 \times 3)$ , for Database I and 21, i.e.,  $(1 + 5 \times 4)$ , for Database II.

### 3.2.3 Combined features

This work proposes a combined features to represent hand gesture. Combined features are obtained by appending the LCS features with the block-based features. This combined feature set carries contour as well as regional information of the gesture and provides better representation of the static hand gesture compared to only LCS or only block-based feature sets.

### 3.3 Feature selection

In this work, GA [23] is used to select optimized feature subset from combined feature set by removing redundant

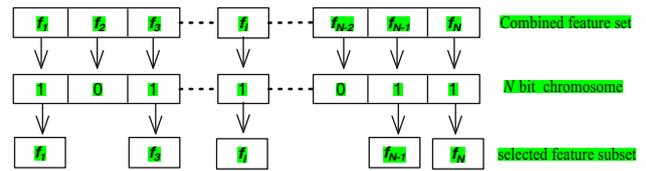


Fig. 3 Feature subset selection using chromosome bit pattern

features. The direct binary coding system as shown in Fig. 3 is used here for feature subset selection.

The detailed procedure for optimum feature subset selection using GA is given below.

1. Initial population: Initially, generate  $N_p$  number of chromosomes, where each chromosome is a binary bit pattern of  $N$  bits.
2. Feature subset selection: A feature subset is selected as shown in Fig. 3 for each chromosome bit pattern.
3. Fitness evaluation: For each chromosome, the selected features are used to train the classifier and corresponding mean square error (MSE) is calculated. The fitness value of each chromosome is obtained using fitness function as in (1),
 
$$F = 0.8 \times (\text{MSE}_{\text{train}})^{-1} + 0.2 \times [\text{abs}(N - N_s)] \quad (1)$$
 where  $\text{MSE}_{\text{train}}$ ,  $N_s$  and  $[\text{abs}(N - N_s)]$  are training MSE, number of selected features and number of reduced features, respectively.
4. Termination criteria: When the termination criteria (reaching of maximum generation) are satisfied, the process ends and finds the optimum solution which gives the maximum fitness value; otherwise, it proceeds with the next generation.
5. Genetic operation: In this step, the system searches for better solutions by genetic operations, including selection, crossover, mutation and replacement.

### 3.4 Classification

Classification is a crucial and final stage of gesture recognition method. This work develops a k-mean- and LMS-based improved RBF neural network for better classification of gesture images using selected feature vectors. The centers of the proposed RBF neural network are automatically selected using k-mean algorithm as presented in Algorithm 1. This center-selection algorithm selects equal number of centers for each class of training dataset. Initially, the weights  $w^*$  of the RBF network are estimated using well-known linear least-square estimation (LLSE) method. The estimated weights [15] are calculated by (2).

$$w^* = (G^T G)^{-1} G^T t \quad (2)$$

**Algorithm 1** : Center-selection algorithm.

- 1: Let  $s$  be the number of gesture classes. Initially, set  $i = 1$ .
- 2: Let  $D_i$  be the training data of  $i$ th class and  $k$  be the number of clusters in  $D_i$ . Initially, set the first  $k$  number of samples from training data  $D_i$  as centroid of each cluster.
- 3: Compute the Euclidean distance between each sample of training data  $D_i$  and the centroid of each cluster. Assign each sample of training data  $D_i$  to the cluster according to nearest centroid.
- 4: Compute the centroid of each cluster.
- 5: Repeat steps (3)-(4), until centroid of the each cluster does not change anymore.
- 6: Store  $k$  number of centroids as centers of RBF neural network for  $i$ th class and set  $i = i + 1$ .
- 7: Repeat steps (2)-(6), until  $i \leq s$ .
- 8: Total ( $k \times s$ ) number of centers are selected for  $s$  classes.

