wjert, 2018, Vol. 4, Issue 1, 92 -101. *Original Article* ISSN 2454-695X

World Journal of Engineering Research and Technology
WJERT
www.wjert.org SJIF Impact Factor: 4.326

THE NOVEL METHOD FOR RECOGNITION OF AMERICAN SIGN LANGUAGE WITH RING PROJECTION AND DISCRETE WAVELET TRANSFORM

Murat Taşkıran*, Sibel Çimen and Zehra Gülru Çam Taşkıran

Electronics and Communication Engineering, Yildiz Technical University, 34220, Esenler, İstanbul, Turkey.

Article Received on 03/11/2017 **Article Revised on 24/11/2017** **Article Accepted on 15/12/2017**

***Corresponding Author**
Murat Taşkıran
Electronics and Communication Engineering, Yildiz Technical University, 34220, Esenler, İstanbul, Turkey.

ABSTRACT

Sign Language is a language that allows individuals with hearing or speech impairment to communicate with themselves and their surroundings and has the feature of not being a universal language. This language, which is not universal, is suitable for image processing as it is expressed by hand signals in general terms. In this study, firstly, we performed the skin color detection process to determine the position of the hand gesture, and then extract the edge regions of the hand gestures using the sobel filter and discrete two-dimensional wavelet transform. The obtained hand gesture image was used with the ring projection technique and the two-dimensional image was reduced to one-dimensional. The feature vectors were obtained by applying discrete wavelet transform to obtained the one-dimensional matrise. Classification of hand gestures was made with the Generalized Regression Neural Network (GRNN) using the obtained feature vectors. As a result of this study, 90.44% test accuracy was obtained.

KEYWORDS: American sign language; determination of skin color; ring projection; discrete wavelet transform; generalized regression neural networks.

INTRODUCTION

The sign language, a visual language, is a language that allows individuals with hearing or speech impairments to communicate with each other or with their environment. This

language consists of hand movements, mimics and body language. One of the most important features is that this language is not universal. Sign language varies from country to country. American Sign Language (ASL) is a sign language developed by hearing-impaired children in the western United States. In this study, the processing of feature vectors using image processing techniques of American Sign Language was performed and classified by neural networks. When the studies in the literature were examined, in 2011, Mahmoud Zaki and Samir Shaheen's work[1] on the feature vector and the Hidden-Markov Model (HMM) were used as the classifier in Basic Component Analysis. In Zafrulla's work in 2011 with the team,[2] the feature vector was extracted with the Kinect camera and HMM was used to make the movement meaningful. Zhou Ren and his team[3] extracted feature vectors with depth comparison with Kinect Sensor in 2013 and classified them using conventional shape matching algorithms. In the study of Ching-Hua et al.[4] in 2014, the American Sign Language movement sensor was tried to be recognized and classified with k-NN and SVM. In Upendran and Thamizharasi's work in 2014, Basic Component Analysis was used to extract the feature vectors and k-NN was used as the classifier.[5] In 2015, Cao and his team[6] obtained feature vectors with depth comparison using Microsoft Kinect and implemented the algorithm by using random forest and restricted connection angle algorithm as a classifier. Finally, in 2016, a real-time American Sign Language sensor was proposed by Jayshree Pansare and Maya Ingle[7] in India, using the Haar technique. Funde and Thepade[8] proposed a new feature extraction algorithm consisting of a hybrid of cosine and wavelet transforms. In this study, the skin color detection process was performed to determine the position of the hand gesture, and then the edge regions of the hand gestures were extracted using the sobel filter and discrete two-dimensional wavelet transform. The obtained hand gesture image was subjected to a ring projection technique and the two-dimensional image was reduced to one-dimensional. The feature vectors are obtained by applying the one-dimensional matrice discrete wavelet transform obtained. Classification of hand gestures was made with the Generalized Regression Neural Network (GRNN) using the obtained feature vectors.

