Scientific
Research
Publishing

An Efficient Framework for Indian Sign Language Recognition Using Wavelet Transform

Mathavan Suresh Anand[1,2], Nagarajan Mohan Kumar[3], Angappan Kumaresan[3]

[1]Anna University, Chennai, India
[2]Department of Computer Science and Engineering, Sri Sai Ram Engineering College, Anna University, Chennai, India
[3]Department of Electronics and Communication Engineering, SKP Engineering College, Anna University, Tirvannamalai, India
Email: suresh.anandm@rediffmail.com

Abstract

Hand gesture recognition system is considered as a way for more intuitive and proficient human computer interaction tool. The range of applications includes virtual prototyping, sign language analysis and medical training. In this paper, an efficient Indian Sign Language Recognition System (ISLR) is proposed for deaf and dump people using hand gesture images. The proposed ISLR system is considered as a pattern recognition technique that has two important modules: feature extraction and classification. The joint use of Discrete Wavelet Transform (DWT) based feature extraction and nearest neighbour classifier is used to recognize the sign language. The experimental results show that the proposed hand gesture recognition system achieves maximum 99.23% classification accuracy while using cosine distance classifier.

Keywords

1. Introduction

The sign language is a fundamental communication tool among people who suffer hearing impairments. In recent years, many tools are designed for sign language recognition using human gestures particularly using hand

gestures. In this section, some recent techniques are reviewed. A new approach for hand gesture recognition is discussed in [1] for ISLR. It is composed of three stages: pre-processing, feature extraction and classification. At first, brightness and contrast of the captured images are adjusted in the pre-processing stage followed by RGB to gray scale conversion. Discrete Cosine Transform (DCT) features are extracted from the segmented skin region, where pixel-based or region based extraction method is used. Finally self organizing map based neural network classifier is employed for sign and numeral recognition.

Wavelet transform based ISLR system is implemented in [2] using video sequences. In pre-processing the given RGB video is converted into gray colour space and high frequency noises are removed by Gaussian low pass filter. Then, segmentation is performed using canny edge detection, DWT and Otsu's thresholding. Elliptical Fourier descriptors are employed for hand gesture feature extraction. Finally, sugeno type fuzzy inference system is used for hand gesture recognition by using extracted features. KNN search based sign language recognition is described in [3]. Initially, the given hand gesture image is transformed into gray colour space. Then, edge detection is performed using canny and the discontinuities among the edges are joined based on KNN classifier. Finally fuzzy rule set is used for sign recognition.

Generic Cosine Descriptor (GCD) based Taiwanese sign language recognition is discussed in [4]. To extract hand shape representation, Generic Fourier Descriptor (GFD) is used as region based descriptor. GCD is generated by using DCT instead of DWT in GFD. Finally Euclidean distance classifier is exploited for sign language recognition. Arabic sign language recognition is reviewed in [5] by means of automatic gestures translation into manual alphabets. In the pre-processing stage, the video sequences are converted into video frames, then skin and background regions are differentiated. Edge detection is done for selected frame followed by feature vector calculation, where distance between the orientation points is calculated as feature vectors. Two different classifiers, namely minimum distance classifier and multilayer perceptron neural network classifier are employed for recognition.

Adaptive Neuro Fuzzy Inference System (ANFIS) based Arabic sign language recognition system is implemented in [6]. Initially, image denoising is performed using median filter and iterative thresholding is used to segment the hand gesture and then smoothed by Gaussian smoothing. Translation, scale, and rotation invariant gesture feature are extracted to train ANFIS for classification. Vision features based sign language recognition is designed in [7]. Three vision features are extracted from four components such as hand shape, place of articulation, hand orientation, and movement. To obtain features from hand configuration and hand orientation, kurtosis position and principal component analysis are employed. Hand movement is represented by motion chain code. Hidden Markov model classifier is used to recognize the corresponding sign language.

A pair of data gloves is used for motions of hands and fingers detection for Korean sign language (KSL) in [8]. Based on hand gesture, KSL recognition is performed and transformed into normal Korean text. To achieve better recognition, fuzzy min-max neural network classifier is employed. Hand gesture features based ISLR is discussed in [9]. Hand region is segmented using YCbCr skin colour model. Wavelet packet decomposition, principle curvature based region detector and complexity defects algorithms are performed to extract shape, texture and finger features of each hand. Finally, each hand gesture is recognized using multiclass nonlinear support vector machine. Similarly, dynamic gestures are recognized by employing dynamic time warping with the trajectory feature vector.

