Sign Language Recognition

Tülay Karayılan
Department of Computer Engineering Yıldırım
Beyazıt University
Ankara, Turkey
155101122@ybu.edu.tr

Özkan Kılıç
Department of Computer Engineering Yıldırım Beyazıt
University
Ankara, Turkey
ozkankilic@ybu.edu.tr

Abstract— **Millions of people around the world suffer from hearing disability. This large number demonstrates the importance of developing a sign language recognition system converting sign language to text for sign language to become clearer to understand without a translator. In this paper, a sign language recognition system using Backpropagation Neural Network Algorithm is proposed based on American Sign Language. The neural network of this system used extracted image features as input and it was trained using back-propagation algorithm to recognize which letter was the given letter with accuracy of respectively 70% and 85% with two proposed classifiers.**

Keywords—human-computer interaction; hand gesture recognition; artificial neural network; multilayer perceptron; back-propagation; Static Hand Gesture Translator; American Sign Language (ASL); finger spelling

I. Introduction

Over 5% of the world population, which means 360 million people, including 32 million children and 328 million adults, has hearing disability according to World Health Organization (WHO) statistics [1]. Hearing impaired people generally use sign languages for communicating with other people. But mostly hearing people do not know sign language. When considering large number of people who suffer from hearing disability, it is revealed how important providing them opportunity to communicate with hearing people who do not have knowledge of sign language.

A need to develop such a sign language recognition system arises day by day. The important key points of such a sign language system are reducing cost and obtaining more accurate rate efficiently. Developing a sign language system based on machine learning for automatically recognition sign language and converting sign language to text helps hearing people to communicate and understand hearing impaired people.

The proposed system uses the images in the local system or the frame captured from webcam camera as input. Processed input image is given the two classifiers which use Artificial Neural Network, Backpropagation Algorithm. One of them uses raw features and the other one uses histogram features. Finally the predicted result is produced as text.

In this paper, converting sign language to text by an automated sign language recognition system based on machine learning is proposed to satisfy this need. Artificial neural Backpropagation Algorithm is used for the system.

The rest of paper is organized as follows: Following section gives insight on previous studies on sign language recognition systems based on machine learning. In Section III, proposed methodology of the sign language recognition system is explained in detail. Section IV presents the experimental results which are obtained from the proposed methodology. Finally, conclusion of the paper is shown in Section V.

II. LITERATURE SURVEY

Chuan-Kai Yang, Quoc-Viet Tran and Vi N.T. Truong have proposed a system recognizing static hand signs of alphabets in American Sign Language from live videos and translating into text and speech. AdaBoost and Haarlike classifiers have been used for the classification during training process. After the training process, the classifier can recognize different hand postures. Process of testing the system consists of three stages: preprocessing stage, classification stage and text to speech stage. In “Preprocessing Stage”, frames from the video stream are extracted and methods of image processing are used to obtain the features from the image. In “Classification Stage”, the processed images in the preprocessing stage are used as input and classification is done by using Haar Cascade Algorithm. In “Text To Speech Stage”, text recognized by the classifier is converted to speech by using SAPI 5.3. Performance measures of the system are: 98.7% precision, 98.7% recall, 98.7% sensitivity, 99.9% specificity and 98.7% F-score. [2]

Yi Li has proposed Hand Gesture Recognition System using on Microsoft Kinect for Xbox. The system is built on Candescent NUI project and uses Open NI framework for data extraction from the Kinect sensor. There are three main processes in the proposed system: Hand Detection, Finger Identification, and Gesture Recognition. In “Hand Detection” process, firstly hands are separated from background by using depth information. Then two clusters of hand pixels are obtained by using K-means Clustering Algorithm to be able to detect hand by merging two clusters. Afterwards, convex hulls of hands are determined by using Graham Scan Algorithm and detection of hand contours is done by using contour tracing algorithm.

