ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2022.0357

ANALYSIS ON DIFFERENT METHODS FOR SIGN LANGUAGE IDENTIFICATION

M. Prema and P.M. Gomathi
Department of Computer Science, P.K.R. Arts College for Women, India

Abstract

Sign language (SL) is the most organized and structured type of hand/arm gesture in the communicative hand/arm gesture taxonomies. The ability of machines to comprehend human actions and meanings has numerous uses. SL identification is one area of focus. SL is employed by the deaf and hard-of-hearing communities to communicate. Hearing-impaired persons communicate via visual indicators instead of vocal communication and sound patterns. SL also uses facial expressions and body postures as a medium of communication. Pattern matching, computer vision, natural language processing are the key factors in SL identification. This study reviews state-of-the-art methodologies employed in current SL identification research, comparing the various algorithms at each stage. Discuss the challenges and limitations of gesture identification research in general, as well as SL identification in particular. Overall, this paper gives a thorough introduction to the topic of automated SL identification, paving the way for future research.

Keywords:

SL, Gesture Taxonomies, Deaf Community, Computer Vision, Pattern Matching, Natural Language Processing

1. INTRODUCTION

People utilise hand signals to express their thoughts and feelings, and it serves as a reminder of what they're saying. As a means of communication, SL makes use of a system of hand gestures, visual motions, and signs. SL is a useful tool for the deaf and speech-impaired community to utilise in everyday communication. In order to communicate, signers employ a variety of body parts, including their fingers, hands, arms, heads, and facial expressions. The hearing community, on the other hand, does not use SL, and only a small number of people can understand it [1] [2].

Until now, the deaf community and the rest of the community have not been able to communicate effectively because of this barrier. The majority of SL is based on the upper body, from the waist down. In addition, the shape of the same symbol might alter dramatically depending on where it appears in a sentence. Conversational gestures, controlling gestures, manipulative gestures, and communication gestures are all examples of different sorts of hand gestures. Gestures employed to convey meaning are referred to as SL. It is possible to employ SL as a testing ground for computer vision algorithms due to its great structural complexity [3] [4].

This involves the entire process of tracking and detecting gestures, as well as translating them into words with semantic meaning. In SL communication, non-verbal cues are just as important as hand motions (i.e., manual signing). Hand, facial or body movements can be classified into gestures in SL, which are employed to describe emotions. Research over the last decade has employed on establishing a system for diverse SLs to recognize signs and it has been found that such a system is tough for numerous disciplines like gesture recording method, human action interpretation. Since most sign identification systems are designed to recognize gestures in isolation, this can lead to inaccurate results [5] [6].

Gestural phrases are translated into intelligible speech or text by continuous SL identification systems. True human indications are constant, and any interruption in communication will have a negative impact. Unlike spoken languages, which involve co-articulation of several body signals, SLs have huge vocabularies and reference language that make identification of isolated or continuous signing difficult. This is why the sign identification system is so complicated. There have been a number of new techniques to sign identification recently, but effective automatic SL system identification remains an open issue [7]-[10].

2. LITERATURE REVIEW

This section addresses the many ways for identifying SL. Tubaiz et al. [11] recommended a glove-based Arabic SL identification system employing sequential data classification. Compile an 80-word sensor-based dataset of 40 sentences. Two DG5-VHand data gloves were employed to record hand movements. A camera synchronizes hand gestures with SL sentences. Data temporality is captured and emphasized using simple preprocessing and feature extraction approaches. Then the data is classified using MKNN (Modified k-Nearest Neighbor). The suggested MKNN uses feature vector context to classify accurately. The recommended solution recognized sentences at 98.9%. The findings are compared to a previous vision-based technique using the same texts. The recommended method outperforms vision-based systems in terms of classification rates.

Raheja et al [12] employed a real-time dynamic hand gesture identification system. Pre-processing transformed the collected video to HSV color space, which was then segmented using skin pixels as markers. The results were improved by using Depth information in parallel. Gesture classification was done using Support Vector Machines (SVM) to extract Hu-Moments and motion trajectories from the system studied with a webcam and MS Kinect. Hearing impaired people could use this technique to teach themselves and communicate.

Patel and Ambekar [13] recorded hand motions, analyzed them in MATLAB and outputted speech and text. English and Hindi are chosen for speech and text. The moment approach is employed to assess the image characteristics value. The performance of two classification techniques (PNN and KNN) is compared. This study will facilitate communication between normal and deaf persons.