Genetic Algorithm (GA) based hand gesture recognition is implemented for ISLR in [10]. It is composed of four stages; real time hand tracking, hand segmentation, feature extraction and gesture recognition. Hand tracking is performed by Camshift method and segmentation is achieved in HSV (Hue, Saturation, and Intensity) colour model. Finally, gesture recognition is attained by GA algorithm, where single and also double handed gestures are also recognized. A review of the most recent works related to hand gesture interface techniques such as glove based technique, vision-based technique, and analysis of drawing gesture is discussed in [11]. DCT based sign language recognition system is illustrated in [12] using hand and head gestures. In order to extract features from video, DCT features and features from corresponding binary area are taken into account. To recognize gesture using extracted features, simple neural network classifier is employed.

In this paper, an efficient ISLR system based on DWT and nearest neighbour classifier is proposed. The previously described works in the literature focuses mainly the shape descriptor, edge operators and DCT for hand gesture recognition. Recently, DWT is applied to all research areas where image or signal is an input. This study uses DWT for extracting texture information of hand gesture along with the area of hand gesture as binary features for the analysis of Indian sign language. The remainder of this paper is organized as follows. The system

design for ISLR system is explained in Section 2. The main three modules: segmentation, feature extraction and classification are discussed in Section 3, 4 and 5 respectively. Section 6 describes the experimental results and conclusion is made in Section 7.

2. System Design

The proposed ISLR system framework is shown in **Figure 1**. Before extracting features and classification, a segmentation approach is applied to segment only the hand gesture images. Then DWT based features are extracted and given to a simple KNN classifier for sign recognition.
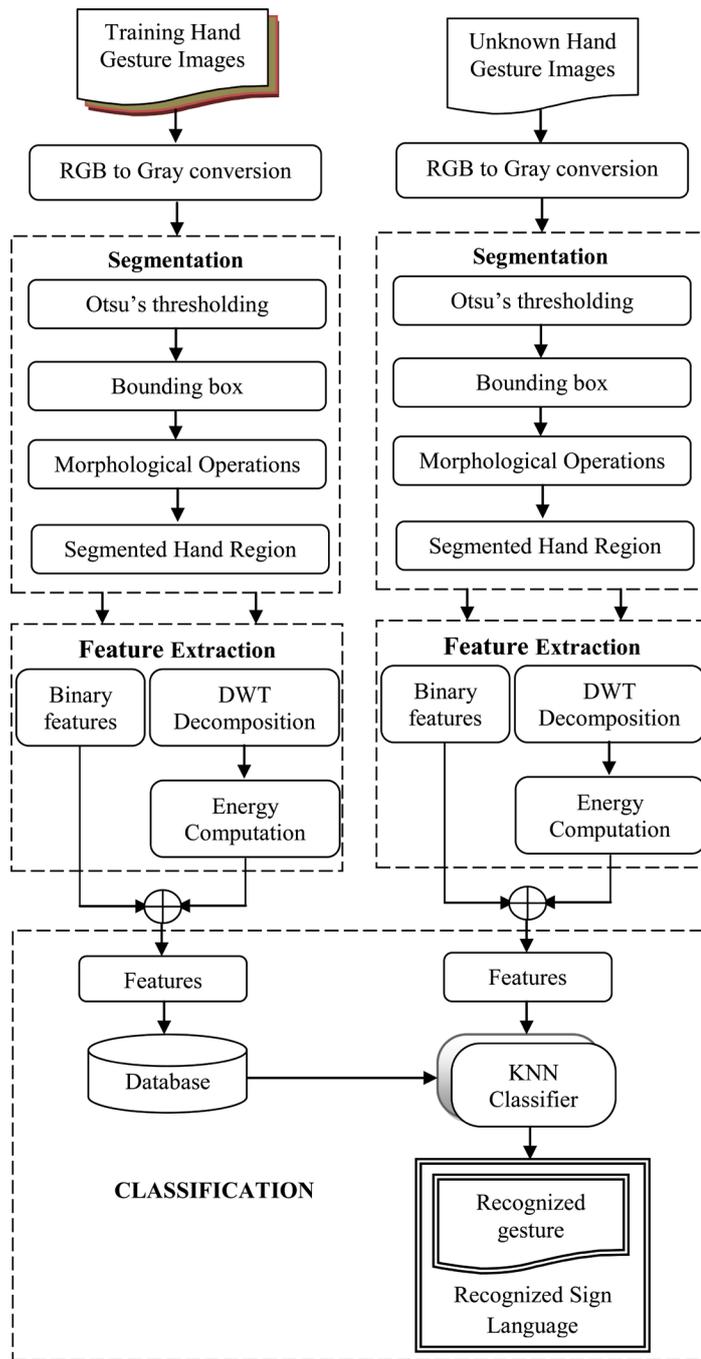

Figure 1. System block diagram for sign language recognition.

