Journal of Data Science and Modeling, Vol. 1, No. 2, 191-203, June 2023

Research Manuscript

Static Sign Language Recognition Using Depth Data Based on Geometric Features

Mostafa Karbasi[1], Zahra Aghajani*[2], Bahareh Asadi[3],

[1] Assistant professor, Department of electrical and computer engineering, Ghiaseddin Jamshid KashaniUniversity, Abyek, Iran.

[2] MSc Student, Department of electrical and computer engineering, Ghiaseddin Jamshid KashaniUniversity, Abyek, Iran.

[3] Assistant professor, Department of electrical and computer engineering, Ghiaseddin Jamshid KashaniUniversity, Abyek, Iran.

Received: 22/10/2021 Accepted: 06/09/2022

Abstract: Deaf people or people with hearing loss have a major problem in everyday communication. There are many applications available in the market to help blind people to interact with the world. Voice-based email and chatting systems are available to communicate with each other by blinds. This helps to interact with persons by blind people. Also, many attempts have been made with Sign Language (SL) translators to solve of communication gap between normal and deaf people and ease communication for deaf people. In this paper, the geometric feature is used as feature extraction for static sign recognition. Support Vector Machine (SVM) classifier is used for training and testing to develop a system using static signs. So, the accuracy result for static signs using the Geometric feature is 62.92% which needs to be improved by other feature extraction and classifiers.
Keywords: Sign Language SL, Support Vector Machine SVM, Deaf People.

Mathematics Subject Classification (2010): 91B32, 16Y80.

*Corresponding Author: z.aghajani@jku.ac.ir

1. Introduction

A sign language (also signed language) is a language that predominantly utilizes manual communication to exchange meaning. SL can include at the same time joining hand shapes, introduction and movement of the hands, body, or arms, and facial impressions to express a speaker's idea (Wikipedia).

Linguistic research in sign language has shown that signs mainly consist of four basic manual components (Stokoe et al. (1965)).

SL recognition is usually defined as a complicated model with combined positions and placement of the palm (Yao and Fu (2012)). Tremendous efforts are required to translate signs to text and voice with large vocabulary. It is still a challenging problem and an open question of Artificial Intelligence (AI) and computer graphics.

Last two decades, many systems have been developed for understanding the mapping between the hand shape, structure of the joint, and palm movement for SL recognition (Yao and Fu (2012)). In continuous sign language recognition, we have to deal with strong coarticulation effects, i.e., the appearance of a sign depends on preceding and succeeding signs, and large inter- and intra-personal variability. In Ong and Ranganath (2005); Wu and Huang (1999) reviews on research in sign language and gesture recognition are presented.

We use a vision-based approach which does not require special data acquisition devices. We choose the SVM method to train the system with our image database.

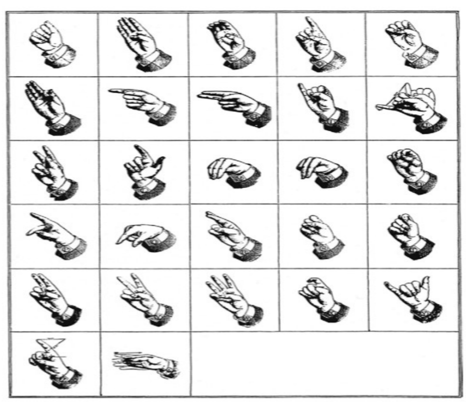

Fig 1 The American Manual Alphabet (Stokoe (2005))

We notice many advantages and disadvantages in all sign language recognition, which are related to database, hand detection, feature extraction, gesture recognition, and especially devices recorded signs. In this paper, we address the limitations and improvement of existing systems.

2. Literature Review

Normally, feature means an underlying arrangement of estimated data and assembles determined qualities. Rearranging data or selecting the proper features from data is called feature extraction. In addition, it could be useful in pre-processing steps for pattern recognition, image processing, and machine learning that require feature extraction. Dimension reduction is directly related to feature extraction, but it has to be processed while the data are too long, and sometimes it seems redundant.