Hassan et al. [14] demonstrated intelligent identification of static, manual, and non-manual HSL. Fourier descriptor was employed to extract features from RGB digital images. Artificial Neural Network categorizes the data based on the features. High identification accuracy achieved using the particle swarm optimization technique (PSO). The optimized features had a

90.5% identification accuracy compared to 74.8% for the manually selected features. Thus, intelligent HSL identification was successful with high average accuracy.

Liu et al. [15] suggested an LSTM-based end-to-end SLR approach in 2016. This system accepts the trajectories of 4 skeletal joints as inputs and does not design explicit features. We employed Kinect 2.0 to build a vast isolated Chinese SL repertoire. Our approach outperforms standard HMM-based techniques in investigations.

Zadghorban, and Nahvi [16] introduced a new technique for recognizing word boundaries in Persian SL videos. Compared to existing methods in the literature, our algorithm decomposes SL footage into sign words using motion and hand shape information. Separated words are categorized and recognized using a hidden Markov procedure and a hybrid KNN-DTW algorithm. In the absence of a proper Persian SL database, the authors created one with three signers performing sentences and words. The above database was employed to simulate recommended word boundary detection and classification techniques. According to the results, the accurate words boundary detection method achieved 93.73 %, while words identification utilizing hands motion and shape characteristics achieved 92.4 %.

Sawant and Kumbhar [17] introduced a Matlab-based SL Identification system that can recognize 26 Indian SL motions. Four components are pre-processing and hand segmentation, feature extraction, sign identification and text to voice conversion. Image processing is employed to segment. Eigenvalues and Eigenvectors are retrieved and employed in identification. The PCA method was employed to recognize gestures and translate them to text and audio. The recommended technology enables deaf-dumb people to communicate with regular people.

Li et al. [18] developed an innovative frame study for detecting solitary Chinese SL. Specific Hand Shape (SHS) descriptor confers distinguish hand forms. An encoder-decoder LSTM procedure is employed to improve sign identification using the SHS descriptor. An isolated Chinese SL database of 80 words and a hand shape database is employed to test the recommended approaches. The suggested SHS descriptor outperforms the classic HOG descriptor in terms of discrimination and identification efficiency.

Ibrahim et al. [19] developed an automatic visual SLRS that convert Arabic words into text. Hand segmentation, tracking, feature extraction, and classification are recommended stages. Hand segmentation uses a dynamic skin detector based on face color tone. The hands are then tracked using a suggested skin-blob tracking approach. A dataset of 30 isolated words from the daily school life of hearing-impaired children was created to evaluate the suggested method. The suggested system has a 97% identification rate in the signer-independent mode. Also, the suggested occlusion resolution strategy outperforms previous methods by precisely specifying the position of the hands and head with an improvement of 2.57% at $\tau = 5$.

Zhang et al. [20] suggested a frame study using Hidden Markov Procedures (HMMs) that employed the original sign video trajectories and hand-shape information. First, we propose an augmented shape context trajectory feature that can capture spatiotemporal information. Second, we use Kinect mapping routines to get hand areas and HOG per frame (pre-processed by PCA). Moreover, to improve prediction accuracy, we vary the hand shapes rather than the number of hidden states for each sign procedure. For identification, we propose combining trajectories and hand shapes probability. Finally, we test our strategy using our self-built Kinect dataset, and the results show its efficacy.

Hore et al. [21] recommended three unique ways to handle the challenge of ISL gesture identification by merging Neural Network with Genetic Algorithm, Evolutionary Algorithm and Particle Swarm Optimization (PSO) individually. The NN input weight vector has been optimized for the lowest error. The recommended technique was compared to NN and MLP-FFN classifiers. Several performance indicators were calculated which includes accuracy, precision, recall, F-measure and kappa. The investigation findings showed the suggested approach outperformed prior study in recognising ISL motions. On average, the NN-PSO outperformed the other techniques by 0.9956 (Kappa Statistic) and 99.98 (Accuracy).

Tharwat, et al. [22] recommended a new approach that eliminates the need for uncomfortable gear like gloves to facilitate hand identification. The technology uses 2D photos to extract gestures. The SIFT approach mines constant features that are robust in motion. LDA technique is also applied in reducing issues of the extracted features and improves class separation, therefore improving the introduced system accuracy. The Arabic sign characters will be identified using SVM, k-NN, and minimum distance classifiers. Experiments evaluate the suggested system performance revealed near-perfect accuracy of the results. The trials also demonstrated the suggested system robustness against rotation, with an identification rate around 95%. The study also revealed that the system is comparable to the study.

Yang et al. [23] presented an approach for computing the likelihood of HMM is presented to reduce computation complexity. A coarse segmentation method is suggested to yield the maximal level number, together with grammar and sign length constraints. The recommended method outperforms traditional approaches in identification and computation on a KINECT dataset of Chinese SL having 100 sentences of 5 signs each.

