A Vision Based Approach for Pakistan Sign Language Alphabets Recognition

NabeelSabir Khan (Corresponding Author)

School of Science and Technology (SST),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail: nabeel.bloch@umt.edu.pk

Adnan Shahzada

Politecnico di Milano, Milan, Italy

E-mail: shahzada@elet.polimi.it

Saleem Ata

School of Engineering (SEN),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail: saleemata@umt.edu.pk

Adnan Abid

School of Science and Technology (SST),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail: adnan.abid@umt.edu.pk

YaserDaanial Khan

School of Science and Technology (SST),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail: yaser.khan@umt.edu.pk

Muhammad ShoaibFarooq

School of Science and Technology (SST),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail: shoaib.farooq@umt.edu.pk

Dr. M. TahirMushtaq

School of Engineering (SEN),

University of Management and Technology, Lahore

PO box 54770, Lahore, Pakistan

Tel: +9242111300200 E-mail:tahir.mushtaq@umt.edu.pk

Inayat Khan

Centre of Excellence in Science and Applied Technologies, Islamabad, Pakistan

E-mail: inayatk@gmail.com

Abstract

There is a persistent communication barrierbetween the deaf and normal community becausea normal person has no or limited fluency with thesign language. A person with hear-impairment hasto express himself via interpreters or text writing.This inability to communicate effectively between thetwo groups affects their interpersonal relationships.There are about 0.24 million Pakistanis who are eitherdeaf or mute and they communicate through PakistanSign Language(PSL). In this research work a systemfor recognizing hand gestures for Pakistan Sign Languagealphabets in unimpeded environment is proposed. A digitalcamera is used to acquire PSL alphabet's images withrandom background. These images are preprocessedfor hand detection using skin classification filter. Thesystem uses discrete wavelet transform (DWT) for featureextraction.Artificial neural network(ANN) withbackpropagation learning algorithm is employed torecognize the sign feature vectors. The dataset contains500 samples of Pakistan Sign Language alphabets withvarious background environments. The experimentsshow that the classification accuracy of the proposedsystem for the selected PSL alphabets is 86.40%.

Keywords: Pakistan Sign Language (PSL), Discrete Wavelet Transform (DWT), Computer Vision, Artificial Neural Network (ANN), Skin Classification Filter, Back Propagation Learning Algorithm

1. Introduction

Dumb and deaf people face great difficulty while interactingwith normal people. As the hearing impaired ordeaf people cannot talk like normal people, so they aredependent on a kind of visual communication most of thetimes. A sign language is a way of communication thatuses visual modality instead of acoustic sound patterns forexchanging information[1]. People with hearing and speechimpairment use different sign and gestural languages intheir daily face-to-face communication. There are differentmutually unintelligible sign languages around the world,because the languages were developed independently ofother deaf communities. American Sign Language, BritishSign Language, French Sign Language, Australian SignLanguage and many others have developed in various deafcommunities. Similarly, Pakistan Sign Language(PSL) isdeveloped in Pakistan by the deaf and dumb communitywhich has its own vocabulary and syntax.

A sign language recognition system can be useful inbridging the communication gap between hearing impairedand the normal people. Because normal people are usuallyunaware of sign language grammars, It is very difficultfor them to understand what a dumb or deaf person istrying to communicate. As a result,communication of adumb person is usually limited only within the family orthe deaf community[2]. A sign language recognition systemmay facilitate the deaf community by translating the signlanguage to natural language. We propose a mechanismthrough which signs can be captured, recognized andtranslated to the text. Researchers have generally usedeither the direct devices or the computer vision basedinterfaces for Sign Language recognition systems. Thedirect devices include data gloves, styli and other positiontracking devices[3], while vision based approaches use acamera to capture the hand gestures and movements. Themain advantage of using computer vision based approachis that the user need not wear or use any complicated andexpensive device.