Many features should be calculated to find the model parameters (Kang et al. (2008)). Required hand features have been found from the image, and it is very vital for model parameters and computing. Some normal features are hand silhouettes used by Shimada et al. (1998) and Sonka et al. (2014). Karbasi et al. (2016) briefly explained different hand detection techniques used by different features, and classifiers. They mentioned the advantages and disadvantages of each technique which have been used by different researchers. (Karbasi et al. (2016)) focused on hand and head detection based on the Kinect camera as a new technology in the human-computer interaction domain. Contours features are implemented by Starner and Pentland (1997) while main points are circulated with the hand (Joints, palm, fingertips) executed by Lee and Kunii (1995), Fillbrandt et al. (2003), Starner and Pentland (1997) and Imagawa et al. (1998). Distance-transformed images are shown by Teh and Chin (1988) and Silhouettes have been used as a feature for parameter computation in both three dimensions according to the models (Shimada et al. (1998)).

Karbasi et al. (2016) and Karbasi et al. (2015) worked on hand detection with depth data through a Kinect camera to improve the accuracy for their research. Firstly, they implanted their method based on distance variable to detect the hand front of the user. Secondly, They used a hybrid method to combine depth data and color data to improve the accuracy of the results for hand detection.

Over the years, there have been many patterns for the hand shape category. The hand formation and area, are concerned with its geometric features or contour-based approaches (Belongie et al. (2002)).

The automatic sign language recognition (ASLR) system is based on Bayes' decision rule. The word sequence which best fits the current observation to the trained word model inventory (i.e., the acoustic model in automatic speech recognition (ASR)) and language models (LM) will be the recognition result (Fig 2)(Dreuw et al. (2007)).

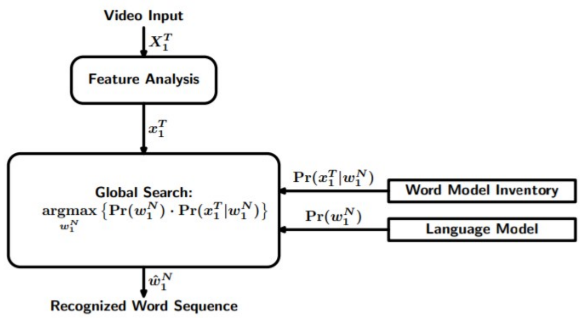

Fig 2 Bayes' decision rule used in ASLR Dreuw et al. (2007)

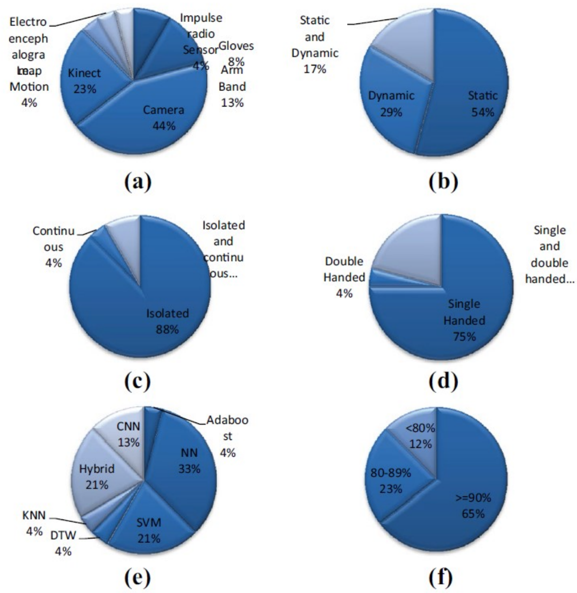

*Fig 3. **a)** Usage of different data acquisition techniques used in American Sign Language (ASL) systems. **b)** Research work carried out on static/dynamic signs in ASL. **c)** Percentage of research work carried out on the basis of signing mode in ASL. **d)** Percentage of research work carried out on the basis of single/double handed signs in ASL. **e** Percentage of research work carried out on technique used for recognition of signs. **f)** Accuracy of research for different ASL systems.(Wadhawan and Kumar (2021))*

All methods mentioned above are very hard to approach and time-consuming in the context analysis. In this way, below, researchers are using the standard deviations of the context (Lin and Chang (2011) and Keskin et al. (2013)). For the above approach, the real hand model is used to show the hand in 21 different sections, and at the end, a random decision forest (RDF) trained on synthetic depth will make a good hand model.

