ML Based Sign Language Recognition System

Amrutha K
Dept. of Computer Science
CHRIST(Deemed to be University)
Bangalore, India
amrutha.k@res.christuniversity.in

Prabu P
Dep. Of Computer Science
CHRIST(Deemed to be University)
Bangalore, India
prabu.p@christuniversity.in

2021 International Conference on Innovative Trends in Information Technology (ICITIIT) | 978-1-6654-0467-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICITIIT51526.2021.9399594

***Abstract*—This paper reviews different steps in an automated sign language recognition (SLR) system. Developing a system that can read and interpret a sign must be trained using a large dataset and the best algorithm. As a basic SLR system, an isolated recognition model is developed. The model is based on vision-based isolated hand gesture detection and recognition. Assessment of ML-based SLR model was conducted with the help of 4 candidates under a controlled environment. The model made use of a convex hull for feature extraction and KNN for classification. The model yielded 65% accuracy.**

Keywords— Automated SLR, preprocessing, feature extraction, segmentation, and classification models.

1. Introduction

Sign language is the standard form of communication among the speech and hearing impaired. The region-wise division of the sign language helps the users to have a facile method to convey information. As the larger population of society does not understand sign language, the speech, and hearing impaired usually rely on the human translator. The availability and affordability of using a human interpreter might not be possible all the time. The best substitute would be an automated translator system that can read and interpret sign language and convert it into an understandable form. This translator would reduce the communication gap that exists among people in society.

The SLR should be trained with many sign language data and its grammar for a smooth and uninterrupted sign language conversion. Every gesture created so far has a specific meaning and an application. Every sign language used all over the world is rich in grammar and vocabulary. SLR can be considered as a modified HCI model, where the system can read and process the hands' movement [1]. Such models would pave a path for barrier-free communication [1].

Sign languages of different regions have different semantics, and the hand gesturing also varies. As a result, the complexity of the model also varies. This diversity among the sign language acts as an obstacle in developing a universal SLR model. Once a model is created, even though region-wise, the speech and hearing impaired will get more freedom in communication in any field

This paper explores different techniques used in each module of an ML-based SLR system.

II. Automated Sign Language Recognition System

An automated sign language recognition system is a model that can recognize sign gestures. The gestures are usually given with hand movement and supporting factors of facial expression and body postures[2]. The hand gestures' capturing happens when the signer stands in front of the camera, facing it. The camera captures the sign and feeds the input into the interpreter model.

The model must be pre-trained with a large amount of information to increase the prediction accuracy. The image size fed into the model also varies for a different concept. Figure 1 illustrates the rudimentary structure of an automated SLR system.

Once given as input, the image undergoes a series of processing steps until the model recognizes it. Even though it is a stand-alone model, the SLR model faces many difficulties while predicting the sign. The signer's sudden movements of the hand, occlusion of the hands, and the signer's speed are a few of the factors that affect the prediction apart from the illumination condition [3].

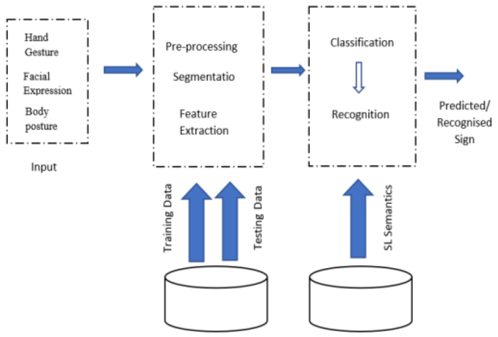

Fig. 1. ML-based SLR model

A. *Preprocessing*

The first step of an SLR model after capturing the image is to preprocess it. The purpose of preprocessing is to enhance the raw image captured with the help of a camera. When the image is captured, there are chances of blurring or other distortions like uneven lighting conditions. There are many methods through which these unwanted features can be removed from the image. Brightness corrections, grayscale transformations are a few of the most commonly used preprocessing methods. The preprocessing intensify the details such as edges and regions for recognition. J Sinha and K Das, in their paper [4], have made use of three types of preprocessing techniques, namely skin segmentation and histogram matching, eigenvalues, and eigenvector during the feature extraction stage and Eigenvalue with Euclidean Distance matrics for classification. Most of the SLR system suffers from the illumination problem. As a solution, authors R Gross and V Brajovic proposed a novel image preprocessing algorithm [5] that reduces illumination variation in the image. With the help of a single brightness image, the algorithm calculates the brightness field and then applies it to all the frames with varying illumination.