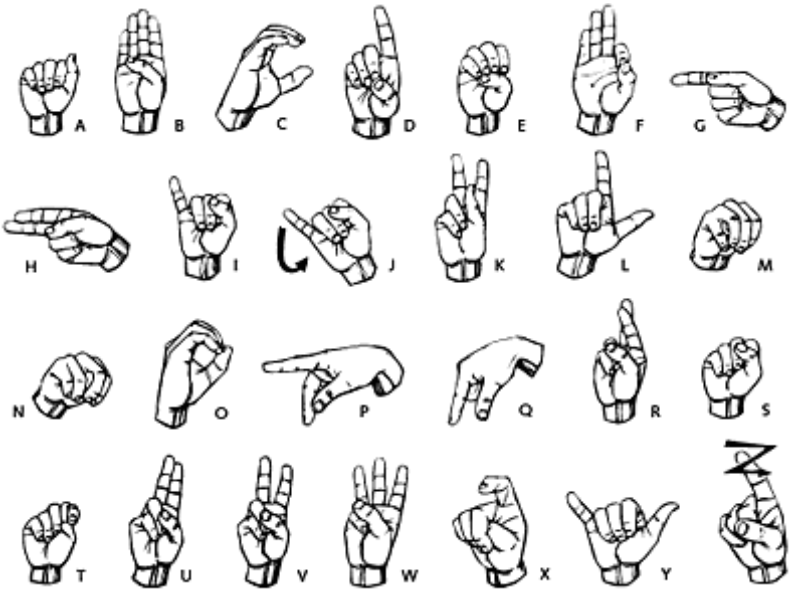

Figure 1: American Sign Language.

MATERIALS AND METHODS

In this study, data set[9] of Massey University, Institute of Information and Mathematical Sciences, collected in 2012, was used. The flow chart of the our proposed method is as shown in Figure 2.

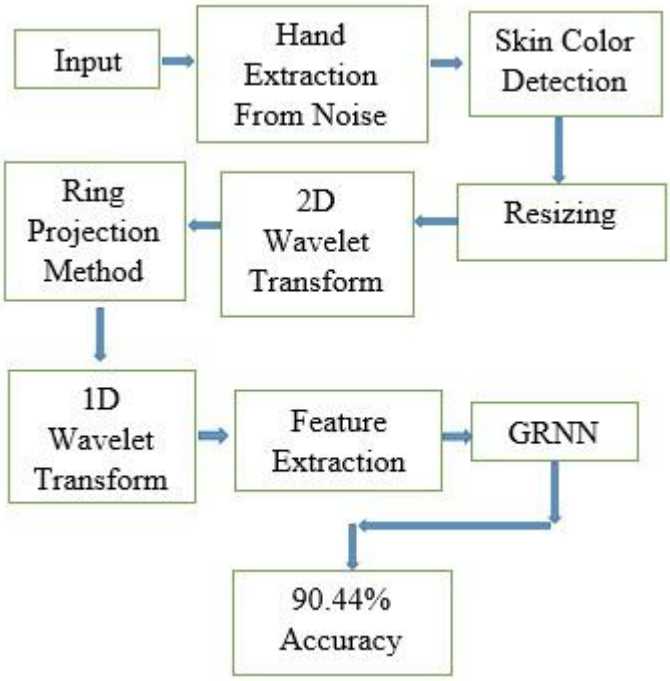

Figure 2: Flow Chart of the Proposed Method.

The data were converted to YCbCr space for skin color determination. YCbCr is a color space storing the brightness signal with Y, and color information (blue and red) with Cb and

Cr. The YCbCr color space emerged during the worldwide effort to create a digital video standard. Y is defined in the range 16-235 of 8 bits. Cb and Cr are defined as 16-240. The YCbCr color model includes detachable shiny features that make it easy to separate vibrancy and brightness. For this reason, it is a suitable color space for image processing. The relation of this space to the RGB color space can be shown as in (1).[10]

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112.000 \\ 112.000 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The threshold values used in the skin color determination studies in the literature have been investigated[11] and consequently the limitations of (2) and (3) have been used in this study in determining the skin color.

$$70 \leq Cb \leq 130 \quad (2)$$

$$134 \leq Cr \leq 173 \quad (3)$$

In order to determine the hand in a clear way, it is necessary to subtract the real part from the areas where noise can be counted. For this, 1/1200 view is taken from the whole image. This can be regarded as a filter used to avoid a problem in understanding the hand gesture determined by shadows or noises that may be due to the condition of the hand gesture. It is aimed to clean the image of 256x256 in large scale by extracting the true value of 1/1200 which is the value found by trial and error method. Skin color was determined and both hand gestures were obtained and background was removed.

Figure 3: Color Skin Detected Hand.

The images were resized to 256x256 and the edge extraction was done with the sobel filter so that the image processing and image dimensions do not differ.

