Sign Language Recognition using Neural Networks

Sabaheta Đogić [1], Gunay Karli [1]

[1]*International Burch University, Francuske revolucije bb, Ilidža, Sarajevo, B&H*

Abstract **– Sign language plays a great role as communication media for people with hearing difficulties.In developed countries, systems are made for overcoming a problem in communication with deaf people. This encouraged us to develop a system for the Bosnian sign language since there is a need for such system. The work is done with the use of digital image processing methods providing a system that teaches a multilayer neural network using a back propagation algorithm. Images are processed by feature extraction methods, and by masking method the data set has been created. Training is done using cross validation method for better performance thus; an accuracy of 84% is achieved.**

Keywords **– Sign Language recognition, Artificial Neural Network, Image Processing.**

1. Introduction

Sign language is the mainlanguage of the Deaf population, their primary communication medium. There is no universal sign language except for Gestuno or IS (InternationalSign) used at International meetings organized by deaf people, where spontaneous signs are chosen with the aim to make the language universal [1]. For every country there is different sign language, which might be very similar so that it causes the communication difficulty for the Deaf people. Sign language is systemic language includingfinger spelling, followed by motions, reading from the lips, and almost all non verbal ways of expression that plays the important role in communication for the Deaf people, by tending to present the language visuallywith the use of signs [2].The sign users, however, still have difficulties in communication with the speaking persons, who are not sign users.
Therefore, sign language recognition (SLR) is a multidisciplinary research area involving pattern recognition, computer vision, natural language processing and linguistics [3].

To solve a problem of communication among Deaf, many researchers have been working on recognition of various sign languages, and the translator is considered accessible to decode sign language into usual spoken language and vice versa. Several systems made to perform these processes are distinguished by the efficiency, feasibility and performance when they are implemented [4].
Glove based systems use data gloves to archive the accurate position of hand gestures as its positions are directly measured. Such system is very expensive and brings much cumbersome experience to the users However, vision based system uses camera for data acquisition and it possesses a problem as these systems should be background invariant, light insensitive, person and camera independent to achieve real time performance. But because of few requirements they are considered easy, natural and less costly compared to glove based approach [5].
A real time sign language recognition system was proposed by CorneliuLungociu [6] who created the data set using the web camera for the 80 alphabet signs from English alphabet, and whose system proved to be accurate 80%. For preprocessing the images computing Fourier Descriptor was performed. Yannick L. Gweth, Christian Plahl and Hermann Ney [7] employed appearance base GHMM system derived from the MLP based feature with the result of 84% accuracy of signs taken by the camera under laboratory conditions. Md. AtiqurRahman, Ahsan-Ul-Ambia and Md. Aktaruzzaman [8] presented a system for recognizing static gestures whose average recognition accuracy was 80, 28%. They used Gaussian filter for removing noise from gray scale and Euclidian Distance as classifier. Hienz et al [9] for a vision based system extracted feature vectors from video frame which recognized 262 signs with accuracy of 94%. Rule based classification was done on the images captured by single video camera using modular frame grabber system. Ch.Raghava Prasad [10] trained an image database of Indian sign language, by preprocessing the images at various levels to improve the recognition rate. Texture features are extracted using Gabor Filter [11] and shape features are extracted using Chan-Vese [12] active contour models, and in this way formed feature matrix. Average recognition rate was around 98.61%, on a typical data set.

After the analysis of the Bosnian sign language status and its practical use, Husnija H. [13] came to the

conclusion that the experts working with the deaf people in Bosnia contain general theoretical knowledge and experience, but they think they are not competent with the practical use. By standardizing the sign language necessary preconditions should be formed for comprehensive learning such as; its structure and grammar rules, its way of use and acquiring a competence in sign language and at the end general theoretical facts and their experience from using sign language.

Bosnian sign language is admitted by the law in the 15 September 2009, thus determines the rights for deaf people to use it. It is conceptually precise language that contains the old signs as well as new innovative signs. It is a language which includes finger spelling, lips reading , motions, sometimes the pronunciation and almost all forms of non-verbal expressions that plays a key role for communication between the users of sign language. In this way language is presented visually, and achieves complete expression [14].