Researchers have applied numerous techniques such ashidden markov model, principal component analysis, statisticalmeasures and artificial neural networks. M. AtiqurRahmanet. al. have used height, area, centroid, anddistance of the centroid from the origin (top-left corner)of the image are used as features and then extractedfeatures are used to train a Back propagation NN[4]. A.Giegal has used principal component analysis to recognizeAmerican Sign Language alphabets[5]. A neural networkbased speech translator is developed by Mansi Gupta et.al. that undergoes the process of conversion from RGB toLAB to binary form. Then the number of black pixels foreach block is computed and saved. After this the Euclideandistance between the signed input and stored data iscompared for classification[6]. SignTutor is an interactivesystem proposed by OyaAranet. al. for Sign Language

Tutoring that uses a glove-based interface[7]. Ali Karamiet. al. have implemented a system that uses statisticalmeasures for dimension reduction and ANN for classificationof Persian sign language[8]. The system proposedby Ayoub Al-Hamadiet. al. uses translation, rotation andscale invariant features for recognizing postures[9].Gesture and sign language recognition has been thefocus of many researchers but unfortunately there is nosignificant work done for Pakistan sign language at themoment. There is a system named "Boltay Hath" that aimsat recognizing Pakistan Sign Language but it uses dataglovesas its interface.[10]. The statistics given by PopulationCensus Organization (Government of Pakistan), thereare more than 3.3 million Pakistanis who are disabled.About 0.25 million among them suffer from hearing losswhich is around 7.4% of theoverall disabled population inthe country[11]. A significant part of the deaf populationis young and sign language recognition system can turnthem into useful human resources for certain positions.

Our contribution in the paper is three fold: 1)Theproposed system is the only computer vision based systemfor PSL recognition to date. 2)The system does not requireany hardware like cyber-gloves that in effect reduces thecost and cumbersomeness of wearing hardware devices.

3)The proposed system works efficiently with randombackground gestures.

This research was conducted to propose an intelligent,robust and efficient system for the dumb and deaf peoplethat will help them to communicate with other people intheir gestural language. The proposed system recognizesthe Pakistan sign language alphabets using image processing,discrete wavelet transforms and artificial neuralnetworks. The organization of the paper is as follows:Section 2 explains Pakistan Sign Language's origin and alphabets.There is a brief introduction of Discrete WaveletTransform in section 3 and section 4 explains the proposedsystem. Experimental results are discussed in section5. The conclusion of whole research work is given in section 6.

2. Pakistan Sign Language

Pakistan Sign Language has its own vocabulary, syntaxand semantics. It has undergone continuous improvementand evolution like all other languages. Spoken languagesof an area have a significant impact on the growth and developmentof the sign language and variety of blends occurwhenever they interact with each other. Signed English is adialect of sign language that is formed by the combination

of British Sign Language and English. Sign Exact English(SEE) is gestural language that matches each spoken wordof American English language. Likewise Signed Urdu hasemerged by the combination of PSL, Urdu and otherregional languages(Punjabi, Sindhi, Pushtu, Baluchi)[10].

Urdu Language consists of 38 alphabets but usuallywe find 37 unique hand gestures for most significantalphabets as shown in Fig. 1. Short messages are sentand received using these handshapes. Hearing-impairedpeople in Pakistan make use of these handshapes fromPSL vocabulary combined with small gestures to expressthe words in Urdu. As Urdu is a combination of languagesso English Sign alphabets are also used with PSL. Forexample, to represent Saturday, ’S’ is represented usingboth hands from English sign alphabets followed by thesign of Saturday. Such variations are bound to exist due tothe existence of different cultural backgrounds in the sameregion through our history. Urdu itself is spoken in manydifferent ways in different regions of Pakistan and it differsin vocabulary, phonology and grammar too. Similarly PSLhas regional variations in many items. There are manyexamples that one sign is acceptable in one region but notpreffered in another region in the same country but thatdoes not conclude that the sign is wrong for that particularregion.