In “Finger Identification” process, firstly common pixels

978-1-5386-0930-9/17/$31.00 ©2017 IEEE

which are on the hand contour and both the convex hull, from which detection of fingertips are done by using the three-point alignment algorithm are calculated for each candidate fingertips collected. Then finger names are determined according to their relative distances, and a direction vector is assigned to each finger. In "Gesture Recognition" process, there are three layers including finger counting classifier, vector matching and finger name collecting. When a single hand gesture is used, system gives accuracy which varies from 84% to 99% . When same gesture is performed with both hands, accuracy varies from 90% to 100%.[3]

Anis Diyana Rosli, Adi Izhar Che Ani, Mohd Hussaini Abbas, Rohaiza Baharudin, and Mohd Firdaus Abdullah have proposed a spelling glove work recognizing the letters of American Sign Language alphabet. The system has been designed targeting deaf-mute people to communicate with normal people. Firstly, the alphabet of the sign language is formed by the designed glove. When the sign language is formed, the bending sensor detects the position of each finger and yields various resistance value. Then Microcontroller that is connected to the spelling glove categorizes the position of bending of finger based on output voltage produced. After Microcontroller finds the combination position of each finger in library, LCD displays correct alphabet. Recognition rate of the system is 70%.[4]

Md. Mohiminul Islam, Sarah Siddiqua and Jawata Afnan have proposed another Hand Gesture Recognition study based on American Sign Language. The system works in four steps for gesture recognition including image acquisition, preprocessing, feature extraction and feature recognition. In "Image acquisition" step, a database of 1850 images of 37 signs is created by collecting image samples of each sign of the sign language from different people. ”Preprocessing” step prepares the image received from camera for feature extraction step by removing noise and cropping image to obtain portion from wrist to fingers of a hand for sign detection. ”Feature extraction” step applies different algorithms for feature extraction of hand gesture recognition system including K convex hull for fingertip detection, eccentricity, elongatedness, pixel segmentation and rotation. Artificial Neural Network, Backpropagation Algorithm is used for training. Gesture recognition rate of the system is 94.32%. [5]

Jun-Wei Hsieh, Teng-Hui Tseng, Wan-Yi Yeh and Chun-Ming Tsai have proposed a sign language recognition system in order to detect English letters and numbers. The color data, skeleton data and depth data which are obtained from the input Kinect are used for detecting palm area of hands. Then Otsu thresholding method is used for extracting palm and morphology closing operation is used for closing the holes in the palms. Then SURF descriptors and features are extracted.

Finally, Brute-force and SVM are used for recognition the letters and numbers in the sign language. The accuracy rate obtained by classification of the numbers and letters with SVM is 100%. The alphabet is also trained by SVM with a recognition rate of 70.59%. [6]

M. Deriche, S. I. Quadri and M. Mohandes have proposed an image based recognition system for Arabic Sign Language. Region growing technique and Gaussian skin model are used respectively for face detection and hand tracking. Hidden Markov Models is used for classification of the signs. Proposed system has accuracy rate of 93%. [7]

Raja S. Kushalnagar, Lalit K. Phadtare and Nathan D. Cahill have proposed a system for synthesis of American Sign Language . The proposed system uses MS Kinect and Open NI library. Skeletal and depth data is read in from the Kinect, then detection of the palm orientation and classification of hand shape is done. A three dimensional extension of the shape context classification algorithm is used for the classification of hand shape. The classification method classifies correctly 10 shapes of 40 hand shapes of the Hamnosys set. [8]

A novel training method for sign language recognition has been proposed by Shuqiong Wu and Hiroshi Nagahashi. The system proposes a new training method for Haar-like features based on AdaBoost classifier, including a hand detector which combines a skin-color model, Haar-like features and frame difference based on AdaBoost classifier for detecting moving right or left hand and a new tracking method which uses the hand patch extracted in the previous frame in order to create a new hand patch in the current frame. The detecting rate of the system is 99.9% and the rate of tracked hands which are extracted in proper size is more than 97.1%. [9]

Chana Chansri and Jakkree Srinonchat have presented a study of recognizing Thai sign language.The proposed system receives the color and depth information from the Kinect sensor for hand detection. Then Histograms of Oriented Gradients technique is used for feature extraction of images. Finally the extracted features are trained bu using Neural Network. The accuracy rates are obtained from different distances from Kinect sensor such as 08.m, 1.0m and 1.2 m are in order of 83.33%, 81.25%, 72.92%.[10]

Majid Zamani and Hamidreza Rashidy Kanan have come up with another methodology based on the saliency of images in order to recognize alphabets and numbers of American Sign Language. In the proposed methodology firstly, input images are processed by the saliency detection and then the obtained output from saliency detection is processed by Linear Discriminant Analysis and Principal Component Analysis for reducing size, maximizing an external class distances and minimizing an internal class distances. Afterwards, collection of vectors for each image obtained from previous stage are trained by using Neural Network. The average recognition rate obtained from the study is 99.88%.[11]

III. METHODOLOGY

The proposed system works as following steps: processing image, classifying the image to decide which letter of sign language the detected hand sign and producing text.