Figure 4: Edge Detection.

The original images that have been resized are subjected to one-stage discrete 2D wavelet transform. 2D wavelet transform is derived from one-dimensional wavelet transform. The concept of discrete one-dimensional wavelet transform can be shown as in Figure 5.

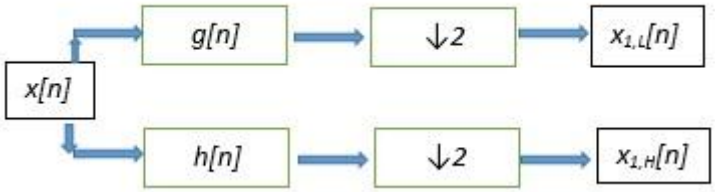

Figure 5: Concept of Discrete Wavelet Transform.[12]

x[n] is the input signal, *h[n]* is high pass filter, *g[n]* is low pass filter, *x1, L[n]* is the output of the low pass filter, *x1,H[n]* is the output of the high pass filter, ↓ 2 is the 2 factor sample reduction.

2D wavelet transform is one dimensional wavelet transform over n and m. This is also shown in Figure 6.

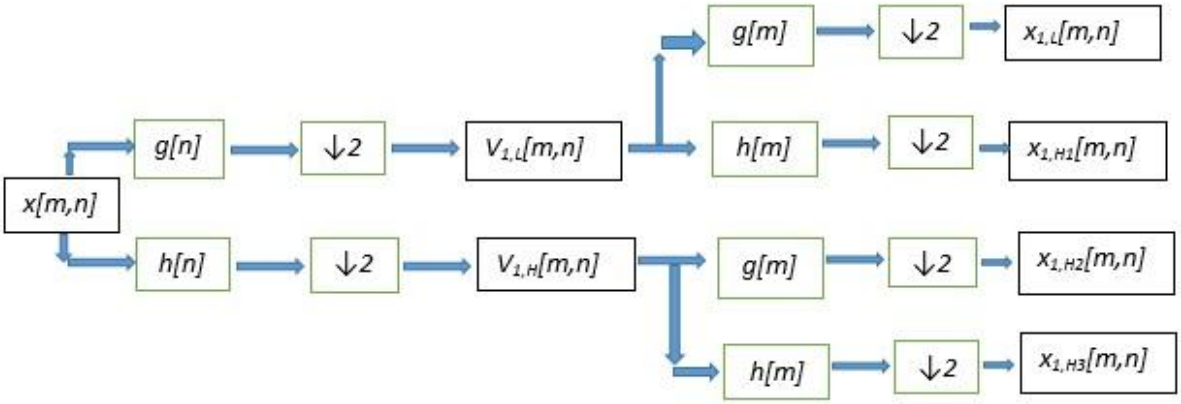

Figure 6: Concept of 2-D Discrete Wavelet Transform.[12]

When the 2D wavelet transform is applied to an image, 4 images are obtained at 1/4 of the original size.[12] The Daubechies-1 filter is used for this wavelet transformation.

Figure 7: 2-D Discrete Wavelet Transform.

The horizontal coefficient vector is used since the edge information can be obtained more clearly from the data obtained from 2D wavelet transform. The edge information obtained as a result of the sobel filter applied to the image of the hand gesture is added to the image obtained as a result of the 2D wavelet transformation. This was done to clearly reveal the positions of the fingers in the hand gesture. The obtained image is shown in Figure 8.

Figure 8: An image, Edge Information Added on 2-D Discrete Wavelet Transform.

The ring projection technique is used to reduce the 2D image, which is the output of the wavelet transform, to a single dimension.

The ring projection technique is feature extraction, independent of the rotation process. During this process, the center point of the view is determined, and rings are drawn inside the radius from this center. The state of the sum of the gray level pixels of the rings passes through the function f gives the attributes. This process is expressed in (4).[13]

$$RP(r) = \int_0^{2\pi} f(r\cos\theta, r\sin\theta)d\theta \tag{4}$$

For each r-value, the feature vector is obtained by summing the gray level brightness values of the pixels falling within the T-angle range using the f-function.

$$RadP(r,\theta) = \int_0^{R}\int_0^{\theta}(r\cos\theta, r\sin\theta)d\theta\, dr \tag{5}$$

The total number of features obtained in the Radial Projection method is always 360.[13] When the images reduced to this single dimension are subjected to one-dimensional discrete 1-dimensional wavelet transform, the detailed coefficient vectors obtained as the feature vector are used. Figure 9 shows the vector of coefficients in detail.