In order to implement the proposed sign recognition system effectively by using hand gesture images, a pre-processing step is applied to the original captured image. Because the captured image has many background information's that are irrelevant for sign recognition. The pre-processing includes colour space conversion, thresholding and morphological operations to segment the hand gesture from the whole image. The process of sign recognition from the hand gesture images are divided into three modules; segmentation or pre-processing, feature extraction, recognition or classification. In the pre-processing module, only the hand gesture is segmented from the whole image using Otsu thresholding [13] and morphological operations. The detailed procedure for the segmentation of hand gesture is explained in Section 3. After segmentation, the texture features are extracted using DWT and then KNN classifier is used to classify the type of hand gesture using the extracted features. The feature extraction and classification procedure are briefly explained in Section 4 and 5 respectively.

3. Segmentation

The key process of any sign recognition system is the segmentation of appropriate hand image region because it separates task-relevant processed region from the image background prior to subsequent feature extraction and recognition stages. **Figure 2** shows the outputs of segmentation process.

In order to perform segmentation, at first the given RGB hand gesture images are transformed into gray scale. To facilitate the segmentation process simple, Otsu's thresholding [13] is employed in this study. It is a global thresholding approach depends on gray intensity value of the images and selects the threshold by minimizing the within-class variance of the two groups of pixels separated by the thresholding operator [13]. The hand gesture region is segmented from its background region by thresholding it using the computed threshold from Otsu method. After thresholding, connected component analysis is applied to construct a rectangle containing only the hand gesture image. Then, the thresholded image is cropped from the acquired image using the formed rectangle. As the cropped image having many holes, morphological flood fill operation is used to fill the regions. Finally, the cropped image is superimposed on the original acquired image to get the hand gesture image.

4. Feature Extraction

Wavelet technology is frequently applied in many image processing areas such as feature extraction, super-

(a) Input image   (b) Gray scale image   (c) Thresholded image

(d) Bounding box image   (e) Cropped image   (f) Morphological processed image
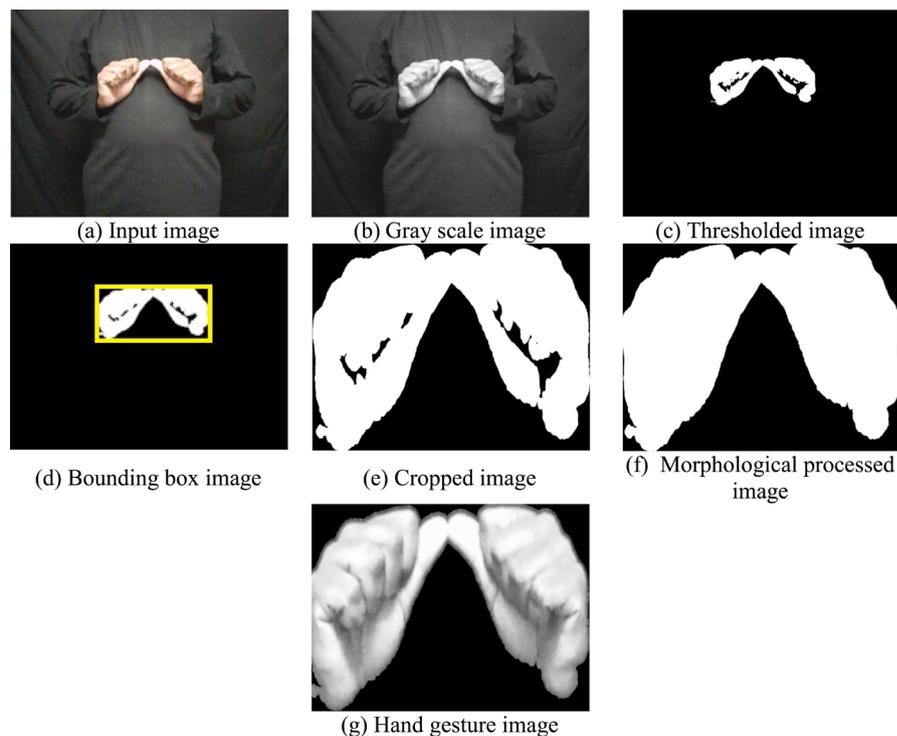
(g) Hand gesture image