where  $G^T$  is the transpose of  $G$ ;  $(G^T G)^{-1}$  is the inverse matrix of  $(G^T G)$ ;  $t = [t_1, t_2, \dots, t_l]^T$  is target vector; and  $G = [G_1(x), G_2(x), \dots, G_i(x), \dots, G_m(x)]^T$  is projected input vector  $x$  in hidden space. Note that  $G_i(x)$  is the  $i$ th hidden node (or RBF unit) output and typically,  $G_i(\cdot)$  is chosen as a Gaussian function

$$G_i(x) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma^2}\right) \quad i = 1, 2, 3, \dots, m \quad (3)$$

where  $x$  is an  $n$ -dimensional input vector,  $c_i$  and  $\sigma_i$  are the center vector with the same dimension as  $x$  and spread factor of  $i$ th RBF unit, respectively, and  $m$  is the number of hidden node (or number of centers) of the network. To improve the performance of the proposed technique, this work also proposes a technique to reduce the error between score-generated output and actual target at the training phase by updating the initial estimated weights of RBF network using LMS algorithm. The proposed weights update technique is given in Algorithm 2. The updated weight matrix and selected centers are stored at the end of training phase and used for testing purpose.

#### 4 Experimental results and discussions

Performance of the proposed gesture recognition method is tested on two hand gesture databases. The details of Databases I and II are given in Sect. 2. The gesture recognition performance is evaluated on the basis of four statistical indices [24] including classification accuracy (Acc), sensitivity (Sen), positive predictivity (Ppr) and specificity (Spe). To investigate recognition performance, each database is partitioned into training and testing sets (with a total of 1200 randomly selected images of all gestures used for training and the remaining 1200 images for testing). The experiments are separately conducted utilizing LCS [2], block-based [3], combined and selected combined feature sets to evaluate the recognition performance for each feature sets. This work uses GA to select optimized feature subset from the com-

**Algorithm 2** : The proposed weights update technique.

- 1: Initially, set number of iterations  $n = 1$ .
- 2: Set data point's number  $j = 1$ .
- 3: Compute estimated output vector  $y(j)$  by multiplying estimated weights  $w^*(n)$  with  $G(x)$ . Where  $G(x)=[G_1(x), G_2(x), \dots, G_m(x)]^T$  is projected input vector  $x$  at hidden space.
- 4: Compute the error vector which can be defined as  $E_1(j) = t(j) - y(j)$ . Where  $t$  and  $y$  are separately target vector and estimated output vector of the network. Assign  $E(n) = E_1(j)$ .
- 5: Update estimated weights using LMS algorithm [15] as follows

$$w^*(n+1) = w^*(n) + \eta G(x)E(n)$$

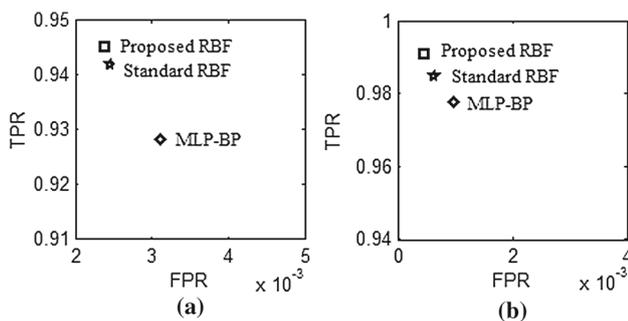
where,  $n$  is iteration number,  $\eta$  is the learning rate parameter. Assign  $w^*(n) = w^*(n+1)$ .

- 6: Increase  $j$  by 1 and repeat steps (3)-(5) until  $j$  reaches  $N$ . where,  $N$  is the total number of data points.
- 7: Compute the mean square error which can be defined as  $MSE(n) = \frac{1}{N} \sum_{j=1}^N E_1^T(j)E_1(j)$ .
- 8: Increase  $n$  by 1 and repeat steps (2)-(7), until  $MSE(n)$  is reduced to 0.0001 or  $n$  reach to 10000.