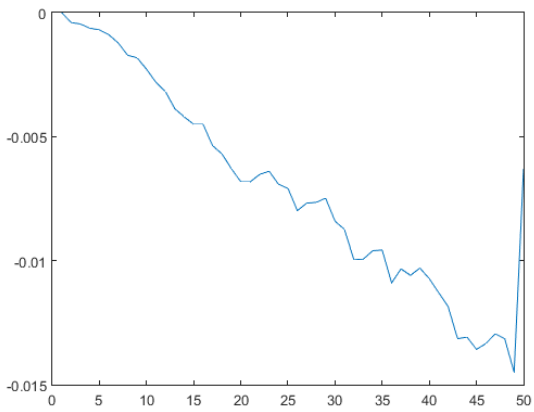

Figure 9: Detailed Coefficient Vector of 1-D Discrete Wavelet Transform.

This feature vector consists of 50 elements. In the used sign language data set, a data set consisting of a total of 900 samples was used, taking 25 samples each for each hand gesture, with 10 digits and 26 letters. 50% of this data set is used for training and 50% is used for test data. The use of the Generalized Regression Neural Networks (GRNN) has been found suitable for classification in consideration of the given multidimensional space distribution.

The GRNN is a radial based network which is a customized structure used for function approximation problems. This network, with its secret neurons, can often provide an

approach to continuous functions by giving a good performance result. It does not need continuous training as it is in Multi-Layer Perceptrons. It may converge to a function depending on the relationship between input and output data and training data. For this reason, the error in approach to function converges to zero as the training data grows. It uses the kernel approach for convergence. According to the kernel approach, the regression of a dependent y variable relative to an independent x variable approximates the value with the most likelihood for y given the x inputs and training set. The approach is determined to approximate the mean square error to the lowest value. The only parameter that can be set in GRNN networks is the smoothing (spread = σ) factor for the kernel function.[14] Figure 10 shows the symbolic architecture of GRNN.

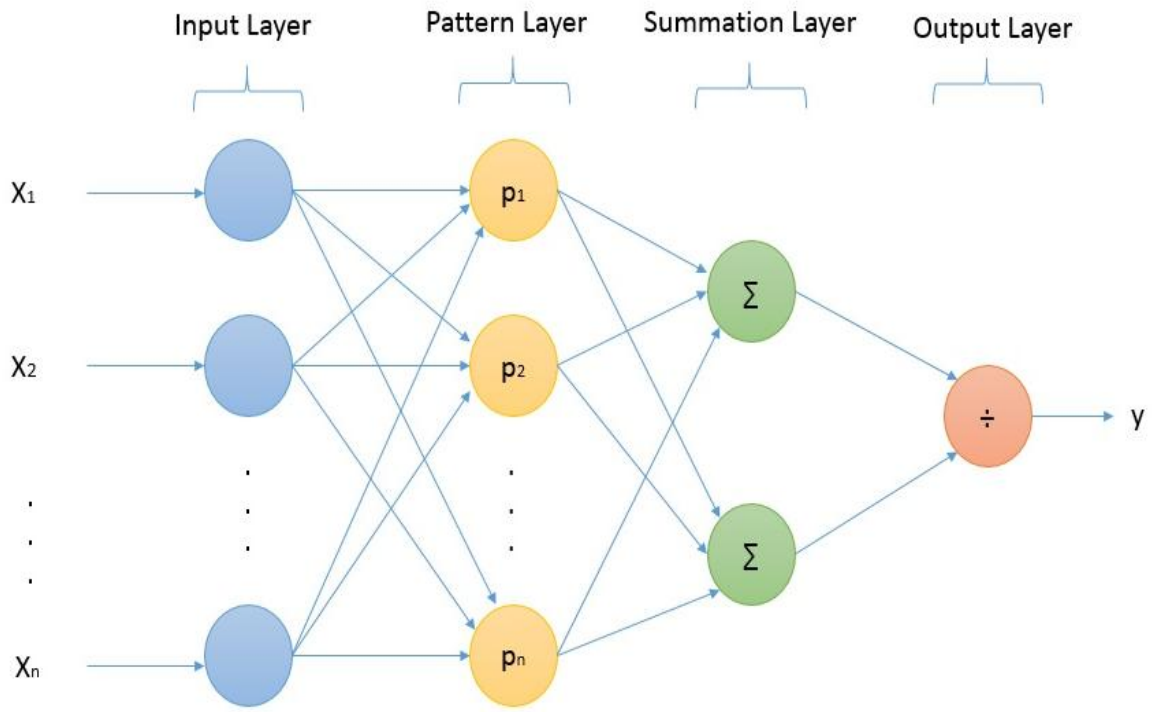

Figure 10: Symbolic GRNN Architecture.