3. Methodology

In Fig 4 Hierarchy of signs is simply shown, the whole steps of how we can analyze the hand gesture in SL.

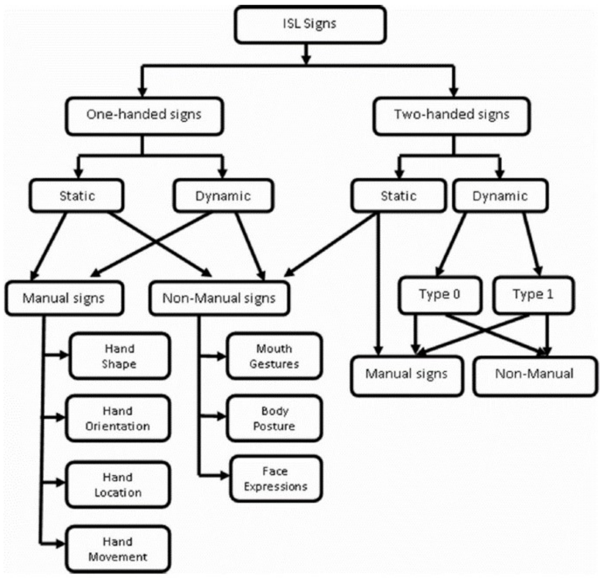

Fig 4 Hierarchy of signs from Wadhawan and Kumar (2021)

In this paper, we just focus on one hand sign in static mode.

Fig 5 a) Single haded static manual sign. b) non- manual sign (Wadhawan and Kumar (2021))

The model-based technique is the same as the non-model one, which includes contours (Lowe (1991)). For different shapes. We need to have distance features and similar features with the same hands. It seems necessary to check the scale features in the image because you will find different sizes for the same hand shape. The similar hand silhouette with dissimilar sizes will provide the matching features values .

There are six features for a geometric feature, which are eccentricity, compactness, orientation angle, semi-major, and semiminor axes, area and roundness. The greyscale image delivered these moment-shaped features. The description of moments of the grey value-function f(x, y) of an object as expressed by (A201) (Kilian (2001)).

$$m_{p,q} = \iint x^p y^q f(x,y)\, dxdy \tag{3.1}$$

Moments are, for the most part, arranged by their order. The order of every moment relies upon the indices p and q of the moment $m_{p,\ q}$ similarly, the sum p + q of the indices expresses the arranging order of the moment mp, q. In view of that, seven shape features have been gotten from the requests existing apart from everything else to portray the face and the two hands blobs.

The joining is ascertained over the area of the object. Normally each other pixel-based feature rather than the grey value could be utilized to compute the snapshots of the object. Utilizing binary picture, the grey value function $f(x,y)$ develops

$$f(x,y) = \begin{cases} 1 & Object \\ 0 & Background \end{cases} \tag{3.2}$$

It can be neglected in the subsequent formulas.

1. Blob area: The zero order moment describes the area A of the object:

$$A = m_{0,0} = \iint f(x,y)\, dxdy \tag{3.3}$$

2. Semi major and semi minor axes denoted by a and b respectively:
   a = height of the ellipse
   b = width of the ellipse

3. Orientation angle in which the blob has its biggest extension θ:

$$\theta = \frac{1}{2}\, arctan\left(\frac{2m_{1,1}}{m_{2,0} - m_{0,2}}\right) \tag{3.4}$$

4. Compactness C, (Swan and Ridgway, 2012):

$$C = \left(\frac{A}{Perimeter^2}\right) \tag{3.5}$$

5. Roundness k:

$$K = \frac{Perimeter^2}{2\pi A} \tag{3.6}$$

where
k = 1 for circle and k > 1 for other objects

6. The eccentricity, ε, can be straight, resulting from the semi-major and semi-minor axes a and b of the object. ε can also be directly calculated from the central moments of second order by:

$$\epsilon = \frac{\sqrt{(m_{2,0} - m_{0,2})^2 - 4m_{1,1}^2}}{m_{2,0} + m_{0,2}} \tag{3.7}$$

The eccentricity, ε, can have values from '0' to '1'. It is '0' for a perfectly round object and '1' for a line-shaped object (Dai et al. (1992)).