Fig.1.PakistanSignLanguageAlphabets

3. Discrete Wavelet Transform (DWT)

Introduction of the wavelet transform was motivated bythe need of further developments from Fourier Transform(FT). Although, FT gives us the information about thefrequencies present in a signal but does not explain aboutthe locality of the frequency components. The wavelettransform was introduced at the beginning of the 1980sby Morlet and have many applications in the areas asmathematics, physics, signal processing, medical imagingand image processing. The information provided by thecontinuous wavelet transform (CWT) is too redundant toreconstruct the signal and requires a lot of computations.

This is where discrete wavelet transform (DWT) comesinto play that not only provides the sufficient informationto analyze and reconstruct a signal but also reduces thecomputational requirements.

The foundations of the DWT were laid by Croiser, Esteban,and Galand by devising a technique to decomposediscrete time signals back in 1976. Crochiere, Weber, andFlanagan also worked on the similar lines in that year.

They named their analysis scheme as subband coding. Inthe discrete case, filters of different cutoff frequencies areused to analyze the signal at different scales[15]. In DWTthe signal is passed through iterations of low and highpass filters and then rescaling is done. Filtering is used todetermine the signal resolution that represents the amountof detail information in the signal. Scaling on the otherhand is achieved up sampling and down sampling of theoriginal signal [16], [17].

As image is a two dimensional(2D) signal, we can implement2D discrete wavelet transform by applying lowand high pass filters to decompose the image in differentcoefficients and down samplers to reduce the data withoutlosing any of the information. The decomposition results incutting down the spatial resolution in half since only halfthe number of samples now characterizes the entire signalbut because frequency band of signal now contain only halfof the previous bands, the frequency resolution is doubled.

This process can be repeated for further decompositionof the image. Each decomposition level will have halfnumber of samples and double frequency resolution thanthe previous level. This 2D-DWT leads to a decompositionof approximation coefficients at some level in four components:

• The approximation Image:(Both horizontal and verticaldirection trends)

• Vertical Detail Image: (Low frequencies in horizontaland high frequencies in vertical direction)

• Horizontal Detail Image: (High frequencies in horizontaland low frequencies in vertical direction)

• Diagonal detail image: (Detail image in both, horizontaland vertical directions).

The DWT of image I(x,y) of size M x N is:

$$W_{\emptyset}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \emptyset_{j_0,m,n}(x,y) \quad (1)$$

$$W_{\varphi}^{i}(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \varphi_{j,m,n}^{i}(x,y) \quad (2)$$

wherej_0 is an arbitrary starting scale and $W_{\varphi}^{i}(j, m, n)$coefficients define an approximation of I(x,y) at scale j_0.

The $W_{\varphi}^{i}(j, m, n)$coefficients add horizontal, vertical anddiagonal details for scales $j \geq j0$ and i = H, D, V.

Low frequency components of the image constitute theapproximation image whereas high frequency componentsmake the three (Vertical, Horizontal and Diagonal) detailedimages.

4. The Proposed System

The proposed system to recognize Pakistan Sign Language(PSL) alphabets can be divided into two majorphases:

• Feature Extraction

• Classification

4.1. Feature Extraction

The process of Feature extraction consists of the followingcomponents:

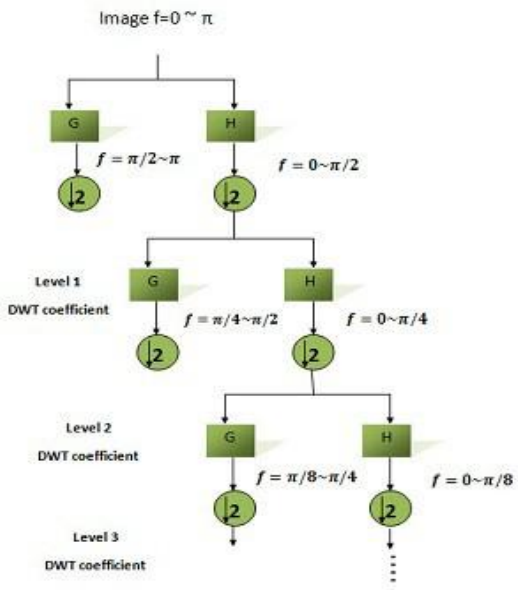

Fig.2.Discrete Wavelet Transform (DWT)baseddecomposition

4.1.1) Illumination and Color Balancing:

The system inputsRGB(Red, Green, Blue) images taken under normallighting conditions using ordinary digital camera sothey may have many inconsistencies and variations amongthem. In real world, the assumption that given an imagewith sufficient amount of color variations, the averagevalue of the R, G, and B components of the image shouldaverage to a common gray value is held very well. Therefore,we can use this assumption to force the images in ourdataset to have the same gray level value for each of therethree(RGB) channels to reduce the effects of luminanceinconsistencies. We have used Gray World algorithm [12]for color balancing and illumination compensation. It calculatesthe average values of R, G and B color componentsof the image and then determines scaling factors for allthree components based on the deviations they have withthe common gray value of the image.

***4.1.2) Hand Segmentation**:*

The next step is to extract thehand out of the image for which color based segmentationis used. The resultant RGB image from the Gray WorldAlgorithm is then converted to YCbCr(Y is luminanceand both Cb and Cr are the Chromatic components)color space [13]. YCbCr form is derived by from thecorresponding RGB space as follows:

$$Y = 0.2989R + 0.5866G + 0.1145B \quad (3)$$

$$Cr = 0.7132(R - Y) \quad (4)$$

$$Cb = 0.5647(B - Y) \quad (5)$$

The following Chromatic components of the YCbCrcolor space image are used to detect pixels that appearto be skin using the range of Cb and Cr that Chai andNgan [14] have found to be the representative of the skincolor.

$$77 \leq Cb \leq 127$$

$$133 \leq Cr \leq 173$$

The skin pixels are marked as blue by setting the pixel[R G B] values to [0 0 255] correspondingly. It is thenconverted to a binary image by setting all the skin pixelsas 1 and non-skin pixel as 0. A 4x4 median filter is appliedthen to fix the isolated false classified skin pixels.TheFigure.3 shows the result of segmentation.

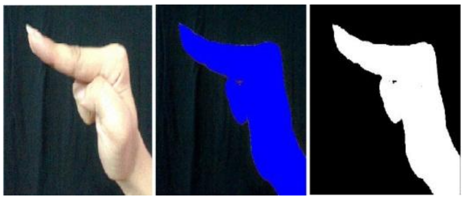

Fig.3.ColorBaseSegmentation

4.1.3) Cropping and Resizing:

Cropping is done by eliminatingall the zero valued rows and columns from thebinary images and then the images are resized to 300x400pixels as shown in the Figure. 4.

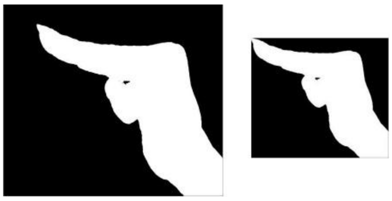

Fig.4.CroppingandResizing

4.1.4) Discrete Wavelet transform:

Our system makes useof DWT property of retaining the same information componentdespite of less number of samples to reduce thedimensions and extract features out of an image. Haarwavelet transform was used to derive the interest pointsfrom the sign images because it shows better results whiledescribing the human body parts[18]. As we are usingartificial neural networks for the classification, it is notpossible to use all the image elements(pixels) as the input.Therefore, we have used 2D-DWT at multiple levels toreduce the number of inputs to the neural network. Wehave experimented with different levels of decompositionthrough DWT to find the best coefficient matrices thatmay produce the coefficients nearest to the original onewhen reconstructed. In this paper, we have used the coefficientsof approximation along with the detail coefficientson the 6th level. Feature extraction through DWT can bevisualized by the Figure. 5

Fig5. Discrete Wavelet Transform (DWT):Approximation,Horizontal, Vertical andDiagonal detailImages

4.2. Classification

Artificial neural network (ANN) is used as the signlanguage alphabets classification model in our proposedsystem. ANN is an information processing system thattakes its inspiration from biological neural network. Artificialneural network aims to generalize the mathematicalmodels of neural biology and imitates the human learningand cognitive processes. Number of constituent neuronlayers and the pattern of connections between them isreffered as the Architecture of the network. The way connectionstrengths(weights) are adjusted is called learningalgorithm and activation function determines and controlsthe output of the neurons against the set of inputs [19].