Figure 2. Output images at each stages of segmentation process.

resolution, compression, recognition, and de-noising. It transforms a time domain of a signal into a time-frequency domain. Using DWT for sign recognition system, the features can be represented as a sum of wavelets at different time shifts and scales. A time varying function $f(t) \in L^2(R)$ can be represented using scaling function ($\phi(t)$) and wavelet function ($\psi(t)$) as follows [14]:

$$
\begin{aligned}
f(t) &= \sum_k C_0(k)\phi(t-k) + \sum_k \sum_{j=1} d_j(k) 2^{-j/2} \psi\left(2^{-j}t-k\right) \\
&= \sum_k C_{j0}(k) 2^{-j0/2} \phi\left(2^{-j0}t-k\right) + \sum_k \sum_{j=j0} d_j(k) 2^{-j/2} \psi\left(2^{-j}t-k\right)
\end{aligned}
\tag{1}
$$

where $L^2(R)$ stands for square integrable function space, which satisfies $\int_{-\infty}^{+\infty}|f(t)|^2 \mathrm{d}t < +\infty$. C_0 and d_j represent the scaling coefficient (approximation coefficients) at scale 0, and detail coefficients at scale j respectively. The variable k is the translation coefficient and scales denote the different (low to high) decomposition bands. Mallet [15] developed a wavelet filter bank approach which is shown in **Figure 3**.

In **Figure 3**, H is a high-pass filter, G is a low-pass filter, represents down-sampling, $cA1$ and $cD1$ are the approximation and detail wavelet coefficients at level 1. To obtain 2D wavelet decomposition, rows are decomposed by 1D wavelet transform and then columns are decomposed which produces one approximation and three detailed sub-bands. Further decomposition is applied on the approximation sub-bands.

In order to obtain texture information from the corresponding hand gesture images, the segmented hand region is subjected into DWT decomposition. The level of decomposition used in this study varies up to 7. Subsequently, energy is computed from each wavelet sub-bands as feature vector by using Equation (2).

$$
E_k = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \left| \text{wavlet\_subband}_k(i,j) \right|
\tag{2}
$$

where $\left| \text{wavlet\_subband}_k(i,j) \right|$ is the pixel value of the kth sub-band; M is width of the kth sub-band and N is the height of the kth sub-band. In addition to wavelet energy features, to achieve accurate recognition system, the area of segmented hand gesture region is also used as one of the features. The course of action is repeated for all training hand gesture images and the resultant features are stored in a database for further action.

5. Classification

In order to recognize the hand gesture, a simple nearest neighbour classifier is used. As it is an instant based classifier, it does not require explicit generalization. It instantly creates hypotheses for classification from the given training instances. The hand gesture image, which has to be classified (test sample) into one of the predefined gestures undergoes the same feature extraction technique as already described. **Figure 4** shows the classification module of the proposed ISLR system.

The extracted features of test image and stored training image feature database are fed into the nearest neighbour classifier. First, the distances between the test features and each of the training feature vectors are calculated. Then, test sample is recognized into one of the training gesture class by computing minimum distance between them. In this study, four distance measures such as Euclidean, city block, correlation and cosine are analyzed. Let us consider, the features of unknown hand gesture image are $x = (x_1, x_2, x_3, \cdots, x_n)$ and training features in the database are $y = (y_1, y_2, y_3, \cdots, y_n)$. The computation of the distance measures are as follows:

$$
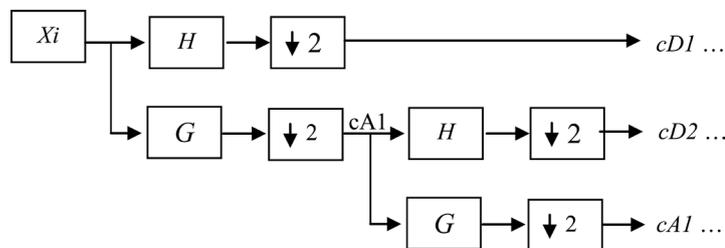\text{Euclidean} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}
\tag{3}
$$

Figure 3. DWT algorithm for input signal *Xi*.