4. Result

This is completed by selecting six features (semi-minor, semi-major axes, blob area, eccentricity, compactness, orientation angle, and roundness). These features are moment-based features from the greyscale image.

It is visible that there are variations in the geometric features among the two hand postures of 'A' and 'W' as shown in Fig 6. Table 1 shows the features values extracted from hand postures of 'A' and 'W' to be passed to the recognition stage.

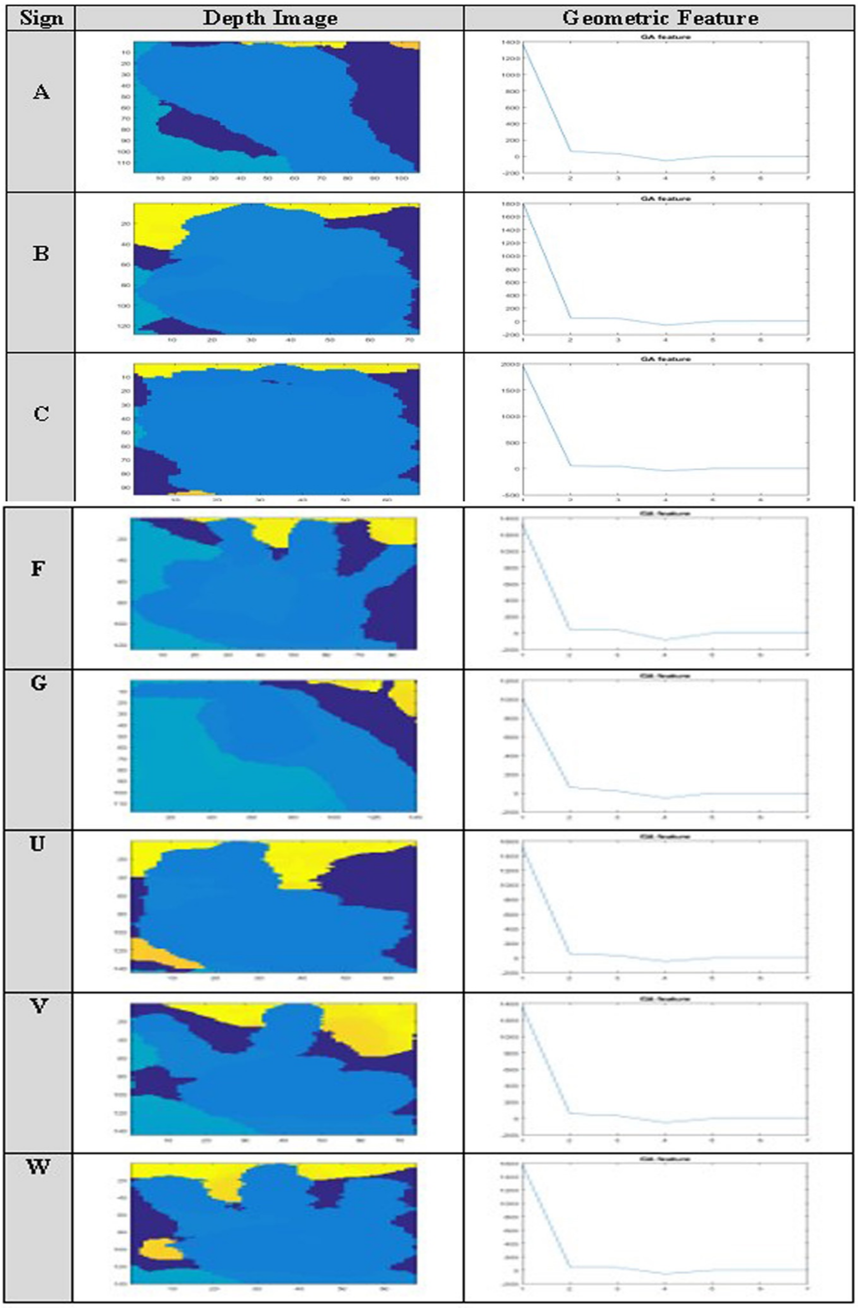

Fig 6. Feature Extraction using Geometric Feature for ‘A’,’B’,’C’,’F’,’G’,’U’, ‘V’,’M’

TABLE I. Extracted Geometric Feature for Letter 'A','B','C','F','G','U', V','M'

Alphabets	Area	Axis		K	E	Compactness	Angle
		Major	Minor				
A	1024	66.75	22.11	-41.86	0.94	0.85	0.59
B	1830	54.89	43.19	52.38	0.61	0.94	0.85
C	962	60.74	22.81	-54.13	0.92	0.87	0.55
F	1330	44.22	41.20	-84.63	0.36	0.85	0.58
G	1008	60.17	22.61	-52.53	0.92	0.89	0.65
U	1517	56.67	36.91	-47.83	0.75	0.89	0.73
V	1358	58.47	36.11	-48.56	0.78	0.79	0.41
W	1582	54.31	42.84	-50.07	0.61	0.81	0.41

The geometric features for the selected characters appear different because of the different sign hand shapes. In terms of area, letters B and 'W' have the largest area due to the wide shape of the signs for these letters. The letter C appears to be large as well in the depth image, but the area is smaller. This is because the shape of the sign is smaller (curved fingers and thumb, with the rest of the palm behind the fingers when viewed from the side), but the color map used could not show this variation as the sensitivity of the color map is not sufficient to demonstrate the size.

The major and minor axis refers to the length of them if we are to approximate an ellipse to cover around the area of interest. The angle of the ellipse indicates the rotation direction of the ellipse, while eccentricity measures the ratio between the major and minor axis lengths. This information indicates the size of the palm and how it is rotated during the sign gesture. This would indicate some features to classify different hand signs. Based on observation from the experiments, the angle approaches 0.85 when the ellipse is aligned with the vertical axis (in the case of the B character), and approaches 0.4 when it is more aligned to the horizontal axis (as in the cases of W and V). Shapes that appear more elliptical have a higher eccentricity, as in the case of characters 'A' and 'G'.