Finger spelling that presents alphabet by one or two hands,is supplement to sign language, and it is very useful when particular meaning has no sign. Sign language interpreters are experts in recognizing grammatical structure and syntax. They have excellent knowledge and skills, which are a key role for the process of translation, but interpreters, are having very expensive and have negative effect on privacy and self independency. This is neither practical nor possible for all situations. For this reason developing a system that can facilitate the automatic translation between sign language and spoken language, can remove barriers in communication.

This paper presents a Bosnian sign language recognizer that has been developed to recognize signs of BSL using Back propagation Neural Network technique. In this recognition system, we have used a set of features for training and testing of ANN. This system allows fast training and intelligent learning of new gestures, and it is designed with the aim to simply use a project without any specialized hardware, or without acquiring from the user to wear special instruments such as gloves, or certain color clothing.

2. Materials and Methods

The proposed system works in three phases to perform sign language recognition. In the first phase by the help of MATLAB [15], the data set of images is being processed for the neural networks to train it. Second phase is creating dataset by the help of masking method to prepare it for training. Furthermore, supervised learning uses back-propagation algorithm in training data. Next phase is cross validation method that divides data into one segment to train a model and the second segment to validate the model, leaving the chance for every sign to appear in both training and testing model, thus achieving greater accuracy. For the image processing we have created (Graphical User Interface) in the MATLAB, a tool for numerical competition and visualization and fourth-generation programming language. GUI consists of 6 push buttons and it shows different phases of the system: input image, part for image processing, and a button for preparing the images to be trained.

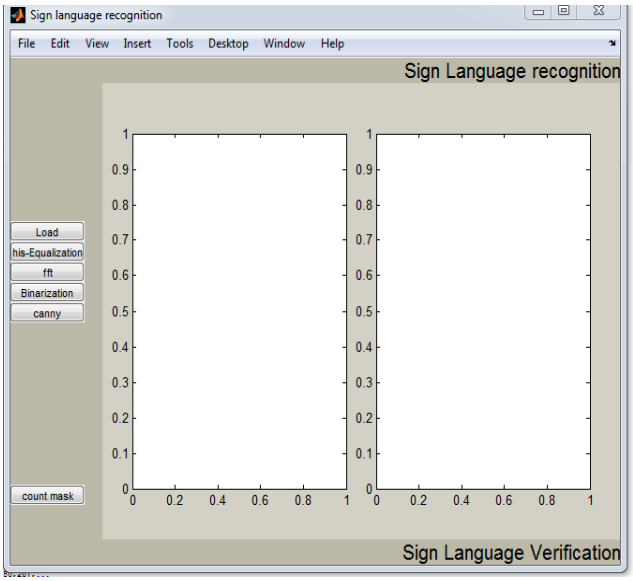

Figure 1. GUI for sign language recognition

Image acquisition

Images are taken by the camera from 3 different users. These users are trained to present every sign for Bosnian language, so that the database of 3 samples for each sign is formed. Images are taken under different lightening conditions, with the same background behind. The image database consists of totally 90 images, 60 images are used for training purpose, while remaining 30 images are used for testing purposes.

Three steps are performed in the processing: color space conversion; Image enhancement using histogram equalization method, Fast Fourier transform; Image segmentation is done using binarization method and finding the edges of the image.

a) Color space conversion: the images are converted to gray scale image thus removing all color information, leaving only the luminance of each pixel. Then all images are resized to 375x500 dimensions allowing algorithm better performance.

b) Image enhancement: Histogram equalization is one of well-known enhancing technique done for contrast enhancement because of its effectiveness. This is very important step in order to avoid the procurement of non-existing artifacts in the output image, by re-mapping the gray levels of image [16].

c) Fast Fourier transform is used to represent image in the frequency domain by breaking an image into real and imaginary components. This is done because it is faster to apply filters in the frequency domain, then in image domain [17].

d) Binarization is used to separate foreground object from background of the image thus determining a gray threshold according to some criteria and assigning each pixel to one class (background or foreground) [18].

e) Image Edge detection is used for detecting forefront of the image by reducing the amount of data and filtering unimportant information to highlight important ones. There is a different algorithm, so we choose to use canny edge detection algorithm. It works on a principle of blurring images to remove noise, and then the local maxima of the edges are marked on the large magnitudes of image. Next step is to detect final edges by smothering the edges that are not connected to a certain edge [19].