$$\text{City block} = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n| \tag{4}$$

$$\text{Cosine} = \frac{\sum_i x_i y_i}{\sqrt{\sum_i x_i^2}\sqrt{\sum_i y_i^2}} \tag{5}$$

$$\text{Correlation} = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{\left[N\sum x^2 - (\sum x)^2\right]\left[N\sum y^2 - (\sum y)^2\right]}} \tag{6}$$

6. Results and Discussions

In order to validate the proposed ISLR system, 13 English alphabet hand gestures images such as A, B, C, D, G, K, L, P, T, V, W, X and Z are captured by using C170 Logitech 5 MP digital camera in image database. These images are captured at above 105 cm from ground position and 80 cm horizontal distance from camera and person, where person wore a black colour jacket to differentiate hand gesture from background area. In this study, 50 images are captured for each letter and totally 650 images are acquired. **Figure 5** shows the captured hand gesture images.
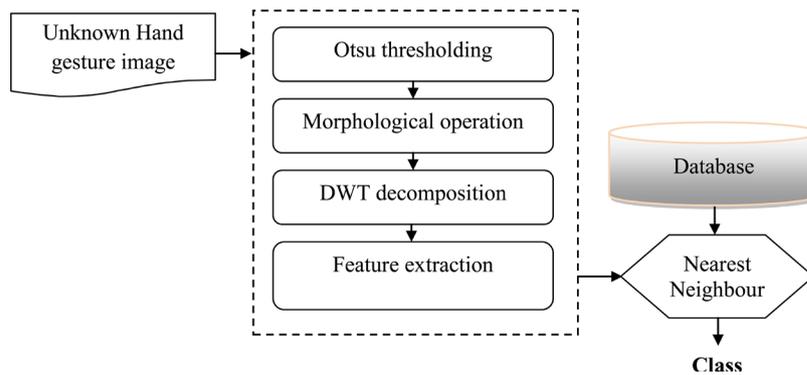

Figure 4. Proposed ISLR system-classification module.
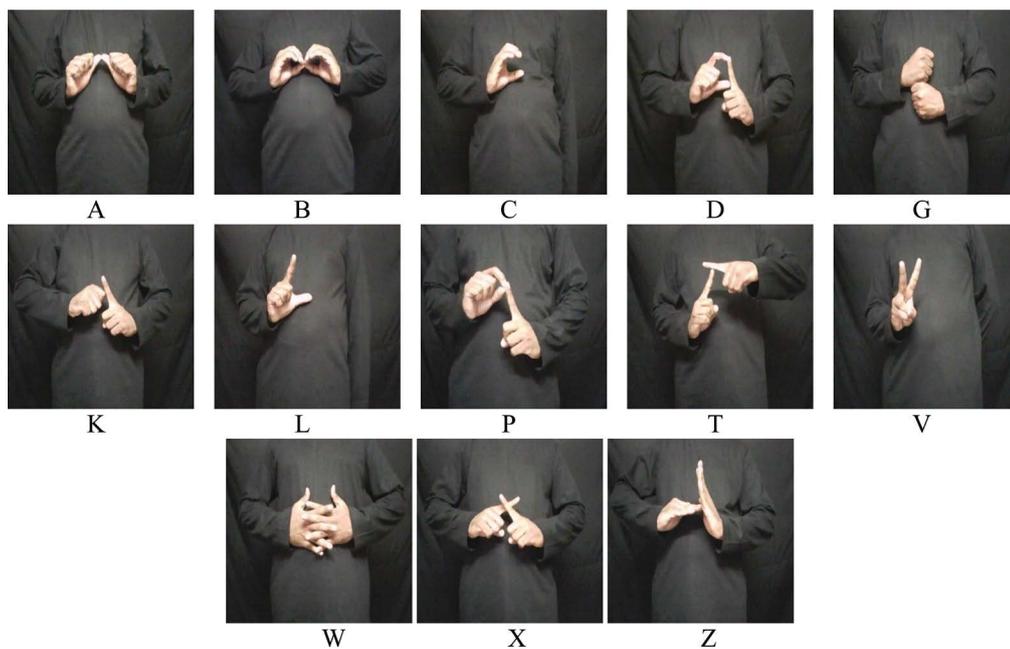

Figure 5. Captured hand gesture images of 13 English alphabets.