Compactness indicates the homogeneity of the shape based on its perimeter and area. A high compactness value indicates that the shape is uniform, while a low compactness value indicates that the shape is irregular. In the experiments, it was found that signs with protruding fingers scored lower in compactness compared to homogeneous shapes. This can be seen in the cases of 'V' and 'W', where the compactness score is lower than character 'B'.

The value of K represents the roundness of the region of interest. This value is higher in the case of B, where the shape is almost round. This roundness value is lowest in the remaining hand signs with fingers protruding or irregular shape or palm angle, significantly affecting the overall roundness of the area of interest.

Generic features are more generic in nature and were found not to be a strong feature set for the classification of hand gestures using depth information (As shown in Table 4).

TABLE II. Overall Result for Training of 24 Static Signs result using Geometric Features

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	0	10	0	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
C	0	0	13	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	4	4	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
E	0	0	1	0	12	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
F	0	0	0	0	0	9	4	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
G	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
H	0	0	0	0	0	1	0	6	0	0	0	0	0	0	1	0	0	0	2	0	1	0	0	0
I	2	0	0	0	1	0	0	0	8	0	0	1	0	1	0	0	0	0	1	0	1	0	0	0
K	0	0	0	0	0	1	0	0	0	3	0	0	0	0	5	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	1	0	0	0	0	12	1	0	0	0	0	0	0	0	0	0	0	0	1
M	0	0	0	0	0	0	0	0	0	0	1	9	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	9	3	0	0	0	0	0	8	0	1	0	0	0	0	0	0	0	0	0
O	1	1	0	0	1	0	0	0	1	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
P	0	0	0	0	1	0	0	0	0	0	0	0	0	3	9	0	0	0	2	0	0	0	0	0
Q	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	2	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	14	0	0	0	0	0
U	0	0	0	0	0	0	0	0	0	0	4	0	1	0	0	0	0	0	0	10	0	0	0	0
V	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0
W	0	0	0	0	0	0	0	3	0	0	0	0	2	0	0	0	1	0	0	0	0	9	0	0
X	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	8	0
Y	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	13