4.2.1) ANN Architecture:

A feed forward Multilayer Perceptron(MLP) is used for the classification of alphabets.We used a single hidden layer because we did not find anysubstantial benefit of adding another hidden layer. In caseof neural network, you have to empirically find the rightarchitecture for the classification task at hand. We haverun many simulations with different network structureand finally found the network with 140 input neurons, asingle hidden layer with 75 neurons and 37 output neuronsexhibiting the best results. See Figure. 6.

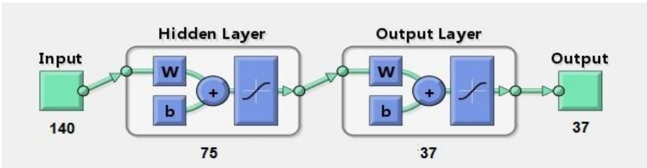

Fig.6.Artificial Neural Network (ANN)Structure

4.2.2) Training Algorithm and Parameterization:

Gradientdescent with momentum and adaptive learning rate backpropagationis used to train the neural network. Thisalgorithm works very well with the noisy and inconsistentdata and improves generalization[20]. Tan-Sigmoid is usedas the activation function for all the layers.

$$T(S) = \tanh(S) = \frac{e^{S} - e^{-S}}{e^{S} + e^{-S}} \quad (6)$$

It generates the output in the range [-1,1].
Training is done in supervised manner and a trainingsample is a pair ((*xp*, *dp*)) where p= 1,...,P. The inputvectors are denoted by (*xp*= *xp*,1, ..., *xp*,*n*) , output isrepresented by (*op* = *op*,1, ..., *op*,*n*) and desired output is(*dp*= *dp*,1, ..., *dp*,*n*). Error in the classification for a singlepattern is defined as:

Ep,*j*= *op*,*j*− *dp*,*j*(7)

The back-propagation algorithm applies a correction$\Delta W_{a,b}$to the synaptic weight $W_{a,b}$which is proportional tothe partial derivative $\frac{\partial E_{p,j}}{\partial W_{a,b}}$that is called gradient descent.

$$\frac{\partial E_{p,j}}{\partial W_{a,b}} = -E_j T_j (S) o_{p,j} \quad (8)$$

The performance of the network is measured by MeanSquared Error(MSE) defined as the average squared errorbetween the network generated outputs and the targetoutputs.

$$MSE = \sum_{p=1}^{P} \sum_{j=1}^{K} E_{p,j}^{2} \quad (9)$$

Random data division is used for validation.

5. Experiments and Simulation Results

A digital camera was used to capture 37 static onehandedPSL alphabets. The experimental data was collectedby varying the hand orientation, its distance andangle with the camera with random backgrounds. Ourdataset contained 500 images of 37 alphabets from which426 images were utilized for training and 74 for testing.The proposed neural network described in the previoussection was simulated using Matlab. The performancecurve and train state of the ANN is shown in Fig. 7 and8.

