Detection of sign language gesture using deep learning and machine learning

Prajwal naik, Nitin Chavan, Piyush Jadhav, Niranjan khavale

Kj's Trinity college of engineering and research pune

Abstract—

Sign language is used as a primary form of communication by many people who are Deaf, deafened, hard of hearing, and non- verbal. Communication barriers exist for members of these populations during daily interactions with those who are unable to understand or use sign language. Based on the depth information of Kinect, this paper studies the real-time dynamic sign language recognition algorithm and improves the dynamic time warping algorithm for the recognition of sign language trajectories. As of the authors knowledge, little to no work exists in the detection of phonetic characteristics in SL, and there is little consensus on what may constitute a phoneme in SL linguistics. In this regard, this work may contribute to further understand the building blocks of SL utterance formation, which may have direct repercussions on how SLs are represented, transmitted and processed electronically.

Index Terms—Sign Language, Clustering, Naive Bayes, Natural Language Processing, Image Thresholding.

I.INTRODUCTION

The rapid development of artificial intelligence and other technologies, gesture recognition as a major humancomputer interaction method has gradually become a hot issue. As a special gesture, sign language is also the main communication method for language-disabled people. According to the World Health Organization, 466 million people worldwide have disabling hearing loss, which can be caused by factors such as birth complications, disease, infection, medication use, noise exposure, and ageing Written communication is a common method of correspondence between Deaf and hearing individuals, but it is considerably slower than both spoken and signed languages. Although deafness is classified as a disability, most members of the Deaf community do not consider themselves disabled, but rather part of a cultural group or language minority. Many studies have used the Microsoft Kinect gaming camera, which has a built-in depth sensor with an infrared projector that triangulates the distance between a signer's hand and the camera. The effectiveness of computer vision-based approaches is limited by environmental factors such as lighting, background conditions, shadows, and camera position. In the literature related to the area, the phoneme concept relies on manual configuration

SELECTION OF PHONETIC UNITS

The selection of the phonetic units in this experiment was obtained through a computational vision algorithm in which a set of sign language videos were analyzed from the DIELSEME database [11]. This open access database is composed of a set of videos that explain the meaning of certain signs of Mexican sign language through examples, and situations where they are normally used. Through a change detection algorithm the selection of these phonetic units was possible, this through the threshold of the image difference, followed by the search of contours to obtain the regions in which two images differ. This value was represented taking the value of the difference threshold and the

series of frames of the video this was draw in a graph.

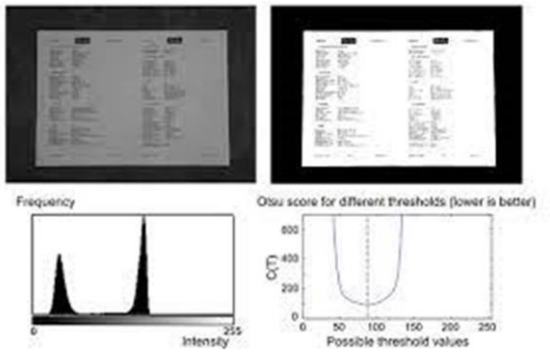

Fig . The difference threshold representing the maximum values in the video

These values are calculated from the computer vision algorithm of the image difference threshold and selected frames and calculate the difference and the threshold of two consecutive frames, as shown in figures 5 to 8. For the detection of phonetics units we validate the results by two methodologies: 1) The first one was validate the information integrity with a set of interpreters, that identifies the set of signs of a message in SL. 2) The second was to group the phonetic units through a clustering algorithm, to analyze if there is enough information to group the phonetic units. As a result of this experimentation, it turned out that the low points kept more information than the high ones, because it was easier for the interpreters to recognize the signs, that is why these frames were considered as phonetic units. As mentioned earlier in this paper, there is no clear definition in the field of linguistics about phonemes in sign languages [13]. However, these units allow us to recover the necessary information to be able to interpret a message correctly and the classification of these phonemes for the processing of language is what allows us to group these units to be able to find patterns of similarity, being able to verify that the phonetic units found in a video can be identifiable in other videos. That is why in this paper we used a soft-clustering algorithm to find these clusters based on the information that can be obtained from the selected frames.