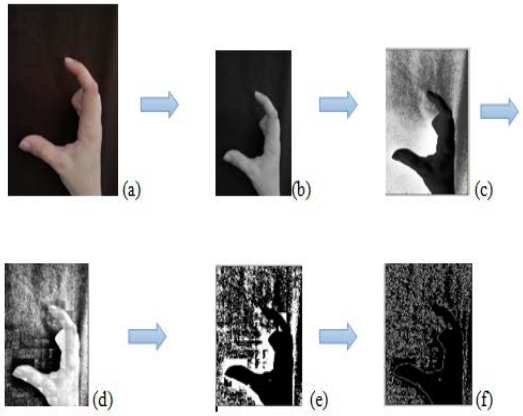

Figure 2. (a) image before color space conversion; (b) gray scale image resized to 375x500 dimensions (c) image after histogram equalization; (d) image after enhancement by FFT; (e) image after adaptive binarization; (f) image done by canny edge detection

Image masking

The mask image for every quadrant is produced misusing the attributes of array and the edge esteem, so the mask image is used like an image filter. Accordingly the mask image ought to be such that the wanted quadrant ought to be white and all other quadrants ought to be black The mask era for the distinctive quadrants are composed manually in excel document, and by creating codes in Matlab, the image array is analyzed by these masks. There are 15 separate masks such that some of quadrants are binary 1s and the rest are binary 0. For the image features to be segmented into the mask quadrants, the feature image must be multiplied with the quadrant that fits to the mask image.
After masking method we achieve a data set which is ready for training by artificial neural networks

Classification of Sign Languages using Neural Network

Neural networks are the result of academic investigation that use mathematical formulations to represent an information processing paradigm that is inspired by the way biological nervous system composed of highly interconnected neurons, work together to solve the problem.

Neural networks or artificial neural networks are defined by the inventor of one of the first neurocomputers, Dr. Robert Hecht-Nielsen who says that neural network is:*"...a computing system made up of a number of simple, highly interconnected processing elements, which process information by their dynamic state response to external inputs."* [20].

These simple processing elements, called neurons, units, cells or nodes process the information by means of signals transmitted by connection links. These links contain the weights, which are multiplied along with incoming signal from any typical neural network to produce output signal, all together representing information being used by the network to solve the problem [21].

Neural networks logic is compared to the human brains because of its approach of solving problems. Neural networks learn by example, which means the examples must be taken carefully, it has to be trained to solve the task, and network finds to solve the task itself, while the computers solve the problem in a way they have to be given small unambiguous instructions about the known problem. It has some other advantages such as an ability to solve problems without finding and describing the method

of such problem solving, without having the knowledge of the nature of solved problems, it only needs to have some examples of similar tasks with good solutions, and it is called adaptive learning or 'programming themselves ' [22].

Neural networks are self-organized in a way they create their own way of representing or organizing information received, as well as its ability to be used as specialized hardware system where major advantage is massive parallel processing [23].

The paper acknowledges Back propagation algorithm presented by Werbos allowing perceptron to be trained in multi-layer configuration, n-1 node neural network to be constructed, and a linearly separable problem to be solved. Back-propagation network works in a way to train data set of some input based on its synaptic weights, then this output is compared to the known output and a mean-square error is calculated. Error value is used to adjust weights from the output layer through hidden layers, until input layer is reached. This process is repeated until the overall error drops below some threshold level, and the network approaches the ideal function asymptotically. The schematic of perceptron is changed by using sigmoidal function, instead of signum function used by earlier versions of perceptrons [24].