bined feature set. The detailed setting of parameters for GA is as follows: population size ( $N_p$ ) : 20, crossover probability ( $c_p$ ) : 0.7, mutation probability ( $m_p$ ) : 0.1, uniform crossover, roulette wheel selection and elitism replacement. The length of the chromosome ( $N$ ) is taken as same as length of combined feature set. For Databases I and II, the selected chromosome length is assigned 313 and 421, respectively. The maximum number of generation is assigned 50. To compare the gesture recognition performances of the proposed RBF with MLP-BP as in [1] and standard RBF as in [16], the experiments are conducted with separate feature sets. In the proposed classifier, the optimal number of centers and spread factor are chosen based on the maximizing training performance. The comparative results of static hand gesture recognition with separate feature sets and classifiers are shown in Table 1. The cooperative results in Table 1 indicate that the combined feature set provides better results compared to each of the individual feature sets for both databases using MLP-BP or standard RBF or proposed RBF classifier. Therefore, the static hand gesture is better represented by combined feature set than any of the individual feature sets. The experimental results also show that GA-based optimized feature subset selection method selects 135, 161 and 148 numbers of features for Database I and 175, 214 and 228 numbers of features for Database II using MLP-BP, standard RBF and proposed RBF classifier, respectively. Therefore, GA-based optimum feature subset selection technique reduces the number of features by more than 48 and 45 % for Databases I and II, respectively, which in turn helps to reduce the feature space complexity. The experimental results also indicate that the recognition performance improves due to removal of unwanted features from combined feature set. The results show that average accuracy, sensitivity, positive

**Table 1** Performance of static hand gesture recognition technique (in %) with separate feature sets and classifiers

Classifier	Features	Database I					Database II				
		Features length	Acc	Sen	Ppr	Spe	Features length	Acc	Sen	Ppr	Spe
MLP-BP [1]	LCS [2]	300	99.26	91.08	91.25	99.61	400	99.75	97.00	97.07	99.87
	Block based [3]	13	98.35	80.17	80.16	99.14	21	99.40	92.83	92.85	99.69
	Combined	313	99.39	92.67	92.75	99.68	421	99.79	97.50	97.55	99.89
	Selected Combined	135	99.40	92.83	92.91	99.69	175	99.81	97.75	97.82	99.90
Standard RBF [16]	LCS [2]	300	99.42	93.08	93.18	99.70	400	99.83	98.00	98.04	99.91
	Block based [3]	13	98.97	87.67	88.06	99.46	21	99.72	96.67	96.76	99.86
	Combined	313	99.49	93.83	93.96	99.73	421	99.85	98.25	98.28	99.92
	Selected Combined	161	99.53	94.33	94.44	99.75	214	99.88	98.50	98.53	99.93
Proposed RBF	LCS [2]	300	99.44	93.33	93.43	99.71	400	99.85	98.17	98.20	99.92
	Block based [3]	13	99.03	88.33	88.78	99.49	21	99.74	96.83	96.96	99.86
	Combined	313	99.51	94.17	94.26	99.75	421	99.91	98.92	98.94	99.95
	Selected Combined	148	99.54	94.50	94.58	99.76	228	99.92	99.08	99.10	99.96

**Fig. 4** Comparison of TPR and FPR of three classifiers for **a** Database I and **b** Database II

predictivity and specificity of the proposed RBF classifier with selected features are 99.54, 94.50, 94.58 and 99.76 %, respectively, for Database I and 99.92, 99.08, 99.10 and 99.96 %, respectively, for Database II which are superior to the earlier reported results obtained using MLP-BP as in [1] and standard RBF neural network as in [16] classifier. The gesture image classification performance of the proposed RBF, MLP-BP and standard RBF is also studied provisions of receiver operating characteristic (ROC) curve [25]. On an ROC curve, false positive rate (FPR) and true positive rate (TPR) (or sensitivity) are separately plotted on the X and Y axes. A point in the ROC space is better than other if its location is closer to the upper-left corner. The comparative results of proposed RBF, MLP-BP and standard RBF using selected combined features for Databases I and II are shown in Fig. 4, and these results indicate that the classification performance of proposed RBF is better compared to MLP-BP and standard RBF classifier.