RESULTS AND CONCLUSION

The results of the studies carried out in the literature and the method we propose are given in Table 1.

Table 1: Results.

References	Results		
	Applied Method	**Classifier**	**Accuracy**
[1]	PCA	HMM	89.10 %
[2]	Depth Comparison and Adaptive Distance	HMM	74.81%
[3]	Depth Comparison	Matching Algorithm	93.2%
[4]	Motion Sensor Information	SVM k-NN	72.78% 79.83 %
[5]	PCA	k-NN	77.3%
[6]	Depth Comparison	Random Forest	90%
[7]	Haar Algorithm	Alphabetical American Sign Language Classifier	88.26%
[8]	Hybrid of Cosine Haar and Wavelet Transform	Similarity Measure	~70%
Proposed Method	Ring Projection and Wavelet Transform	GRNN	90.44%

It seems that the proposed method is more successful in understanding hand gestures when compared to literature studies. Although the work done by Ren and his team[3] has a more successful face than the our proposed method in this study, Ren and his team used Kinect camera is a high-cost disadvantage and only 10 of 36 hand gestures have been worked on. However, we have used all characters and also 10 numbers, 36 hand gestures in total.

Future work is aimed at real-time operation of this work and creation of a hardware that can be easily used by the end user.

REFERENCES

1. Zaki M M and Shaheen S I, “Sign language recognition using a combination of new vision based features”. Pattern Recognition Letters, 2011; 32: 572-577.
2. Zafrulla Z, Brashear H, Starner T, Hamilton H, Presti P, “American sign language recognition with the kinect”. Proceedings of the 13th international conference on multimodal interfaces, ACM, New York, NY, USA, 2011; 279–286.
3. Ren Z, Yuan J, Meng J, and Zhang Z, “Robust Part-Based Hand Gesture Recognition Using Kinect Sensor”. IEEE Transaction on Multimedia, 2014; 15(5): 1110-1120.
4. Chuan C H, Regina E, and Guardino C, “American Sign Language Recognition Using Leap Motion Sensor”. 13th International Conference on Machine Learning and Applications, 2014; 541-544.

5. Upendran S and Thamizharasi A, "American Sign Language Interpreter System for Deaf and Dumb Individuals". International Conference on Control, Instrumentation, Communication and Computational Technologies, 2014; 1477-1481.
6. Dong C, Leu M C, and Yin Zhaozheng, "American Sign Language Alphabet Recognition Using Microsoft Kinect". IEEE Conference on Computer Vision and Pattern Recognition Workshps, 2015; 44-52.
7. Pansare J R and Ingle M, "Vision-based approach for American Sign Language Recognition using Edge Orientation Histogram". International Conference on Image, Vision and Computing, 2016; 86-90.
8. Funde A G and Thepade S D, "Recognising Sign using fractional energy of cosine haar hybrid wavelet transform of American Sign images", Conference on Advances in Signal Processing, 2016; 290-294.
9. www.massey.ac.nz/~albarcza/gesture_dataset2012.html.
10. Chen Y K and Chen K J, "Pavement detection in YCbCr color space and its application." IEEE International Conference on Information and Automation, 2015; 663-667.
11. Phung S L, Bouzerdoum A and Chai D, "A novelskin color model in ycbcr color space and its application to humanface detection". In IEEE International Conference on Image Processing (ICIP'2002), 2002; 1: 289–292.
12. Pan G C, "A Tutorial of Wavelet for Pattern Recognition". Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, ROC.
13. Ataş M, Kaya Y, and Uyar M, "An Efficient Rotation Invariant Feature Extraction Method Based on Ring Projection Technique". 21st Signal and Communications Applications Conference, 2013.
14. Sezer E, "Epilepsi Teşhisi için EEG Sinyal Analizi". Selçuk University, Graduare School of Natural and Apllied Sciences, Master's Thesis, 2008.