B. *Segmentation*

Segmentation is the process of dividing the images into smaller portions from which information can be retrieved [6]. The partition is based on similar properties or features they exhibit. The similarity can be found in colour, texture, or even colour-based histogram. The labeled segments collectively form the entire picture[7]. This is an unavoidable step as features can be extracted from each of these regions. The given figure explains the broad classification of segmentation. As a large amount of information is fed from the hand region, it must be separated from each video frame. The hand segmentation can be based on the colour, threshold, or edge, depending on the image requirement and type. Both 2-dimensional grayscale image and 3- dimensional RGB image can be segmented into different regions. Clear boundaries separate these regions. A few of the most commonly used segmentation methods used for SLR models are discussed in this paper.

Colour Based segmentation

As a large amount of information is fed from the hand region, it must be separated from each video frame. The hand segmentation can be based on colour, threshold, or edge. There are many existing methods for hand segmentation. In their paper, M P Paulraj et al. [8] proposed separating hands without gloves. The singers were given proper instructions on their outfits while experimenting [9]. In their paper, Tofighi, G et al. introduced An adaptive histogram template of skin [10]. This method helps to extract the histogram of the signer's hand using colour and texture. The image should have just the hand during this segmentation as it cannot segment from a complex background [11][12]. In their paper, J R Pansare and M Ingle [11] proposed an American SLR system where the extracted frame undergoes a preprocessing phase of removing noise and morphological operation. A skin detector model has then applied along with the grey thresholding method to segment a perfect hand from the image.

In their paper, J Han et al. proposed a skin segmentation and tracking model for the SLR system [13]. The skin colour model is build using SVM active learning along with region segmentation. The authors have also used Kalman Filter for tracking any skin types. In their paper, J Zieren K-F Kraiss proposed a person independent SLR system [14]. The authors have used statistical colour models for skin detection[14][15] to capture the face and hand skin quickly, but the system fails during uncontrolled or cluttered backgrounds. A probability threshold is kept for understanding different skin colours.

Edge-based segmentation

While performing threshold-based segmentation under the perfect condition of light, there is a chance of the segmentation influencing the digital image. The objects detected might be too small or too big that the image's edge might be vague. For avoiding this bias, edge-based segmentation can be used to extract the features [7]. The major limitation of the edge-based segmentation is that if the model is not a noise-tolerant type, then the algorithm might create a false edge or even miss out on an existing edge [15]. Many authors have proposed accurate SLR models based on the edge detection method.

E E Hemayed proposed an edge-based SLR for Arabic sign language recognition using a Prewitt edge detector in their paper. The detector forms a matrix with the highest gradient, and the vectorized form is taken for differentiating the gestures [16]. MVD Prasad et al. proposed a fusion of two edge-based models: morphological subtraction model and canny edge detection model. The Erosion and Dilation method segments the hand and the face area from the frame, and then the Canny edge detector is used for more precision in shapes [17][18].

C. Feature Extraction

The preprocessed image will be of high dimensionality and will take a lot of computational costs if taken directly for classification. This matter can be resolved with the help of feature extraction. Reduction in the dimension forms compact information that can be fed for classification and recognition. Fewer dimensions with more details will result in high accuracy in prediction. Feature selection is also a method through which the information can be retrieved. Still, while performing feature selection, there is a chance for information loss as sometimes, not all information can be taken into consideration. In feature extraction, the dimensions reduce, but it does not face any information loss. The features present in an image can be relevant, irrelevant, or redundant[19].

PCA

Principal Component Analysis is the most commonly used unsupervised feature extraction technique. PCA is usually used when the information is in the form of redundant data. The principal components that do not contribute to the main element are removed as a part of dimension reduction.