The process to validate the phonetic units It was the following, a total of ten LSM videos taken from the DIELSEME dataset were analyzed, the general objective was to be able to validate the phonemes obtained by the algorithm, through the recognition of different signs, interpretation or understanding of different videos.

This will be possible through the encoding of several videos with different signs and contexts. With the participation of four interpreters to obtain the recognition of the signs and to obtain the loss of information. In figure 5 to 9 are four examples of the videos where there was a greater loss of information.

II. METHODS

A search was conducted on the Web of Science, Association for Computing Machinery, and Institute of Electrical and Electronics Engineers (IEEE) databases for articles published up until December 2019. Search terms of sign language recognition and sign language classification were used to find papers. Paper titles, key terms, and abstracts were examined for study relevance prior to selection. Exclusion terms such as image, vision, and camera were used to narrow results. Non-English articles and those not published in peerreviewed sources were excluded from the study. Since the primary focus of this review was to examine methods for SLR used in mobile applications, only studies involving wearable devices were included.

III. RESULTS

A total of 72 research studies were selected for this analysis based on the search and exclusion criteria. Key parameters related to each study are summarized in Table I. Included are the publishing year, sensor configuration, recognition model, lexicon size, number of subjects, recognition accuracy, and whether the study focused on recognizing isolated signs or continuous sentences. The research questions are discussed individually in the following six sections.

Hand Segmentation and Tracking Method Kinect2.0 can capture the depth information of the image. Segmentation of images using depth information can overcome the effects of traditional hand segmentation in unstable environments such as changes in illumination or complex background changes. This paper uses a hand segmentation method that combines depth threshold and skin color threshold.

The TLD (Tracking-Learning-Detection) algorithm is selected as the method of hand tracking in this paper. It is an algorithm that works well for occlusion, deformation, etc. Moreover, the TLD algorithm continuously updates the "significant feature points" of the tracking module and the target model of the detection module and related parameters through an improved online learning mechanism, thereby making the tracking more stable, robust and reliable.

Boundary Constraint with Relaxed Endpoint In the actual operation of the DTW, it will be found that if the search range is too large, there are many areas where no path branches. Therefore, constraining the search boundary can remove the invalid area, thereby reducing the amount of calculation and improving the calculation efficiency. If the search area is too large, the limited area will have less effect. However, if the area is too small, it will have a greater impact on the recognition rate. It has been found through experiments that the boundary effect is best when the slope is between 0.5 and 1.

IV. EXPERIMENT AND ANALYSIS

First, 70 sign language words were selected for the experiment. The 70 sign language words contain 20 onehanded sign language words and 50 two-handed sign language words. The sign language words vary in length and duration is between 1s and 4s. As shown in Table 1. According to the 70 sign language words, an experimental database was established: a total of 5 sign language players were selected, and the sign language of the above 70 sign language words was collected for each sign language by Kinect2.0. There were 350 sets of sign language data.

As the number of sample templates increases, the time taken to match using the DTW algorithm will increase. Although using endpoint relaxation boundary constraint can't reduce the time complexity of DTW matching, can effectively reduce the amount of computation in the DTW matching process. The partial sequence is eliminated by the lower bound function, which directly reduces the number of template samples. There are more sample symbols in twohand sign language, and each sign language needs to calculate the left and right time series separately, so the effect of LB_BC lower bound distance filtering is more obvious.

V. CONCLUSIONS

In this paper, we analyze that it is possible to group phonetic units from the movements made for the articulation of the Mexican language, in this way, it could be possible to discover a categorization that is not established in the area of linguistics.

The high processing of information and the large volumes of data lead us to believe that this process can become long.

The paper studies the matching algorithm of dynamic sign language: DTW algorithm. Firstly, the principle of DTW is optimized by endpoint relaxation boundary constraint and early termination matching, and the LB_BC lower bound function is used to filter some of the template to be tested. Next, the paper introduces the sign language recognition strategy. First, the matching success rate is improved by assigning weights to the bone points according to the magnitude of the motion of the bone points on the hands. A method for extracting key frames in a trajectory density curve using a sliding window is proposed. Then, the feature vector of the gesture of the key frame is extracted, and the sign language recognition is performed by the method of identifying the motion trajectory and identifying the key gesture.