5. CONCLUSION

In this paper, it is detected that hand shape, signs have variations. Consequently, feature effects on the automatic SL interpreter have been proved. Dissimilar feature vectors were produced for static signs to assess recognition accurateness.

In Section 4, SVM was used as a classifier for testing and training the system using one-hand static signs from the MSL database. Table 2 and Table 3 shows that the accuracy result for the geometric feature is 62.92%. The static sign language recognition results show that, the system accuracy depends mainly on the feature properties.

TABLE III. Overall Result for testing of 24 Static Signs Using Geometric Features

	A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
A	11	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
B	0	10	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
C	0	0	11	0	1	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
D	0	0	0	2	2	1	0	0	1	0	0	0	1	1	2	0	0	0	0	0	2	0	0	0
E	0	0	2	0	7	0	0	3	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0
F	0	0	0	0	0	6	5	0	0	0	2	0	1	0	0	0	0	0	0	1	0	0	0	0
G	0	0	0	0	0	2	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
H	0	0	0	0	1	0	0	4	0	0	0	1	0	0	0	0	0	0	2	0	2	0	0	0
I	1	1	0	0	2	0	0	0	7	0	0	0	0	2	1	0	0	0	0	0	1	0	0	0
K	0	1	0	0	0	0	0	0	1	2	0	0	0	3	3	0	0	0	0	0	0	0	0	0
L	0	0	0	0	0	0	0	0	0	0	11	2	0	0	0	0	1	0	0	0	0	0	0	1
M	0	0	0	0	0	0	0	0	0	0	2	8	0	0	0	0	0	0	0	0	0	0	0	0
N	0	0	0	0	0	3	4	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0
O	1	0	0	0	3	0	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0
P	0	1	1	0	1	0	0	0	0	0	0	0	0	1	8	0	0	0	3	0	0	0	0	0
Q	0	0	1	0	4	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	1	0	0	0
R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	1	0	0
S	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	0	0	0	1	0	0
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0
U	0	0	0	0	0	0	1	3	0	0	5	0	1	0	0	0	0	0	0	5	0	0	0	0
V	0	1	0	0	0	0	0	1	4	0	0	0	0	0	0	0	0	0	1	0	7	0	0	0
W	0	0	0	0	0	0	0	4	0	0	0	0	1	0	0	0	2	0	0	0	0	8	0	0
X	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	9	0
Y	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	13

TABLE IV. The Recognition Accuracy of 8 Characters using GA Features.

Classification / Signs	Samples	Hits	Missed	Accuracy(%)
A	29	23	6	79.3
B	27	15	12	55.5
C	19	11	8	57.8
F	19	6	13	31.5
G	18	13	5	72.2
U	5	5	0	100
V	7	7	0	100
W	8	8	0	100

References

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, **24(4)**, 509-522.

Dai, M, Baylou, P., and Najim, M. (1992). An efficient algorithm for computation of shape moments from run-length codes or chain codes, *Pattern Recognition*, **25(10)**, 1119-1128.

Dreuw, P., Rybach, D., Deselaers, T., Zahedi, M., and Ney, H. (2007) Speech Recognition Techniques for a Sign Language Recognition System, Interspeech in RWTH Aachen University, Germany.

Fillbrandt, H., Akyol, S., and Kraiss, K-F. (2003). Extraction of 3D Hand Shape and Posture from Image Sequences for Sign Language Recognition.*2003 IEEE International SOI Conference. Proceedings (Cat. No.03CH37443)*, 181-186.