Convex Hull

The convex hull is an image processing method where robustness and flexibility are required. The relevant vertices are connected to form a closed polygon structure called the convex region [20]. The region thus formed is taken for further processing.

In the SLR model convex hull is used to detect and capture the finger's movement and palm. Usually, sign language is communicated with the help of hand gestures [21]. The boundary of the region represents the vertices as well as the edges. Features from each hand and fingers can be considered individually, and the prediction can be performed.

D. Classification

Classification is the process of collecting images and label them for recognition. Classifiers also predict the class of the given dataset. The categorization of classifiers can be done based on their learning capacity.

Lazy learner classifier

These classifiers do not learn when the data is fed into them; instead, they store them. The model classifies when the test data is provided to it. The classification is based on the most related data in the storage. Example of lazy learner classifiers include KNN and case-based reasoning

K- Nearest Neighbour

KNN one of the best classification algorithm as it performs well for regression problems. They classify the image s based on the feature vectors extracted from the training images[22]. If the input is a multimodal system, then KNN is the best option for classification. The computational complexity of the KNN can be reduced by reducing the size of the design set [23] or by using fast-KNN[24][25]. The K-neighbours are formed based on some distance metrics. A voting process takes place among the elements, and then a group is formed,

KNN algorithm is the best choice when the image to be classified represented using local features [26]. The performance of the KNN algorithm depends on the type of distance metrics used by the algorithm. The most widely used distance metric is the Euclidean Distance [26]. Other distance metrics include Cosine, Minkowsky, and Chi-Square.

The most popular Euclidean distance method can be calculated from the feature vector X= $(x_{1,} x_2 \ldots x_n$) and Y=$(y_1, y_2 \ldots y_n)$, which shows that the vectors are oof n-dimension.

Then the Euclidean Distance

$$Dist(E)\ (X, Y) = \sqrt{\frac{\sum_{i=1}^{n} x_i - y_i}{n}} \tag{1}$$

The Cosine distance measures the homogeneity among the elements and can be calculated using

$$Dist(c)\ (X, Y) = \frac{\vec{x}\cdot\vec{y}}{|\vec{x_1}||\vec{y}|} \tag{2}$$

The third popular distance metric, Minkowsky, can be calculated using the given formula

$$\mathrm{Dis}_{(m)}(X, Y) = \left(\sum_{i=1}^{n} |x_i - y_i|^m\right)^{\frac{1}{m}} \quad (3)$$

If m=1 indicates Manhatten Distance and m=2 indicates Euclidean Distance. Even though the variable - 'm' can assume any value, it is always set as either 1 or 2.

Eager Learning Classifiers

These classifiers form a model before the training data is fed into it. As a result, the model requires a large amount of training time and less time to predict. The most common examples of eager learning algorithms include SVM, Decision Tree, and Naïve Bayes, and ANN [27].

Support Vector Machines

A support Vector Machine is a complex network that can accommodate large classes of neurons and polynomial classifiers[28]. The hyperplane classifies the elements into classes[29]. According to [29][30], this hyperplane also helps to reduce the misclassification problem.

An object from a cluttered environment can be easily spotted with the help of a support vector machine. The SVM must be initially trained with a large amount of positive and negative images. Objects like face, hands, and other regions of interest can be easily identified in an uncontrolled environment [31].

SVM can classify colour images that can be represented using methods like RGB or CYMK. The colour distribution is represented with the help of colour histogram[32].

Decision Tree

This classifier model is in the form of a tree where the branches are formed using the divide and conquer rule. The height of the tree extends until the termination point is met. The top attribute of the tree is selected using the information gain process. The models give poor performance if new and unknown data is fed into it and most of the time suffers from overfitting. These limitations can be avoided with the help of pre and post pruning method.

Naïve Bayes

This model assumes that the attributes are conditionally independent. Using this model, a large dataset can be trained and tested with significantly less computational time. One of this model's flaws is that it suffers from the 'zero probability problem': that is, when the conditional probability of a particular attribute is zero, then the model fails to give a valid prediction.