The 90 of images are prepared with 3 non-identical images of every letter using Bosnian language alphabet. It has 15 input layers, two hidden layers with different number of nodes that are crucial for obtaining the networks` learning capabilities, and one output layer.

The network is trained using back propagation algorithm, a supervised learning algorithm for multilayer feed forward network. Supervised means that network is not over until it re-adjusts the weights value so the error is minimized or ceased. The performance of the network is evaluated calculating the MSE, correlation coefficient between network outputs, corresponding target outputs and the characteristics of training, validation and testing errors.

Cross validation is used to estimate the quality of the network and to reduce over fitting. The idea of cross validation is to use all the data from the set both as part of training and testing section [25]. We used 3 fold cross validation. After combination of 3 fold is done we receive accuracy for each of them, thus total accuracy is the average of them. To avoid over fitting of the network the initial set is re-partitioned 3 times after 1000 iterations.

As we said, the network is trained in way of changing nodes in hidden layers, as well as functions of an algorithm, to evaluate better network performance.

3. Experimental Results

The system was trained to recognize a set of 30 signs of the Bosnian alphabet. After results are achieved, the accuracy (ACC) has been calculated. Current results are promising, having in mind that few improvements could provide better results. After combination of building network with different nodes and functions, following table give different results.

Table 1. The performance of ANN_1

Number of Neurons in Hidden layer	Transfer Function	Learning Algorithms	ACC
40-1	logsig-purelin	traingdm	0,65
40-1	logsig-purelin	trainlm	0,84
40-1	logsig-purelin	traincgp	0,69

Table 2. The performance of ANN_2

Number of Neurons in Hidden layer	Transfer Function	Learning Algorithms	ACC
30-1	logsig-purelin	traingdm	0,72
30-1	logsig-purelin	trainlm	0,72
30-1	logsig-purelin	traincgp	0,70

Table 3. The performance of ANN_3

Number of Neurons in Hidden layer	Transfer Function	Learning Algorithms	ACC
20-1	logsig-purelin	traingdm	0,83
20-1	logsig-purelin	trainlm	0,79
20-1	logsig-purelin	traincgp	0,71

A significant result is displayed by the overwhelming output of 0.84 (ACC) that had been given out as output by the ANN. The hidden layer is set to possess a total of 40 neurons, „logsigpurelin‟ transfer function and the trainlm algorithm is used.

Our approach could be significantly improved if conditions where images are taken were improved such as the background of the images, brightness and contrast and accurate position. Also if we applied an algorithm to remove breaks and sparks of images produced by the background, system could count the frequencies of images with the help of masks much faster and clearly, thus; make the learning process easier.

4. Conclusion and Future work

Sign language recognition offers improvements of communication for the hearing impaired, guaranteeing less deprivation from society. The fundamental preference of our methodology over alternate endeavors is the need for non-use of extra hardware supplies or special clothes to perceive the signs. As it happens in many applications that utilization computer vision systems to gather information, the fundamental solution of our answer is the image processing part.

The image processing phase of sign recognition procedure obliges a lot of estimations, which presents inertness in the video stream.. The exactness of the framework could be additionally expanded if a more strong edge identification calculation is utilized. By merging hand shapes to design a feature vector that has effectively used to prepare a neural network system, the system classification rate has enhanced promisingly. Fast Fourier transform algorithm models hand textures and canny edge detector demonstrates the edges of hands flawlessly.

By masking Canny edges to give feature vector it gives an expansive size, so we used PCA to effectively reduce the size of the features which have the lowest priority in the training. The classifier is trained for distinctive inputs combinations of 30 signs from Bosnian Sign Language with 3 separate endorsers. Testing is also performed with different combination of gestures and the average recognition rate resulted is around 84,4%.

It was discovered that by utilizing neural systems, the time for preparing will be an element for viable execution of this framework without giving up classification rates. The point of this examination was to make a model framework for the classification of signs, and creating methods which could be later consolidated into a more mind boggling interpretation system.