To validate the gesture recognition performance of the proposed method using selected combined features with proposed RBF classifier and compare it with other techniques,

experiments are conducted separately on Databases I and II with a 10-fold cross-validation test. The advantages of cross-validation are: (i) Test sets are independent, and (ii) the reliability of the performance results could be improved. For 10-fold cross-validation test, all of the gesture images in the Databases I and II are divided into 10 parts, where each part contains 240 images with 24 ASL hand alphabet, performed by all the users. At each instance, one part is taken for test dataset and remaining nine parts are served as train dataset. The overall performances of the proposed method and other techniques are shown in Table 2. From Table 2, it is observed that the overall performance of the proposed method in terms of average accuracy, sensitivity, positive predictivity and specificity are 99.57, 94.88, 94.87 and 99.78 %, respectively, for Database I and 99.90, 98.75, 98.77 and 99.95 %, respectively, for Database II. To test the user-independent performance of the proposed method and compare it with other existing techniques, experiments are conducted separately on Databases I and II in another way with a 10-fold cross-validation test. In this study, all of the gesture images in the Databases I and II are divided into 10 parts, where each part contains 240 images with 24 ASL hand alphabet, performed by only one user. Note that in this study, the gesture images of test dataset are completely user independent from the gesture images of train dataset. User-independent overall performances of the proposed method and other techniques are shown in Table 3. From Table 3, it is observed that for user-independent test, the overall performance of the proposed method with respect to average accuracy, sensitivity, positive predictivity and specificity are 98.52, 82.25, 82.13 and 99.23 %, respectively, for Database I and 98.97, 87.67, 87.94 and 99.46 %, respectively, for Database II. The results in Tables 2 and 3 indicate that the recognition performance of the proposed method

**Table 2** 10-fold cross-validation test results (in %) for Databases I and II where train and test dataset contains the gesture images of all users

Database Technique	Database I				Database II			
	Acc	Sen	Ppr	Spe	Acc	Sen	Ppr	Spe
Block-MLP-BP	98.37	80.42	80.32	99.15	99.44	93.25	93.42	99.71
LCS-MLP-BP	99.25	91.00	91.08	99.61	99.77	97.29	97.34	99.88
Combined-MLP-BP	99.41	92.88	92.90	99.69	99.83	98.00	98.01	99.91
Block-SRBF	98.96	87.54	87.78	99.46	99.80	97.54	97.58	99.89
LCS-SRBF	99.42	93.04	93.12	99.70	99.80	97.63	97.65	99.90
Combined-SRBF	99.46	93.50	93.53	99.72	99.85	98.25	98.28	99.92
HU-MLP-BP [1]	94.87	38.42	35.55	97.32	95.63	47.58	46.97	97.72
DCT-KNN [14]	99.26	91.08	91.79	99.61	99.61	95.38	95.49	99.80
NSDS-KLAD [17]	99.15	89.83	89.95	99.56	99.74	96.92	96.96	99.87
Proposed method	99.57	94.88	94.87	99.78	99.90	98.75	98.77	99.95

**Table 3** User-independent 10-fold cross-validation test results (in %) for Databases I and II

Database Technique	Database I				Database II			
	Acc	Sen	Ppr	Spe	Acc	Sen	Ppr	Spe
Block-MLP-BP	97.13	65.58	64.11	98.50	98.39	80.67	81.20	99.16
LCS-MLP-BP	98.01	76.13	75.62	98.96	98.67	84.08	83.52	99.31
Combined-MLP-BP	98.27	79.21	78.94	99.10	98.87	86.46	86.33	99.41
Block-SRBF	97.18	66.17	66.56	98.53	98.42	81.04	81.97	99.18
LCS-SRBF	98.08	77.00	76.63	99.00	98.73	84.75	84.72	99.34
Combined-SRBF	98.28	79.33	79.17	99.10	98.92	87.08	87.05	99.44
HU-MLP-BP [1]	94.39	32.63	26.99	97.07	95.13	41.58	35.46	97.46
DCT-KNN [14]	97.91	74.92	76.50	98.91	97.23	66.71	68.56	98.55
NSDS-KLAD [17]	97.75	72.96	73.26	98.82	98.44	81.25	82.51	99.18
Proposed method	98.52	82.25	82.13	99.23	98.97	87.67	87.94	99.46