ANN

This model is a three-layered structure with a hidden layer in the middle of the input and output layers. The learning process is by adjusting the parameters of weight and bias. As the complexity of the problem increases, the number of hidden layers also increases. The model maintains a high ground by solving real-time problems. However, as the number of hidden layers increases, the training time and the time to adjust the weight increases.

Table 1 gives a glance at the popular works done for developing SLR models for different sign languages. Their performance is also measured and displayed

TABLE 1: MOST COMMONLY USED ML CLASSIFIERS

Authors	Algorithm used	Algorithm Performance (%)	Data set
Shahabableh [33]	KNN	97	ArSL
	Bayesian	100	
Nandy et al.[33]	Euclidean Distance+KNN	89.8	ISL
Rekha et al. [33].	SVM	89.1	ASL
	KNN	91.7	
Lilha and Sivamurthi [34]	SVM	98.1	ISL
Mohandes [33]	PCA	99.6	ArSL
	SCV		
Issah and Suciat[33]	KNN	86	ASL
Agrawal et al.[35]	SVM	95.3	ISL
	Bayesian KNN	89.9	
Sahoo[33]	KNN	95	ISL
	NN	96	

III. EXPERIMENTAL ANALYSIS

Experimental setup

As an experimental evaluation, the performance of selected models was analyzed with the help of four candidates. Each candidate was made to sign in front of the camera in a controlled environment. The background was stable and plain with constant illumination. With the help of the same candidates, the training dataset was created for each gesture. Each candidate gave five attempts for each gesture. The model was tested with the dataset that included single-handed hand gestures of numbers from 1-5.

The model started with preprocessing, where the background was removed using the threshold method, contour-based segmentation helped in obtaining the contours of the fingers. The convex hull method was used for feature extraction and finally KNN with Euclidean Distance for classification. Figure 2 shows the model recognition of the gestures.

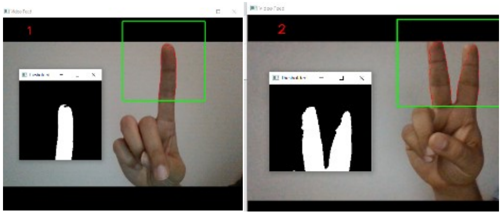

Fig. 2. Recognition of the model

The experiments showed that numbers 1 and 2 were recognized by the model more than other numbers, and the model gave an overall average accuracy of 65%. The formula calculated the accuracy

$$\text{ACC} = \frac{\#\ of\ successful\ attempts}{\#\ of\ gestures} X100 \qquad (4)$$

Table II gives the successful attempts of each candidate for the given dataset.

TABLE II. SUCCESSFUL ATTEMPTS OF EACH CANDIDATE

Candidate	**Numbers**					
	0	***1***	***2***	***3***	***4***	***5***
A	3	4	4	3	3	2
B	4	4	4	2	2	3
C	3	5	3	4	3	3
D	2	4	4	4	3	2

The model showed less performance when the distance between the camera and the object is not considerable. The detection and recognition were less when the hand was moved at a fast pace.

CONCLUSION AND FUTURE WORK

The automated SLR system must recognize the signs gestured by the signer in a real-time condition. The article gives an overview of a machine learning-based SLR model. Each step of the recognizer uses different algorithms to extract maximum information with minimum cost. The experiment was conducted using the convex hull method. The tested model showed an accuracy of 65%. The accuracy can be increased further by using a huge dataset and a different classifier.

The given experiment was used to detect and recognize individual and isolated sign. This can be further modified to recognize continuous sign language for more practical use.

REFERENCE

[1] B. L. Loeding, S. Sarkar, A. Parashar, and A. I. Karshmer, "Progress in automated computer recognition of sign language," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 3118, 2004, doi: 10.1007/978-3-540-27817-7_159.

[2] A. Er-Rady, R. Faizi, R. O. H. Thami, and H. Housni, "Automatic sign language recognition: A survey," presented at the Proc. - 3rd Int. Conf. Adv. Technol. Signal Image Process. ATSIP 2017, vol. 9, pp. 1–7, 2017, doi: 10.1109/ATSIP.2017.8075561.