A. Image Processing

The images in the local system or the frame captured from webcam camera are used as input to the system. After processing input image, then classifiers classify the image which class it belongs to.

The system uses two classifiers: one uses raw images features and the other one uses histogram features. Two of the classifiers use Backpropagation Algorithm which is commonly used Artificial Neural Network learning technique to classify images which class they belong to.

Feature extraction from the images is obtained in this way: input images in to the system are converted to a numerical format which means converting each image to a series of RGB pixels. Then images are normalized to be in the same shape. To do that, each observation is resized to 76x66 px. Then resized images are flattened to get 2D array using as features for the classifier. On the other hand histogram features are obtained by getting flattened histogram of input images to the system.

First classifier which is called Raw Features Classifier uses 3072 features. The second classifier which is called Histogram Features Classifier uses 512 features.

Then two classifiers are trained with Backpropagation Algorithm.

Finally the predicted results which are obtained from classifiers are produced as text.

B. *Classification*

1) *Artificial Neural Network*

Artificial Neural Network (ANN) takes its inspiration from human brain which has incredible processing ability because of having webs of interconnected neurons. ANNs are designed by using basic processing unit called perceptron. Perceptron has only one layer and solves linearly separable problems. The problems which are not linearly separable can be solved by Multilayer Perceptron Neural Network (MLP). MLP has multiple layers, including input, hidden and output layers.

The system for recognition sign language is designed as a multilayer perceptron neural network for each classifier. The designed ANN has three layers: an input layer, a hidden layer and an output layer.

- Input Layer was designed to contain 3072 neurons for Raw Features Classifier and 512 neurons for Histogram Features Classifier. The number of neurons was decided to be equal to the number of extracted features of images in the image dataset.
- Hidden Layer was designed to contain 10 neurons for each classifier. This number was used as starting point. The number was increased until it equals to 120 by comparing performance of them and then selecting the best one.
- Output Layer had 3 neurons for each classifier. The designed NN is a multiclass classifier which means returning three class labels (e.g., letters of alphabet: “A”,”B” or “C”). Deciding 3 neurons based on idea that the output layer should have one node per class label in model.

2) Backpropagation Neural Network

Backpropagation Algorithm is the most commonly used ANN learning algorithm. The steps of the algorithm are listed below:

- Initialization of all network weights is done with small random numbers.
- Training data is used as input and Sigmoid Function is used to obtain output for each unit with equation below:

$$o = \sigma(\vec{w} \, . \vec{x}) \qquad \sigma(y) = \frac{1}{1+e^{-y}} \qquad (1)$$

where $\vec{w}$ is vector of unit weight values and $\vec{x}$ is vector of network input values.

- Then error computation step is started. Error signal(δ) is computed for each network output and then propagated to all neurons in the network as input.
- Calculation of error term δ_k for each network output unit *k* is done with equation below:

$$\delta_k \leftarrow o_k(1 - o_k)(t_k - o_k) \qquad (2)$$

where o_k illustrates network output for output unit *k* and t_k indicates desired output for output unit *k.*

- Error term δ_h is computed for each hidden unit h as below:

$$\delta_h \leftarrow o_h(1 - o_h) \sum_{k \in outputs} w_{kh}\delta_k \qquad (3)$$

where w_{kh} illustrates network weight from hidden unit h to output unit *k*.

- Update of each network weight is done where

$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji} \quad \text{where } \Delta w_{ji} = \eta \delta_j x_{ji} \qquad (4)$$

where η is learning rate and x_{ji} illustrates the input from unit *i* into unit *j*. [12]

Backpropagation Algorithm was used for the proposed system as learning algorithm. Marcel Static Hand Posture was used for the system as dataset. The images from the dataset were split into two parts: training and testing. Then 3072 features for Raw Features Classifier and 512 neurons for Histogram Features Classifier were used as input and training was done with Backpropagation algorithm. After training process, the performance of the proposed system was computed by testing the neural network with test data by different metrics as accuracy, precision, recall and F1 score, which is clarified in detail in the following section.