The high processing of information and the large volumes of data lead us to believe that this process can become long and expensive computationally. And it is a gateway for the development of tools to be able to analyze sign language from its discourse from a computational point of view.

VI .ACKNOWLEDGMENT

We want to thank TCOER Staff for the support for the preparation of this paper. As well as the Head of the depatment in Information technology of the TCOER Pune for the support for the development of this project.

VII. REFERENCES

[1] W. H. Organization, “Deafness and Hearing Loss,” 2018. [Online]. Available: https://www.who.int/news-room/factsheets/detail/deafness-and-hearing-loss.

[2] W. C. Stokoe and M. Marschark, “Sign language structure: An outline of the visual communication systems of the american deaf,” J. Deaf Stud. Deaf Educ., vol. 10, no. 1, pp. 3–37, 2005.

[3] J. Wu, L. Sun, and R. Jafari, “A Wearable System for Recognizing American Sign Language in Real-Time Using IMU and Surface EMG Sensors,” IEEE J. Biomed. Heal. Informatics, vol. 20, no. 5, pp. 1281–1290, 2016.

[4] D. P. Corina, U. Bellugi, and J. Reilly, “Neuropsychological studies of linguistic and affective facial expressions in deaf signers,” Lang. Speech, vol. 42, no. 2–3, pp. 307–331, 1999.

[5] W. C. Stokoe, “Sign Language Structure,” Annu. Rev. Inc., vol. 9, no. 23, pp. 365–390, 1980.

[6] H. Lane, “Ethnicity, Ethics, and the Deaf-World,” J. Deaf Stud. Deaf Educ., vol. 10, no. 3, pp. 291–310, 2005.

[7] H. Brashear, T. Starner, P. Lukowicz, and H. Junker, “Using multiple sensors for mobile sign language recognition,” Seventh IEEE Int. Symp. Wearable Comput. 2003. Proceedings., pp. 45–52, 2003.

[8] U. Bellugi and S. Fischer, “A comparison of sign language and spoken language,” Cognition, vol. 1, no. 2–3, pp. 173–200, 1972.

[9] T. Mohammed, R. Campbell, M. MacSweeney, E. Milne, P. Hansen, and M. Coleman, “Speechreading skill and visual movement sensitivity are related in deaf speechreaders,” Perception, vol. 34, pp. 205–216, 2005.

[10] P. Arnold, "The Structure and Optimization of Speechreading," J. Deaf Stud. Deaf Educ., vol. 2, no. 4, pp. 199–211, 1997.

[11] S. Liddell, Grammar, Gesture, and Meaning in American Sign Language. Cambridge University Press, 2003.

[12] R. Butler, D. Ph, S. Mcnamee, D. Ph, G. Valentine, and D. Ph, "Language Barriers: Exploring the Worlds of the Deaf," vol. 21, no. 4, 2001.

[13] I. M. Munoz-Baell and M. T. Ruiz, "Empowering the deaf. Let the deaf be deaf," J. Epidemiol. Community Health, vol. 54, no. 1, pp. 40–44, 2000.

[14] K. Grobel and M. Assan, "Isolated sign language recognition using hidden Markov models," Syst. Man, Cybern. 1997. …, pp. 162– 167, 1997.

[15] P. Garg, N. Aggarwal, and S. Sofat, "Vision-based hand gesture recognition," IIH-MSP 2009 - 2009 5th Int. Conf. Intell. Inf. Hiding Multimed. Signal Process., vol. 3, no. 1,
pp. 1–4, 2009.

[16] B. Garcia and S. a. Viesca, "Real-time American Sign Language Recognition with Convolutional Neural Networks," Convolutional Neural Networks Vis. Recognit., 2016.

[17] T. Starner, J. Weaver, and A. Pentland, "Real-time American sign language recognition using desk and wearable computer based video," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, no. 12, p. 1371,
1998.

[18] A. Al-Shamayleh, R. Ahmad, M. Abushariah, K. Alam, and N. Jomhari, "A systematic literature review on vision based gesture recognition techniques," Multimed. Tools Appl., vol. 77, no. 21, pp. 28121–28184,
2018.