Luqman and Mahmoud [24] recommended a rule-based machine translation system between ArSL and Arabic. It analyses the ArSL sentence morphologically and syntactically to translate it into Arabic. Validate this study manually and automatically using a health corpus. The findings indicate that our translation algorithm accurately translates over 80% of the texts.

Stoll et al. [25] employed Neural Machine Translation (NMT), Generative Adversarial Network, and movement generation to produce autonomous SL. Our method can convert spoken language sentences into sign videos. Instead of richly annotated data, our approach uses minimal gloss and skeletal level annotations for training. We do this by splitting the task into sub-processes. We convert spoken language words into sign posture sequences using a Motion Graph and NMT Network. After that, a generative procedure uses the pose data to build photo-realistic SL video sequences. This is the first continuous sign video generating method without a graphical avatar. Our method is tested on the PHOENIX14T SL Translation dataset. On dev/test sets, we reported a BLEU-4 score of 16.34/15.26 for text-to-gloss translation. Our approach video creation capabilities in multi-signer and high-definition environments are quantified using broadcast quality assessment parameters.

ISSN: 2229-6956 (ONLINE)
DOI: 10.21917/ijsc.2022.0357

Table.1. Comparison of existing methods for SL identification

Methods	Merits	Demerits
Modified k-Nearest Neighbor [11]	Superior in categorization rates	Effectiveness depends on the quality of data
SVM [12]	Obtains 97.5% accurate	For the mobile platform, it will not be an advanced and comprehensive system
PNN and KNN [13]	Produces better performance in terms of accuracy	Implemented only for Hindi
LSTM [15]	Provides better performance	Need to improve the SLR performance
KNN-DTW [16]	Obtains higher accuracy	Time consuming nature
Principle Component Analysis [17]	Eliminate the communication barrier	Does not implement for all the phonemes in Marathi signs.
LSTM [18]	Performs better than the traditional techniques	Need to use another approach to increase the performance
Hidden Markov [20]	Effective and robust	Does not conduct experiments on larger datasets
SVM, K-Nearest Neighbor, and minimum distance [22]	Obtains 99% accuracy	The dataset durability cannot be tested by expanding its size.
HMM [23]	Lower computation	Is not using hand shape information to increase identification rate and constructing an explicit search window for level creation to save running time.
Rule-based machine translation system [24]	Provides an accurate translation	Does not tested with real time dataset

To communicate with others, the deaf and hard-of-hearing use SL. The vast majority of the population does not, however, know how to communicate only through the use of SL. As a result, it is difficult for the average person to understand those who are deaf or hard of hearing. SL identification is a complex problem that requires the use of a variety of techniques. These include segmenting the hand area from the image frame and extracting features for training and identification, accounting for motion in some of the sign gestures and environmental noise, as well as the correct identification of the detection method.

Upcoming research in the field intends to assist deaf and mute persons to communicate more effectively with the rest of the world by automatically recognizing signs in films and translating them into SL. It also shows how to create a word-level parallel corpus for multiple dialects of an SL, and how to create a sentence-level translation corpus that includes both the source and the translated sentence.

3. RESULTS AND DISCUSSION

These sections discuss the findings of the SL identification procedure which is performed using MATLAB and show the effectiveness of CNN which is compared with the HMM and PNN procedures regarding accuracy, precision, recall.

Precision: It denotes the ratio of the proper outcomes and is measured as:

$$Precision=TP/(TP+FP) \quad (1)$$

Recall: It denotes the percentage of total proper outcomes categorized accurately by the suggested technique and is measured as

$$Recall=TP/(TP+FN) \quad (2)$$

Accuracy: It is the ratio of accurately obtained identification, and is calculated as follows,

$$Accuracy=(TP+TN)/(TP+TN+FP+FN) \quad (3)$$

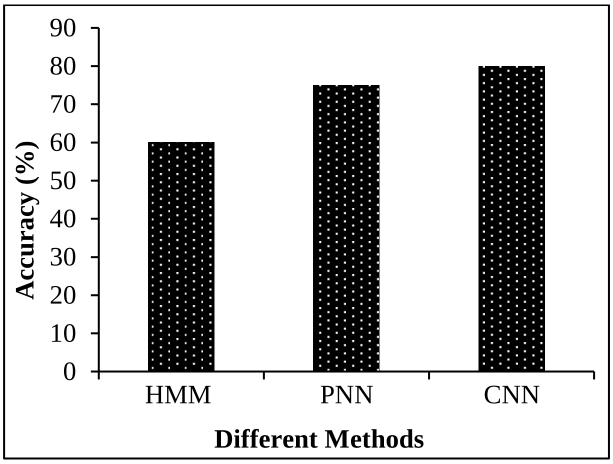

Fig.1. Accuracy results vs. classification methods

Accuracy performance metric comparisons between CNN, HMM, and PNN are shown in Fig.1. From the results, it is concluded that the secured CNN procedure produces the higher Accuracy results of 80% while the HMM and PNN procedures produce only 60% and 75% accordingly.