is better compared to existing techniques such as Block-MLP-BP, LCS-MLP-BP, Combined-MLP-BP, Block-SRBF, LCS-SRBF, Combined-SRBF, HU-MLP-BP [1], DCT-KNN [14] and NSDS-KLAD [17] for both databases. From experimental results, it is also observed that the gesture recognition performances of the proposed method as well as existing techniques are reduced during user-independent test. This is because the hand geometry and flexibility of fingers vary from user to user and the sign of same gesture also varies from user to user. From the experimental results, it is also observed that HU-MLP-BP [1] provides poorer performance compared to all other methods to recognize 24 ASL hand alphabet, because it uses 7 Hu moment invariants as a feature set which is unable to distinguish large number of gestures having similar shapes. However, the proposed method achieves higher recognition performance because it enjoys following advantages: (i) It uses homomorphic filtering [18] and gray world [19] techniques to compensate the lighting variation of grayscale and color gesture images, respectively; (ii) to segment hand region, this work uses histogram-based Otsu segmentation algorithm for grayscale image and skin color detection-based color segmentation technique for color image; (iii) the proposed image rotation technique makes the

image rotation invariant; (iv) a position- and size-invariant combined feature set carries contour as well as regional information of the hand gesture; (v) GA is used to select feature subset by removing unwanted features from combined feature set; and (vi) an improved version RBF neural network is used here to classify hand gesture images.

The proposed static gesture recognition method is implemented using MATLAB 7.6.0 (R2008a). The proposed method uses GA to select optimized feature subset from the combined feature set; therefore, the training temporal cost of the proposed model is high. However, temporal cost of testing is low because stored optimized chromosome bit pattern is used to select feature subset from combined feature set in testing phase. When the proposed model is executed on a Intel Core i5 computer, processor 3.20 Ghz with 4 GB RAM, Windows 7 platform and as the only application running, it takes an average execution time of 0.333 seconds for recognition of each test gesture image. Note that both databases are constructed incorporating the variation of position, rotation and size of the gesture images under different illumination conditions. Therefore, the recognition performances of the proposed method in the above experimental study are invariant of position, rotation and size of the gesture images along

with illumination invariance. From experimental results in Table 3, it is observed that the proposed method recognizes gesture images with an average sensitivity of 82.25 % for Database I and 87.67 % for Database II with user independence. The proposed model can be used for recognition of ASL hand alphabet in a real-time environment.

## 5 Conclusion

This paper presents a novel static hand gesture recognition method which overcomes the challenges of illumination, rotation, size and position variation of the gesture images. This work uses homomorphic filtering and gray world techniques to compensate the illumination variation of the gesture image and proposes an image rotation technique to make segmented gesture rotation invariant. The proposed method explores a combined feature set, by appending LCS feature set with block-based feature set for better representation of static hand gesture. This work also uses GA-based optimum feature subset selection method to remove unwanted features without reducing performance. In this paper, an improved version RBF neural network is also proposed to classify hand gestures using selected combined features. The experiments are conducted separately on two indigenously developed databases of 24 ASL hand alphabet. The experimental results show that the combined feature set provides better recognition performance compared to each of the individual feature sets. For Databases I and II, the GA-based feature selection technique reduces more than 48 and 45 % unwanted features, respectively, which further improves the recognition performance. The performance of the proposed RBF classifier is superior compared to MLP-BP and standard RBF classifier. The performance of the proposed static hand gesture recognition technique which recognizes gesture based on selected combined features with proposed RBF classifier is better compared to existing techniques such as Block-MLP-BP, LCS-MLP-BP, Combined-MLP-BP, Block-SRBF, LCS-SRBF, Combined-SRBF, HU-MLP-BP [1], DCT-KNN [14] and NSDS-KLAD [17]. The proposed hand gesture recognition model can be used as a user-independent recognition system to recognize 24 static ASL hand alphabet with more than 82 % average recognition sensitivity.