IV. EXPERIMENTAL RESULTS

The system for recognizing sign language using multilayer perceptron neural network was implemented by Python programming language using SciPy libraries.

A. *Data Source*

Marcel Static Hand Posture Database was selected as sign language image dataset. [13] The dataset consists of 6 hand postures of 10 persons (a, b, c, point, five, v). The training set is around 100 Mb and the testing database is of 16 Mb in size.

Dataset was split into two parts: one for training (%75), second one for testing (%25). Only A, B and C hand postures of the dataset were used in the system. So, output of network was designed as having three output type: A, B and C.

B. Performance Evaluation

The performance evaluation of the proposed system was computed by different parameters: accuracy, precision, recall and F1 score.

Accuracy is computed by multiplying 100 by number which is the result of dividing the number of correct predictions by the total number of predictions.

$$Accuracy = \ (TP + TN)/(TP + TN + FP + FN) \tag{5}$$

where *TP* denotes the number of examples that are in the desired class and recognized accurately. *FP* denotes the number of examples that are not in desired class and are recognized wrongly as they belong to desired class. *FN* denotes the number of examples that are in the desired class but the examples are recognized wrongly. *TN* contains a number of examples that are not in the desired class and the examples are recognized properly.

Precision is computed by dividing the number of True Positives to the number of True Positives and False Positives.

$$Precision = TP/(TP + FP) \tag{6}$$

Recall is computed by dividing the number of True Positives to the number of True Positives and the number of False Negatives.

$$Recall = TP/(TP + FN) \tag{7}$$

F1 score is computed with harmonic mean of recall and precision values.

$$F1\ Score = 2 * \left(\frac{Precision*Recall}{Precision+Recall}\right) \tag{8}$$

Pruning technique which means trimming size of network by nodes was used to improve performance of the system. So hidden layer size of the system was set 10 as a starting point and increased to 120.

The obtained results of Raw Features Classifier are shown in Table 1.

TABLE I. CLASSIFICATION PERFORMANCE OF RAW FEATURES CLASSIFIER

Hidden Layer Size	Accuracy	Recall	Precision	F1 Score
10	55.778894%	55.7788945%	31.1128507%	39.9448857%
20	55.7788945%	55.7788945%	31.1128507%	39.9448857%
30	55.7788945%	55.7788945%	31.1128507%	39.9448857%
40	55.7788945%	55.7788945%	31.1128507%	39.9448857%
50	70.5192630%	70.5192630%	77.064912%	71.2457746%
60	55.7788945%	55.7788945%	31.1128507%	39.9448857%
70	55.7788945%	55.7788945%	31.1128507%	39.9448857%
80	59.6314908%	59.6314908%	58.0691298%	53.3261237%
90	55.7788945%	55.7788945%	31.1128507%	39.9448857%
100	54.9413735%	54.9413735%	53.1746452%	40.9340909%
110	26.6331658%	26.6331658%	75.1585171%	35.8120455%
120	37.0184255%	37.0184255%	62.4038543%	46.1899104%

The results of Histogram Features Classifier are shown in Table 2.

TABLE II. CLASSIFICATION PERFORMANCE OF HISTOGRAM FEATURES CLASSIFIER

Hidden Layer Size	Accuracy	Recall	Precision	F1 Score
10	80.4020101%	80.4020101%	86.9626659%	83.3971624%
20	81.5745394%	81.5745394%	87.6189313%	84.3219638%
30	79.5644891%	79.5644891%	86.7168957%	82.8313176%
40	82.9145729%	82.9145729%	89.2372706%	85.9213985%
50	84.4221106%	84.4221106%	89.0334947%	86.5886261%
60	83.9195980%	83.9195980%	88.5801359%	86.1167867%
70	85.2596315%	85.2596315%	88.7302882%	86.8768875%
80	85.5946398%	85.594640%	89.7380100%	87.5460364%
90	84.9246231%	84.9246231%	89.8256443%	87.1726224%
100	85.2596315%	85.2596315%	90.2810828%	87.6357624%
110	85.9296482%	85.9296482%	90.3680945%	87.9918534%
120	85.7621441%	85.7621441%	89.8661294%	87.6772961%