The performance of the proposed hand gesture recognition system is analyzed using classification accuracy. The predictive accuracy is computed as follows:

$$\text{Predictive Accuracy } (\%) = \frac{\text{Number of correctly classified hand gesture images}}{\text{total nunber of tested images}} *100 \tag{7}$$

To compute the accuracy of the proposed system, the available hand gesture images of each English alphabet is randomly split into two sets having same number of images. This random split is done 10 times and each time the classification accuracy is computed. **Tables 1-4** shows the classification accuracy of the proposed system using Euclidean, city block, cosine and correlation distance measure up to the 7th level of DWT decomposition.

It is observed from **Tables 1-4** that cosine and correlation distance measures provide 99.69% at the 7th level DWT decomposition. For the same partitions, the city block and Euclidean distance provide only 94.46% and 89.54% respectively. It is also noted that up to the 5th level of DWT decomposition, less than 70% accuracy is attained. As DWT decomposition level increases, the accuracies also increase due to the fact that more texture information is captured at higher level of decomposition. To graphically represent the performance of the proposed ISLR system, average classification accuracy is used. It is the average accuracy of 10 partitions. The average classification accuracies achieved by the distance measures is shown in **Figure 6**.

It is observed from **Figure 6** that the proposed ISLR system achieves over 99% accuracy while using cosine and correlation distance measure. The accuracy of the proposed system for all distance measures are less than 60% at 3rd level of decomposition. However, the accuracy of the system is increased for higher level of decomposition and the maximum accuracy is obtained at the 7th level of decomposition. The average classification accuracies at 7th level of DWT features by the distance measures is shown in **Figure 7**.

Table 1. Predictive accuracies using Euclidean measure.

Partitions	DWT decomposition Level						
	1	2	3	4	5	6	7
1	51.38	52.00	52.62	53.23	60.00	68.92	84.92
2	49.54	50.15	50.15	52.31	56.92	65.23	86.15
3	49.23	49.54	49.85	52.00	55.38	64.62	84.92
4	47.08	47.69	47.08	48.92	52.62	64.00	81.23
5	48.00	48.00	48.92	50.46	55.38	63.38	87.08
6	54.46	55.69	55.69	57.54	62.77	72.00	86.77
7	49.85	49.85	49.85	50.77	53.85	61.85	83.38
8	53.85	53.85	54.15	56.92	60.62	68.31	**89.54**
9	48.31	48.31	48.62	50.77	57.23	66.15	86.15
10	54.46	54.46	55.08	56.62	62.15	70.15	88.00

Table 2. Predictive accuracies using city block measure.

Partitions	DWT decomposition Level						
	1	2	3	4	5	6	7
1	51.38	52.62	54.46	57.54	62.46	73.54	87.38
2	49.85	50.46	51.38	53.85	59.08	69.85	93.54
3	49.85	50.15	51.08	53.54	58.15	70.46	88.92
4	46.46	47.08	47.38	49.85	56.00	69.85	85.23
5	48.31	49.23	50.77	52.62	57.85	69.85	93.23
6	55.08	56.31	57.54	60.31	65.23	75.38	89.54
7	49.85	50.46	50.46	52.31	57.54	69.23	91.08
8	54.77	55.38	55.38	59.38	64.92	74.77	**94.46**
9	49.23	49.85	51.38	52.92	62.15	69.85	90.15
10	54.46	55.08	56.31	59.38	65.85	72.62	92.00

Table 3. Predictive accuracies using cosine measure.

Partitions	DWT decomposition Level						
	1	2	3	4	5	6	7
1	47.69	49.85	51.38	56.31	74.77	90.15	99.08
2	46.15	45.23	49.23	51.38	72.00	91.69	99.69
3	47.08	48.62	51.08	57.54	73.85	88.00	98.77
4	47.38	48.31	49.85	56.00	72.31	89.23	99.38
5	44.92	45.54	50.15	57.54	70.77	90.77	99.38
6	53.54	52.00	55.08	61.54	76.62	88.62	99.08
7	49.85	50.15	49.85	59.69	69.85	91.08	98.77
8	56.92	57.54	59.08	61.85	79.08	91.69	**99.69**
9	49.54	50.46	50.77	58.46	71.08	89.54	99.08
10	49.54	50.77	53.23	60.00	72.62	90.77	99.38

Table 4. Predictive accuracies using correlation measure.