[3] B. Bauer and K. Kraiss, "Towards an Automatic Sign Language Recognition System Using Subunits," presented at the International Gesture Workshop (pp. 64-75). Springer, Berlin, Heidelberg, 2002.

[4] J. Singha and A. Karen Das, "Recognition of Indian Sign Language in Live Video," 2013, arXiv preprint arXiv:1306.1301.

[5] R. Gross and V. Brajovic, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition," presented at the International Conference on Audio-and Video-Based Biometric Person Authentication (pp. 10-18). Springer, Berlin, Heidelberg..

[6] D. Kaur and Y. Kaur, "International Journal of Computer Science and Mobile Computing Various Image Segmentation Techniques: A Review," International Journal of Computer Science and Mobile Computing, 3(5), pp.809-814..

[7] "A Study Of Image Segmentation Algorithms For Different Types Of Images," Int. J. Comput. Sci. Issues, vol. 7, no. 5, 2010.

[8] M. P. Paulraj, S. Yaacob, H. Desa, C. R. Hema, and W. M. R. Wan Ab Majid, "Extraction of head and hand gesture features for recognition of sign language," 2008 Int. Conf. Electron. Des. ICED 2008, no. January, 2008, doi: 10.1109/ICED.2008.4786633.

[9] S. Joudaki , D. B. Mohamad,T. Saba,., A. Rehman., M. Al-Rodhaan, and A. Al-Dhelaan, "Vision-Based Sign Language Classification : A Directional Review Vision-Based Sign Language Classification : A Directional Review," IETE Tech. Rev., vol. 31, no. 5, pp. 383–391, 2014, doi: 10.1080/02564602.2014.961576.

[10] G. Tofighi, S A. Monadjemi, and N. Ghasem-Aghaee, "Rapid Hand Posture Recognition Using Adaptive Histogram Template of Skin and Hand Edge Contour."2010 6th Iranian Conference on Machine Vision and Image Processing (pp. 1-5). IEEE.

[11] J. Ayshree and R. Pansare, "Vision-Based Approach for American Sign Language Recognition Using Edge Orientation Histogram."2016 International Conference on Image, Vision and Computing (ICIVC) (pp. 86-90). IEEE.

[12] Y. Fang, K. Wang, J. Cheng, and H. Lu, "A real-time hand gesture recognition method.", presented at 2007 IEEE International Conference on Multimedia and Expo (pp. 995-998). IEEE.

[13] J. Han, G. Awad, and A. Sutherland," Automatic skin segmentation and tracking in sign language recognition " Published in IET Computer Vision," doi: 10.1049/iet-cvi:20080006, 3(1), pp.24-35..

[14] J. Zieren and K.-F. Kraiss, "Robust Person-Independent Visual Sign Language Recognition." presented at Iberian Conference on Pattern Recognition and Image Analysis (pp. 520-528). Springer, Berlin, Heidelberg,2005.

[15] W.-X. Kang, Q.-Q. Yang, and R.-P. Liang, "The Comparative Research on Image Segmentation Algorithms," presented at First International Workshop on Education Technology and Computer Science (Vol. 2, pp. 703-707). IEEE. 2009, doi: 10.1109/ETCS.2009.417.

[16] E. E. Hemayed and A. S. Hassanien, "Edge-based recognizer for Arabic sign language alphabet (ArS2V-Arabic sign to voice)," in ICENCO'2010 - 2010 International Computer Engineering Conference: Expanding Information Society Frontiers, 2010, doi: 10.1109/ICENCO.2010.5720438.

[17] M. V.D. Prasad, P. V. V. Kishore, E. Kiran Kumar, and D. Anil Kumar, "Indian sign language recognition system using new fusion based edge operator," J. Theor. Appl. Inf. Technol., vol. 88, no. 3, 2016.

[18] J. Serra, "Image analysis and mathematical morphology," by Academic Press, London, 1982, Cytometry, vol. 4, no. 2, 1983, doi: 10.1002/cyto.990040213.