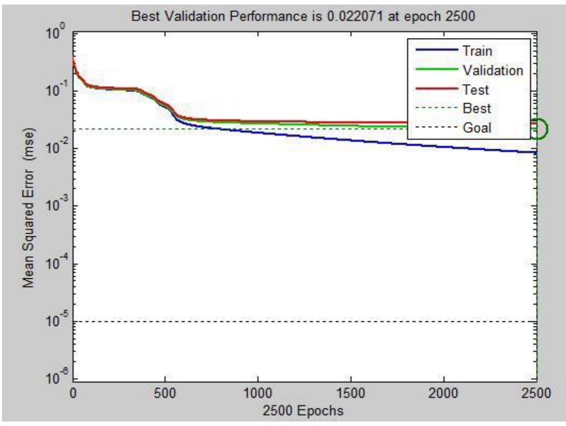

Fig.7.PerformanceCurveduringTraining Process

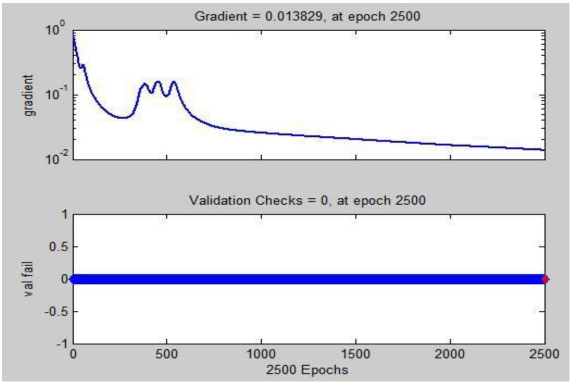

Fig.8.TrainingState

The experimental results for both the training andtesting data are shown in Table. 1.

Table 1: Experimental Results Representation

Data	Total Samples	Correctly Classified	Accuracy (%)
Training	426	365	85.10
Testing	74	63	84.08
Overall	500	425	84.6

It can be observedthat our recognition system was able to classify datawith 84.60% even with the random background. Further,Figure. 9 shows the accuracy with which all individualalphabets are recognized by the system.

Finally the system is compared to the models suggestedby "BoltayHaath". These models include StatisticalInterval Matching (SIM) , Statistical Interval MatchingCombined with LMS and Statistical Interval

MatchingCombined with Democracy "[10]. Our proposed system beats or matches the accuracy of all of the three models mentioned. The result comparisons are shown in Table. 2

Table 2: Comparison with BoltayHaath

Model	Accuracy (%)
Statistical Interval Matching	26
SIM Combined with LMS	84
SIM Combined with Democracy	73
Our Proposed System	84.6

Our results demonstrate that the proposed model hashigh generalization capability and performs efficiently forvaried background environments.

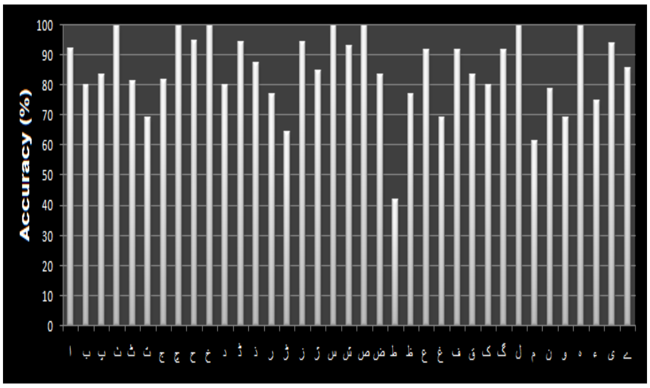

Fig.9.Representation of Classification Resultsof PSL Alphabets

6. Conclusion

An intelligent system for gesture and sign recognition isproposed in this paper to facilitate a better communicationamong normal and the deaf communities. The proposedsystem recognizes the Pakistan Sign Language alphabetswith random backgrounds. A digital camera is used toget images of PSL alphabets instead of data gloves.

The system segments the hand from the images usingskin color tracking and extracts the features by applyingdiscrete wavelet transform (DWT) in the first phase. Inthe second phase, the extracted features are applied toNeural network for classification. We used dataset of 500samples of Pakistan sign Language alphabets in variousbackground environments and the experiments show thatthe proposed system is able to classify the selected PSLsigns with a classification accuracy of 84.60%. The networkis trained using MATLAB NN Toolbox.

References

[1] Armstrong, D.F., Stoke, W.C., Wilcox, S.E., *"Gesture and natureoflanguage"*,CambridgeAcademicPress,1995.