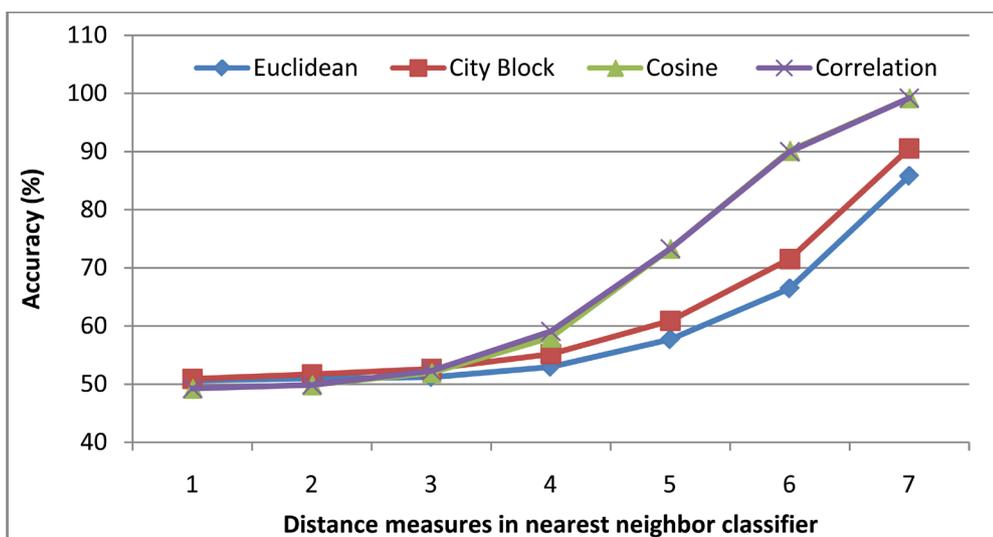| Partitions	DWT decomposition Level						
	1	2	3	4	5	6	7
1	48.62	49.23	52.31	56.62	74.46	89.85	99.08
2	45.85	45.85	48.31	51.38	71.08	91.38	99.69
3	46.46	48.62	52.00	60.62	74.46	88.00	98.77
4	47.69	48.31	49.85	57.23	72.00	88.92	99.38
5	44.92	46.46	49.85	60.31	71.69	89.54	98.77
6	53.23	53.85	54.77	62.46	75.69	88.62	99.08
7	49.23	49.23	51.08	58.77	70.15	91.08	99.08
8	57.54	57.54	59.69	62.15	78.46	92.00	**99.69**
9	49.54	49.23	51.69	59.38	71.69	89.54	99.08
10	49.54	50.46	53.54	61.85	73.54	90.77	99.38

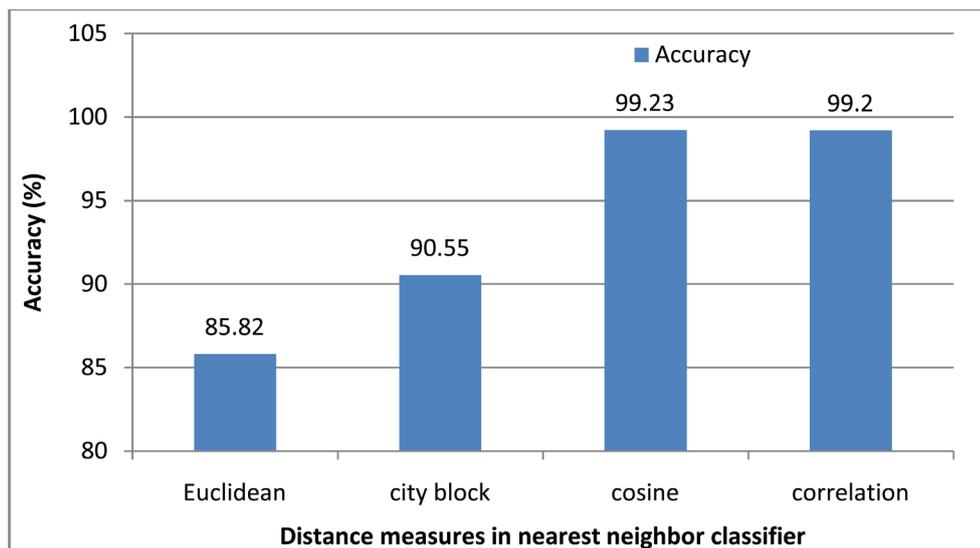**Figure 6.** Average classification accuracies vs. decomposition levels.

Figure 7. Maximum average classification accuracy achieved by the proposed system.