V. CONCLUSION

The system for sign language recognition has been developed using Multilayer Perceptron Neural Network. For the system Marcel Static Hand Posture Database was used. Two classifiers were used for the system. The neural network of Raw Features Classifier accepted 3072 features as input. Histogram Features Classifier received 512 features as input. Then network was trained with Backpropagation Algorithm to recognize which letter the given letter for each classifier. The system gives 70% and 85% accuracy rate from respectively Raw Features Classifier and Histogram Features Classifier.

When considered other studies, the obtained results are average results. The recognition rate can be increased by improving processing image step as a future work.

REFERENCES

[1] “Deafness and hearing loss,” World Health Organization. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs300/en/. [Accessed: 19-Aug-2017].

[2] V. N. T. Truong, C. K. Yang and Q. V. Tran, "A translator for American sign language to text and speech," 2016 IEEE 5th Global Conference on Consumer Electronics, Kyoto, 2016, pp. 1-2. doi: 10.1109/GCCE.2016.7800427

[3] Yi Li, "Hand gesture recognition using Kinect," 2012 IEEE International Conference on Computer Science and Automation Engineering, Beijing, 2012, pp. 196-199.doi: 10.1109/ICSESS.2012.6269439

[4] A. I. C. Ani, A. D. Rosli, R. Baharudin, M. H. Abbas and M. F. Abdullah, "Preliminary study of recognizing alphabet letter via hand gesture," 2014 International Conference on Computational Science and Technology (ICCST), Kota Kinabalu, 2014, pp. 1-5.doi: 10.1109/ICCST.2014.7045002

[5] M. M. Islam, S. Siddiqua and J. Afnan, "Real time Hand Gesture Recognition using different algorithms based on American Sign Language," 2017 IEEE International Conference on Imaging, Vision & Pattern Recognition (icIVPR), Dhaka, 2017, pp. 1-6.doi: 10.1109/ICIVPR.2017.7890854

[6] W. Y. Yeh, T. H. Tseng, J. W. Hsieh and C. M. Tsai, "Sign language recognition system via Kinect: Number and english alphabet," 2016 International Conference on Machine Learning and Cybernetics (ICMLC), Jeju, 2016, pp. 660-665. doi: 10.1109/ICMLC.2016.7872966

[7] M. Mohandes, S. I. Quadri and M. Deriche, "Arabic Sign Language Recognition an Image-Based Approach," Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on, Niagara Falls, Ont., 2007, pp. 272-276.doi: 10.1109/AINAW.2007.98

[8] L. K. Phadtare, R. S. Kushalnagar and N. D. Cahill, "Detecting hand-palm orientation and hand shapes for sign language gesture recognition using 3D images," 2012 Western New York Image Processing Workshop, New York, NY, 2012, pp. 29-32. doi: 10.1109/WNYIPW.2012.6466652

[9] S. Wu and H. Nagahashi, "Real-time 2D hands detection and tracking for sign language recognition," 2013 8th International Conference on System of Systems Engineering, Maui, HI, 2013, pp. 40-45. doi: 10.1109/SYSoSE.2013.6575240R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[10] C. Chansri and J. Srinonchat, "Reliability and accuracy of Thai sign language recognition with Kinect sensor," 2016 13th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), Chiang Mai, 2016, pp. 1-4. doi: 10.1109/ECTICon.2016.7561403

[11] M. Zamani and H. R. Kanan, "Saliency based alphabet and numbers of American sign language recognition using linear feature extraction," 2014 4th International Conference on Computer and Knowledge Engineering (ICCKE), Mashhad, 2014, pp. 398-403. doi: 10.1109/ICCKE.2014.6993442

[12] Tom M. Mitchell,”Artificial Neural Networks”, in “Machine Learning”, McGraw-Hill Science/Engineering/Math, 1997, pp:95-99.

[13] “Image Data Set - Automatic Sign Language Detection,” Google Sites. [Online]. Available: https://sites.google.com/site/autosignlan/source/image-data-set. [Accessed: 19-Aug-2017].