Imagawa, K., Lu, S., and Igi, S. (1998). Color-based hands tracking system for sign language recognition.*Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 462-467.

Kang, S.K., Nam, M.Y. and Rhee, P.K. (2008), Color Based Hand and Finger Detection Technology for User Interaction, *2008 International Conference on Convergence and Hybrid Information Technology*, 229-236.

Karbasi, M., Bhatti, Z., Aghababaeyan, R., Bilal, S., Rad, A. E., Shah, A., and Waqas, A. (2016). Real-time hand detection by depth images: a survey. *Jurnal Teknologi*, **78(2)**.

Karbasi, M., Bhatti, Z., Nooralishahi, P., Shah, A., and Mazloomnezhad, S.M.R. (2015). Real-Time Hands Detection in Depth Image by Using Distance with Kinect Camera. *International Journal of Internet of Things*, **4(1A)**, 1-6.

Keskin, C., Kıraç, F., Kara, Y.E., and Akarun, L. (2013). Real Time Hand Pose Estimation Using Depth Sensors. *In: Fossati, A., Gall, J., Grabner, H., Ren, X., Konolige, K. (eds) Consumer Depth Cameras for Computer Vision. Advances in Computer Vision and Pattern Recognition.* Springer, London.

Kilian, L. (2001). Impulse response analysis in vector autoregressions with unknown lag order. *Journal of Forecasting*, **20(3)**, 161–79.

Lee, J., and Kunii, T.L. (1995). Model-based analysis of hand posture. *in IEEE Computer Graphics and Applications* **15(5)**, 77-86.

Lin, C-C., and Chang, C-T. (2011). A Fast Shape Context Matching Using Indexing, *2011 Fifth International Conference on Genetic and Evolutionary Computing*, 17-20.

Lowe, D.G. (1991). Fitting parameterized three-dimensional models to images,*IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13(5)**, 441-450.

Ong, S.C., and Ranganath, S., (2005), Automatic sign language analysis: a survey and the future beyond lexical meaning., *IEEE Trans Pattern Anal Mach Intell.*, **27(6)**, 873–891.

Shimada, N., Shirai, Y. , Kuno, Y., and Miura, J. (1998), Hand gesture estimation and model refinement using monocular camera-ambiguity limitation by inequality constraints, *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 268-273.

Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image processing, analysis, and machine vision*, Springer New York, NY.

Starner, T., and Pentland, A. (1997). Real-Time American Sign Language Recognition from Video Using Hidden Markov Models. *In: Shah, M., Jain, R. (eds) Motion-Based Recognition. Computational Imaging and Vision*, vol 9. Springer, Dordrecht.

Stokoe, W.C. Jr., (2005), Sign Language Structure: An Outline of the Visual Communication Systems of the American Deaf, *The Journal of Deaf Studies and Deaf Education*, **10(1)**, 3–37.

Stokoe, W., Casterline, D., and Croneberg, C. (1965). *A Dictionary of American Sign Language on Linguistic Principles.* Gallaudet College Press, Washington D.C., USA.

Teh, C-H., and Chin, R.T. (1988). On image analysis by the methods of moments. *in IEEE Transactions on Pattern Analysis and Machine Intelligence*, **10(4)**, 496-513.

Wadhawan, A., and Kumar, P. (2021), Sign Language Recognition Systems: A Decade Systematic Literature Review. *Arch Computat Methods Eng*, **28**, 785–813.

Wu, Y., and Huang, T.S. (1999). Vision-Based Gesture Recognition: A Review. In: Braffort, A., Gherbi, R., Gibet, S., Teil, D., Richardson, J. (eds) Gesture-Based Communication in Human-Computer Interaction. GW 1999. Lecture Notes in Computer Science, vol 1739. Springer, Berlin, Heidelberg.

Yao, Y., and Fu, Y. (2012). Real-time hand pose estimation from RGB-D sensor. *2012 IEEE International Conference on Multimedia and Expo*, 705-710.