7. Conclusion

An automated ISLR system based on DWT is presented in this paper. DWT sub-band energies extracted from the hand gesture images are used as features along with the area of the segmented hand gesture region. Otsu thresholding and morphological operators are used for the segmentation procedure. DWT is applied up to the 7th level for computing sub-band energy features in order to obtain the best feature set. The nearest neighbour classifier used for the classification provides 99.23% accuracy while using the cosine distance metric. The proposed framework is centralized the efforts of hand gesture and computer vision algorithms which in turn develop a low cost and effective ISLR system. The proposed system can be further developed to recognize all alphabets in English.

References

[1] Tewari, D. and Srivastava, S.K. (2012) A Visual Recognition of Static Hand Gestures in Indian Sign Language Based on Kohonen Self-Organizing Map Algorithm. *International Journal of Engineering and Advanced Technology* (*IJEAT*), **2**, 165-170.

[2] Kishore, P.V.V. and Kumar, P.R. (2012) A Video Based Indian Sign Language Recognition System (INSLR) Using Wavelet Transform and Fuzzy Logic. *IACSIT International Journal of Engineering and Technology*, **4**, 537-542. http://dx.doi.org/10.7763/IJET.2012.V4.427

[3] Kalsh, A. and Garewal, N.S. (2013) Sign Language Recognition System. *International Journal of Computational Engineering Research*, **3**, 15-21.

[4] Cooper, H., Holt, B. and Bowden, R. (2011) Sign Language Recognition. In: Moeslund, T.B., *et al.*, Eds., *Visual Analysis of Humans*, Springer, London, 539-562. http://dx.doi.org/10.1007/978-0-85729-997-0_27

[5] El-Bendary, N., Zawbaa, H.M., Daoud, M.S. and Nakamatsu, K. (2010) ArSLAT: Arabic Sign Language Alphabets Translator. *IEEE International Conference on Computer Information Systems and Industrial Management Applications* (*CISIM*), Krackow, 8-10 October 2010, 590-595.

[6] Al-Jarrah, O. and Halawani, A. (2001) Recognition of Gestures in Arabic Sign Language Using Neuro-Fuzzy Systems. *Artificial Intelligence*, **133**, 117-138. http://dx.doi.org/10.1016/S0004-3702(01)00141-2

[7] Zaki, M.M. and Shaheen, S.I. (2011) Sign Language Recognition Using a Combination of New Vision Based Features. *Pattern Recognition Letters*, **32**, 572-577. http://dx.doi.org/10.1016/j.patrec.2010.11.013

[8] Kim, J.S., Jang, W. and Bien, Z. (1996) A Dynamic Gesture Recognition System for the Korean Sign Language (KSL). *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, **26**, 354-359. http://dx.doi.org/10.1109/3477.485888

[9] Rekha, J., Bhattacharya, J. and Majumder, S. (2011) Shape, Texture and Local Movement Hand Gesture Features for Indian Sign Language Recognition. *IEEE 3rd International Conference on Trends in Information Sciences and Computing* (*TISC*), Chennai, 8-9 December 2011, 30-35. http://dx.doi.org/10.1109/tisc.2011.6169079

[10] Ghotkar, A.S., Khatal, R., Khupase, S., Asati, S. and Hadap, M. (2012) Hand Gesture Recognition for Indian Sign Language. *IEEE International Conference on Computer Communication and Informatics*, Coimbatore, 10-12 January 2012, 1-4. http://dx.doi.org/10.1109/iccci.2012.6158807

[11] Ong, S.C. and Ranganath, S. (2005) Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**, 873-891. http://dx.doi.org/10.1109/TPAMI.2005.112

[12] Paulraj, M.P., Yaacob, S., Desa, H., Hema, C.R., Ridzuan, W.M. and Majid, W.A. (2008) Extraction of Head and Hand Gesture Features for Recognition of Sign Language. *IEEE International Conference on Electronic Design*, Penang, 1-3 December 2008, 1-6. http://dx.doi.org/10.1109/iced.2008.4786633

[13] Otsu, N. (1979) A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems*, *Man*, *and Cybernetics*, **9**, 62-69. http://dx.doi.org/10.1109/TSMC.1979.4310076

[14] Liu, Y. (2009) Wavelet Feature Extraction for High-Dimensional Microarray Data. *Neurocomputing*, **72**, 985-990. http://dx.doi.org/10.1016/j.neucom.2008.04.010

[15] Mallat, A.S. (1989) Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, **11**, 674-693. http://dx.doi.org/10.1109/34.192463