The Fig.2 shows the performance comparison between CNN, HMM, and PNN in terms of precision. From the results, it is concluded that the CNN procedure produces the higher precision results of 82% while the HMM and PNN procedures produce only 70% and 75% accordingly.

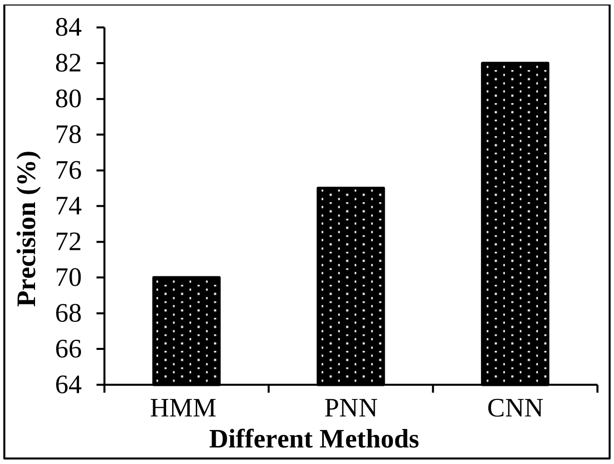

Fig.2. Precision vs. Classification Method

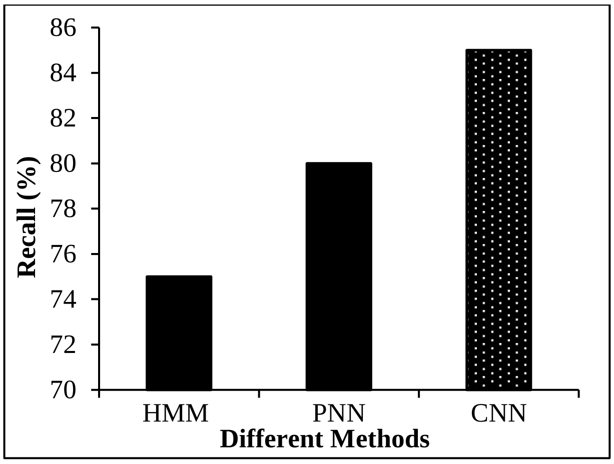

Fig.3. Recall outcomes vs. classification approaches

Accuracy metric contrasted outcomes between CNN, HMM, and PNN are shown in Fig.3. From the results, it is concluded that the CNN procedure produces the higher recall results of 85% while the HMM and PNN procedures produce only 75% and 80% accordingly.

4. CONCLUSIONS

This study refers to the techniques of SL identification and how they study. Each of the strategies presented has its own strengths and disadvantages, and there is no universal SL identification technique. Finally, the open research issues in SL identification are studied. For the deaf community to be involved in the construction of a two-level corpus for multiple dialects of SL, it is decided that future studies in this area should involve the deaf community.

REFERENCES

[1] L. Pigou, S. Dieleman, P.J. Kindermans and B. Schrauwen, "SL Identification using Convolutional Neural Netstudys", *Proceedings of European Conference on Computer Vision*, pp. 572-578, 2014.

[2] P.V.V. Kishore, D.A. Kumar and E.K. Kumar, "Motionlets Matching with Adaptive Kernels for 3-D Indian SL Identification", *IEEE Sensors Journal*, Vol. 18, No. 8, pp. 3327-3337, 2018.

[3] D. Naglot and M. Kulkarni, "Real-Time SL Identification using the Leap Motion Controller", *Proceedings of International Conference on Inventive Computation Technologies*, pp. 1-5, 2016.

[4] Y. Liao, P. Xiong, W. Min and J. Lu, "Dynamic SL Identification based on Video Sequence with BLSTM-3D Residual Netstudys", *IEEE Access*, Vol. 7, pp. 38044-38054, 2019.

[5] L. Geng, X. Ma, B. Xue and Y. Li, "Combining Features for Chinese SL Identification with Kinect", *Proceedings of IEEE International Conference on Control and Automation*, pp. 1393-1398, 2014.

[6] K.M. Lim, A.W. Tan and S.C. Tan, "Block-Based Histogram of Optical Flow for Isolated SL Identification", *Journal of Visual Communication and Image Representation*, Vol. 40, pp. 538-545, 2016.