[19] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction : a survey," Artificial intelligence review pp. 1–54, 2015, doi: 10.1007/s10462-012-9356-9.

[20] M. A. Jayaram and H. Fleyeh, "Convex Hulls in Image Processing: A Scoping Review," American Journal of Intelligent Systems, vol. 6, no. 2, pp. 48–58, 2016, doi: 10.5923/j.ajis.20160602.03.

[21] Y. Wen, C. Hu, G. Yu, and C. Wang, "A robust method of detecting hand gestures using depth sensors," in Proceedings - 2012 IEEE Symposium on Haptic Audio-Visual Environments and Games, HAVE 2012, 2012, pp. 72–77, doi: 10.1109/HAVE.2012.6374441.

[22] J. Kim1, B.-S. Kim2, and S. Savarese3, "Comparing Image Classification Methods: K-Nearest-Neighbor and Support-Vector-Machines,"proceedings of the 6th WSEAS international conference on Computer Engineering and Applications, and Proceedings of the 2012 American conference on Applied Mathematics, vol. 1001, pp. 48109-2122. 2012. .

[23] Y. L. Qiao, J. S. Pan, and S. H. Sun, "Improved k nearest neighbors classification algorithm," Dianzi Xuebao(Acta Electronica Sinica), 33(6), pp.1146-1149.., vol. 33, no. 6, pp. 1146–1149, 2005, doi: 10.1109/apccas.2004.1413076.

[24] W. J. Hwang and K. W. Wen, "Fast kNN classification algorithm based on partial distance search," Electronics letters., vol. 34, no. 21, 1998, doi: 10.1049/el:19981427.

[25] K. Fukunaga and P. M. Narendra, "A Branch and Bound Algorithm for Computing k-Nearest Neighbors," IEEE Trans. Comput., vol. C–24, no. 7, 1975, doi: 10.1109/T-C.1975.224297.

[26] L. Y. Hu, M. W. Huang, S. W. Ke, and C. F. Tsai, "The distance function effect on k-nearest neighbor classification for medical datasets," Springerplus, vol. 5, no. 1, Dec. 2016, doi: 10.1186/s40064-016-2941-7.

[27] S.Rafatirad and M Heidari,. "An exhaustive analysis of lazy vs. Eager learning methods for real-estate property investment."2018.

[28] S. Bernhard, "SVMs-a practical consequence of learning theory.",Corpus ID: 2526394, 1998.

[29] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support Vector Machines for Histogram-Based Image Classification," IEEE transactions on Neural Networks, 10(5), pp.1055-1064. 1999.

[30] V. Vapnik, "The Nature of Statistical Learning Theory.", Springer science & business media,2013

[31] C. Papageorgiou and T. Poggio, "Trainable system for object detection," Int. J. Comput. Vis., vol. 38, no. 1, pp. 15–33, 2000, doi: 10.1023/A:1008162616689.

[32] S. Agrawal, N. K. Verma, P. Tamrakar, and P. Sircar, "Content Based Color Image Classification using SVM," presented at Eighth International Conference on Information Technology: New Generations (pp. 1090-1094). IEEE. 2011, doi: 10.1109/ITNG.2011.202.

[33] A. Kumar Sahoo, M. Sharma, and R. Pal, "Indian sign language recognition using neural networks and KNN classifiers Image processing" View project IoT and 5G View project vol. 9, no. 8, 2014.

[34] H. Lilha and D. Shivmurthy, "Analysis of pixel level features in recognition of real life dual-handed sign language data set," in 2011 International Conference on Recent Trends in Information Systems, ReTIS 2011 - Proceedings, 2011, pp. 246–251, doi: 10.1109/ReTIS.2011.6146876.

[35] S. C. Agrawal, A. S. Jalal, and C. Bhatnagar, "Redundancy removal for isolated gesture in Indian sign language and recognition using multi-class support vector machine," Int. J. Comput. Vis. Robot., vol. 4, no. 1–2, pp. 23–38, 2014, doi: 10.1504/IJCVR.2014.059361.