[2] FoezM.Rahim,TamnunEMursalin, NasrinSultana.*"IntelligentSignLanguage VerificationSystem,UsingImageProcessing, ClusteringandNeuralNetworkConcepts"*.InternationalJournal ofMathematicalEngineering (IJME),Vol.1, No.1-2, Jan-Dec,2009,pp43-56

[3]M.Mohandes,S. A-Buraiky,T.HalawaniandS. Al-Baiyat.*"AutomationoftheArabicSignLanguage Recognition"'* InformationandCommunicationTechnologies:FromTheoryto Applications,2004.Proceedings.2004 InternationalConference onOnpage(s):479-480

[4] Md. AtiqurRahman,Ahsan-Ul-AmbiaandMd.Aktaruzzaman*"RecognitionStaticHand Gestures of Alphabet in ASL"*.2011IJCIT,VOLUME02,ISSUE01.

[5] A. Geigel, *"Recognizing American Sign Language Letters UsingPrincipalComponentAnalysis"*.CPSC320Vision IRC/November2004

[6] MansiGupta, MehaGarg,PrateekDhawan, *"SignLanguageto SpeechConverterUsingNeuralNetworks"*,InternationalJournal of Computer Scienceand Emerging Technologies (E-ISSN:

2044-6004)Volume1,Issue3,October2010

[7] OyaAran, Ismail Ari,*"AlexandreBenoitSignTutor: An Interactive System for Sign Language Tutoring"*.IEEEMultiMedia,January-March2009,pp.81-93

[8] Ali Karami,BahmanZanj, AzadehKianiSarkaleh,*"Persian sign language (PSL) recognitionusing wavelet transform and neural networks"*.ExpertSystemswith Applications:An InternationalJournalVolume38Issue3,March,2011

[9] AyoubAl-Hamadi,OmerRashidandBerndMichaelis,*"Posture Recognition usingCombinedStatistical andGeometrical Feature Vectors basedonSVM"'*,InternationalJournal ofInformation andMathematicalSciences6:12010.

[10] AleemKhalid Alvi,M. YousufBin Azhar, MehmoodUsman, SulemanMumtaz, Sameer Rafiq, RaziUr Rehman, IsrarAhmed. *"Pakistan SignLanguageRecognition UsingStatistical Template Matching"'*.WorldAcademy of Science,Engineering andTechnologyvolume3,2005

[11] Population CensusOrgnaization, Govt. Of Pakistanat http://www.census.gov.pk/index.php

[12] G.D. Finlayson, B. ShieleandJ.L.Crowley. *"Comprehensive colournormalization"*.Proc.ofEuropean Conferenceon Computer Vision,(ECCV).Vol.I, 475-490,Freiburg, Germany,1998.

[13] CharlesPoynton, *"DigitalVideoandHDTV"*,Chapter 24,pp.291-292,MorganKaufmann 2003.

[14] D.Chai,andK.N. Ngan,*"Facesegmentationusingskin-color mapin videophoneapplications"*. IEEE Trans. onCircuits and SystemsforVideoTechnology,9(4):551-564,June1999.

[15] RobiPolikar, Tutorial on: *"MultiresolutionAnalysis: TheDiscreteWaveletTransform"*

[16] Mallat, S. (1989), *"A theory for multiresolutionsignal decomposition: thewaveletrepresentation"*;IEEE Pattern Anal. andMachineIntell.,Vol.11,pp.674-693

[17] Misiti, M., Misiti, Y., Oppenheim, G., &Poggi, J.- M. (2006). *"Wavelet toolbox for use with MATLAB"*.<http://www.mathworks.com>.

[18] Dinh, T. B., Dang, V. B., Duong, D. A., Nguyen, T. T., and Le,D. *"Hand gestureclassification usingboostedcascade ofclassifiers"*. In 2006Internationalconference onresearch, innovation andvisionforthefuture. pp.139-144.

[19] Heykin, S. (1999). *"Neural networks: A comprehensive foundation"*,(2nded.).NewJersey:PrenticeHall.

[20] Munib, Q., Habeeb, M., Takruri, B., and Al-Malik, H. A. (2007).American sign language(ASL) recognition basedon Houghtransform andneuralnetworks.Expert Systemswith Applications,pp.24-37.