[7] A. Mittal, P. Kumar, P.P. Roy and B.B. Chaudhuri, "A Modified LSTM Procedure for Continuous SL Identification using Leap Motion", *IEEE Sensors Journal*, Vol. 19, No. 16, pp. 7056-7063, 2019.

[8] X. Yang, X. Chen, X. Cao and X. Zhang, "Chinese SL Identification based on an Optimized Tree-Structure Framestudy", *IEEE Journal of Biomedical and Health Informatics*, Vol. 21, No. 4, pp. 994-1004, 2016.

[9] P.V.V. Kishore, M.V. Prasad and R. Rahul, "4-Camera Procedure for SL Identification using Elliptical Fourier Descriptors and ANN", *Proceedings of International Conference on Signal Processing and Communication Engineering Systems*, pp. 34-38, 2015.

[10] S. Jiang, B. Sun, L. Wang and Y. Fu, "Skeleton Aware Multi-Modal SL Identification", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Identification*, pp. 3413-3423, 2021.

[11] N. Tubaiz, T. Shanableh and K. Assaleh, "Glove-Based Continuous Arabic SL Identification in User-Dependent Mode", *IEEE Transactions on Human-Machine Systems*, Vol. 45, No. 4, pp. 526-533, 2015.

[12] J.L. Raheja, A. Mishra and A. Chaudhary, "Indian SL identification using SVM", *Pattern Identification and Image Analysis*, Vol. 26, No. 2, pp. 434-441, 2016.

[13] U. Patel and A.G. Ambekar, "Moment-based SL Identification for Indian Languages", *Proceedings of International Conference on Computing, Communication, Control and Automation*, pp. 1-6, 2017.

[14] S.T. Hassan, J.A. Abolarinwa and C.O. Alenoghena, "Intelligent SL Identification using Enhanced Fourier Descriptor: A Case of Hausa SL", *Proceedings of IEEE International Conference on Automatic Control and Intelligent Systems*, pp. 104-109, 2017.

[15] T. Liu, W. Zhou and H. Li, "SL Identification with Long Short-Term Memory", *Proceedings of IEEE International Conference on Image Processing*, pp. 2871-2875, 2016.

[16] M. Zadghorban and M. Nahvi, "An Algorithm on Sign Words Extraction and Identification of Continuous Persian SL based on Motion and Shape Features of Hands", *Pattern Analysis and Applications*, Vol. 21, No. 2, pp. 323-335, 2018.

[17] S.N. Sawant and M.S. Kumbhar, "Real-Time SL Identification using PCA", Proceedings of IEEE International Conference on Advanced Communications, Control and Computing Technologies, pp. 1412-1415, 2014.
[18] X. Li, C. Mao, S. Huang and Z. Ye, "Chinese SL Identification based on SHS Descriptor and Encoder-Decoder LSTM Procedure", *Proceedings of Chinese Conference on Biometric Identification*, pp. 719-728, 2017.
[19] N.B. Ibrahim, M.M. Selim and H.H. Zayed, "An Automatic Arabic SL Identification System (ArSLRS)", *Journal of King Saud University-Computer and Information Sciences*, Vol. 30, No. 4, pp. 470-477, 2018.
[20] J. Zhang, W. Zhou and H. Li, "Chinese SL Identification with Adaptive HMM", *Proceedings of IEEE International Conference on Multimedia and Expo*, pp. 1-6, 2016.
[21] S. Hore, S. Chatterjee, V. Santhi and F. Shi, "Indian SL Identification using Optimized Neural Netstudys", *Proceedings of IEEE International Conference on Information Technology and Intelligent Transportation Systems*, pp. 553-563, 2017.
[22] A. Tharwat, T. Gaber and B. Refaat, "Sift-based Arabic SL Identification System", *Proceedings of Afro-European Conference for Industrial Advancement*, pp. 359-370, 2015.
[23] W. Yang, J. Tao and Z. Ye, "Continuous SL Identification using Level Building based on Fast Hidden Markov Procedure", *Pattern Identification Letters*, Vol. 78, pp. 28-35, 2016.
[24] H. Luqman and S.A. Mahmoud, "A Machine Translation System from Arabic SL to Arabic", *Universal Access in the Information Society*, Vol. 19, No. 4, pp. 891-904, 2020.
[25] S. Stoll, N.C. Camgoz and R. Bowden, "Text2Sign: Towards SL Production using Neural Machine Translation and Generative Adversarial Netstudys", *International Journal of Computer Vision*, Vol. 128, No. 4, pp. 891-908, 2020.