[19] C. Dong, M. C. Leu, and Z. Yin, "American Sign Language alphabet recognition using Microsoft Kinect," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work., vol. 2015-Octob, pp. 44–52, 2015.

[20] C. Keskin, F. Kiraç, Y. E. Kara, and L. Akarun, "Real time hand pose estimation using depth sensors," Proc. IEEE Int. Conf. Comput. Vis., pp. 1228–1234, 2011.

[21] Z. Parcheta and C.-D. Martínez-Hinarejos, "Sign Language Gesture Recognition Using HMM," in Pattern Recognition and Image Analysis, 2017, pp. 419–426.

[22] P. Kumar, R. Saini, S. K. Behera, D. P. Dogra, and P. P. Roy, "Real-time recognition of sign language gestures and air-writing using leap motion," Proc. - 15th Int. Conf. Mach. Vis. Appl., vol. 1, pp. 157–160, 2017.

[23] P. Kumar, H. Gauba, P. Pratim Roy, and D. Prosad Dogra, "A multimodal framework for sensor based sign language recognition," Neurocomputing, vol. 259, pp. 21–38, 2017.

[24] P. Kumar, H. Gauba, P. P. Roy, and D. P. Dogra, "Coupled HMMbased multi-sensor data fusion for sign language recognition," Pattern Recognit. Lett., vol. 86, pp. 1–8, 2017.

[25] S. Mitra and T. Acharya, "Gesture recognition: A survey," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., vol. 37, no. 3, pp. 311– 324, 2007.

[26] J. Hernandez-Rebollar, R. Lindeman, and N. Kyriakopoulos, "A multi-class pattern recognition system for practical finger spelling translation," Proc. - 4th IEEE Int. Conf. Multimodal Interfaces, pp. 185–190, 2002.

[27] L. Li, S. Jiang, P. B. Shull, and G. Gu, "SkinGest: artificial skin for gesture recognition via filmy stretchable strain sensors," Adv. Robot., vol. 1864, pp. 1–10, 2018.

[28] C. Oz and M. C. Leu, "Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker," Neurocomputing, vol. 70, no. 16–18, pp. 2891–2901, 2007.

[29] V. E. Kosmidou, L. J. Hadjileontiadis, and
S. M. Panas, "Evaluation of surface EMG features for the recognition of American Sign Language gestures," Annu. Int. Conf. IEEE Eng. Med. Biol. - Proc., vol. 2, no. 4, pp. 6197– 6200, 2006.

[30] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and EMG sensors," IEEE Trans. Syst. Man, Cybern. Part ASystems Humans, vol. 41, no. 6, pp. 1064–
1076, 2011.

[31] S. Jiang et al., "Feasibility of wrist-worn, real-time hand, and surface gesture recognition via sEMG and IMU Sensing," IEEE Trans. Ind. Informatics, vol. 14, no. 8, pp. 3376–3385, 2018.

[32] J. L. Hernandez-Rebollar, N. Kyriakopoulos, and R. W. Lindeman, "A New Instrumented Approach For Translating American Sign Language Into Sound And Text," Proc. Sixth IEEE Int. Conf. Autom. Face Gesture Recognit., 2004.

[33] T. D. Bui and L. T. Nguyen, "Recognizing postures in vietnamese sign language with MEMS accelerometers," IEEE Sens. J., vol. 7, no. 5, pp. 707–712, 2007.

[34] M. A. Ahmed, B. B. Zaidan, A. A. Zaidan,
M. M. Salih, and M. M. Bin Lakulu, "A review on systems-based sensory gloves for sign language recognition state of the art between 2007 and 2017," Sensors (Switzerland), vol.
18, no. 7, 2018.

[35] W. Tangsuksant, S. Adhan, and C. Pintavirooj, "American Sign Language recognition by using 3D geometric invariant feature and ANN classification," BMEiCON 2014 - 7th Biomed. Eng. Int. Conf., pp. 1–5, 2015.

[36] U. Bellugi and E. S. Klima, "Sign Language," International Encyclopedia of the Social & Behavioral Sciences. 1996.

[37] T. Takahashi and F. Kishino, "Hand gesture coding based on experiments using a hand gesture interface device," ACM SIGCHI Bull., vol. 23, no. 2, pp. 67–74, 1991.