References:

[1] Rubino, F. H. (1975). Gestuno.*International sign language of the deaf.* Cblisle: British Deaf Association.

[2] J.G.Kyle, B. W. (1985). *SIGN LANGUAGE.* United Kingdom: Cambridge University.

[3] Aran, O. (2008). *Vision based sign language recognition: modeling and recognizing isolated signs with manual and non-manual components'*. Istanbul: Bogazici university.

[4] DimitriosTzovaras, T. P. (2007). Image and Video Processing for disability.*EURASIP journal on Image and Video Processing* , 2.

[5] Anetha K, R. P. (2014). Hand Talk-A Sign Language Recognition Based On Accelerometer and SEMG Data. *International Journal of Innovative Research in Computer and Communication Engineering* , 206.

[6] LUNGOCIU, C. (2011). REAL TIME SIGN LANGUAGE RECOGNITION USING. *StudiaInformatica, ,* 75-85.

[7] Yannick L. Gweth, C. P. (2012). Enhanced Continuous Sign Language Recognition using PCA and Neural.*IEEE* , 55-60.

[8] Md. AtiqurRahman, A.-U.-A. a. (2011). Recognition Static Hand Gestures of Alphabet .*Internatioal Journal of Computer and Information Technology* , 75-78.

[9] Hienz, H. G. (1996). 'Real-Time hand-Arm Motion Analysis using a single video camera'.*International conference on Automatic Face and Gesture recognition* , (pp. 323-327).

[10]Ch.Raghava Prasad, D. M. (2014). Recognizing Gestures of Sign Language Using Feed forward Neural .*INTERNATIONAL JOURNAL OF SYSTEMS AND TECHNOLOGIES* , 13-21.

[11] Weldon, T. P. (1996). Efficient Gabor filter design for texture segmentation.".*Pattern Recognition* , 2005-2015.

[12] Chan, T. F. (2000). Active contours without edges for vector-valued images.*Journal of Visual Communication and Image Representation* , 130-141.

[13]Hasanbegovic, H. (2013). Forms of Hand in Sign language in Bosnia and Herzegovina.*HUMAN, Volume 3, Issue 2* , 25-28.

[14] Adilagic, I. (2013). *Uses of Bosnian sign language.* Sarajevo: University of Phylosophy.

[15] Gilat, A. (2011). *MATLAB: An Introduction with Application.*Wiley.

[16] S., N. K. (2012). *Histogram Equalization and Pca Based Face Recognition on Matlab Platform.*Lambert Academic Publishing.

[17] RuchiKawatra, S. A. (2009). Application of Fast Fourier Transformation on image processing software.*International journal of Artificial*

Intelligence and Knowledge Discovery ,Vol 1, No 1 (2011).

[18] Songcan Chen, D. L. (2004). Image binarization focusing on objects.*BICS* , 2411–2415.

[19] Raman Maini, D. H. (n.d.). Study and Comparison of Various Image Edge Detection Techniques.*International Journal of Image Processing (IJIP) , 3-12.*

[20] Caudill, M. (February, 1989). Neural Network Primer: Part I. *AI Expert* .

[21] Jegede, O. A. (2009). Neural Networks and Its Application in Engineering.*Proceedings of Informing Science & IT Education Conference* , (pp. 83-95).

[22] Anderson, J. A. (1995). *NEURAL NETWORKS.* US: Library of Congress Cataloging in Publication Data.

[23] UdoSeiffert, L. C. (2002). *Self-organizing Neural Networks: Recent Advances and Applications.* Germany: Physica-Verlag, A Springer Verlag Company.

[24] Werbos, P. J. (1991). *The Roots of Backpropagation: From Ordered Derivatives to Neural Networks.* Canada: John Willey and Sons.

[25] Jesse Russell, R. C. (2012). *Cross-Validation.* Book on Demand.

Correspoding author: Sabaheta Dogic
Instiitution: International Burch University, Francuske revolucije bb, Ilidža, Sarajevo, B&H
E-mail: sabaheta_dogic@hotmail.com