## References

- Premaratne, P., Nguyen, Q.: Consumer electronics control system based on hand gesture moment invariants. *IET Comput. Vis.* **1**(1), 35–41 (2007)
- Gupta, L., Ma, S.: Gesture-based interaction and communication: automated classification of hand gesture contours. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **31**(1), 114–120 (2001)
- Wachs, J.P., Stern, H., Edan, Y.: Cluster labeling and parameter estimation for the automated setup of a hand-gesture recognition system. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **35**(6), 932–944 (2005)
- Bourennane, S., Fossati, C.: Comparison of shape descriptors for hand posture recognition in video. *Signal Image Video Process.* **6**(1), 147–157 (2012)
- Yang, C.-K., Chen, Y.-C.: A HCI interface based on hand gestures. *Signal Image Video Process.* **9**(2), 451–462 (2015)
- Mitra, S., Acharya, T.: Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **37**(3), 311–324 (2007)
- Yang, R., Sarkar, S., Loeding, B.: Handling movement epenthesis and hand segmentation ambiguities in continuous sign language recognition using nested dynamic programming. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(3), 462–477 (2010)
- Bhuyan, M.K., Bora, P.K., Ghosh, D.: An integrated approach to the recognition of a wide class of continuous hand gestures. *Int. J. Pattern Recognit. Artif. Intell.* **25**(2), 227–252 (2011)
- Liu, W., Fan, Y., Li, Z., Zhang, Z.: RGBD video based human hand trajectory tracking and gesture recognition system. *Math. Probl. Eng.* **2015**(863732), 1–15 (2015)
- Sturman, D.J., Zeltzer, D.: A survey of glove-based input. *IEEE Comput. Graph. Appl.* **14**(1), 30–39 (1994)
- Wang, C., Cannon, D.J.: A virtual end-effector pointing system in point-and-direct robotics for inspection of surface flaws using a neural network based skeleton transform. In: *IEEE International Conference on Robotics and Automation*, vol. 3, pp. 784–789 (1993)
- Genç, S., Baştan, M., Güdükbay, U., Atalay, V., Ulusoy, Ö.: HandVR: a hand-gesture-based interface to a video retrieval system. *Signal Image Video Process.* 1–10 (2014). doi:10.1007/s11760-014-0631-x
- Erden, F., Çetin, A.E.: Hand gesture based remote control system using infrared sensors and a camera. *IEEE Trans. Consum. Electron.* **60**(4), 675–680 (2014)
- Shanableh, T., Assaleh, K.: User-independent recognition of Arabic sign language for facilitating communication with the deaf community. *Digit. Signal Process.* **21**(4), 535–542 (2011)
- Haykin, S.: *Neural Networks: A Comprehensive Foundation*, 3rd edn. Prentice-Hall Inc, New Jersey (2007)
- Ng, C.W., Ranganath, S.: Real-time gesture recognition system and application. *Image Vis. Comput.* **20**(13–14), 993–1007 (2002)
- Dedeoğlu, Y., Töreyn, B.U., Güdükbay, U., Çetin, A.E.: Silhouette-based method for object classification and human action recognition in video. In: Huang, T.S., Sebe, N., Lew, M.S., Pavlović, V., Kölsch, M., Galata A., Kisačanin, B. (eds.) *Computer Vision in Human-Computer Interaction*, pp. 64–77. Springer, Berlin (2006)
- Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd edn. Addison-Wesley Longman Publishing Co. Inc, Boston (2001)
- Lam, E.: Combining gray world and retinex theory for automatic white balance in digital photography. In: *Proceedings of 9th International Symposium on Consumer Electronics, 2005 (ISCE 2005)*, pp. 134–139 (2005)
- Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)
- Chai, D., Bouzerdoum, A.: A bayesian approach to skin color classification in ycbcr color space. In: *Proceedings of TENCON 2000*, vol. 2, pp. 421–424 (2000)
- Canny, J.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(6), 679–698 (1986)
- Michalewicz, Z.: *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd edn. Springer, London (1996)
- de Chazal, P., O'Dwyer, M., Reilly, R.B.: Automatic classification of heartbeats using ECG morphology and heartbeat interval features. *IEEE Trans. Biomed. Eng.* **51**(7), 1196–1206 (2004)
- Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* **27**(8), 861–874 (2006)