Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 179 (2021) 541–549

Procedia Computer Science

www.elsevier.com/locate/procedia

5th International Conference on Computer Science and Computational Intelligence 2020

Systematic Literature Review: American Sign Language Translator

Andra Ardiansyah[a] , Brandon Hitoyoshi[a], Mario Halim[a], Novita Hanafiah[b,*], Aswin Wibisurya[c]

[a]*Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480* [b]*Computer Science Department, BINUS Online Learning, Bina Nusantara University, Jakarta, Indonesia 11480*

[c] *Mobile Application and Technology Program, Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

Sign Language Recognition (SLR) is a relatively popular research area yet contrary to its popularity, the implementation of SLR in daily basis is rare; this is due to the complexity and various resources required. In this literature review, the authors have analyzed various techniques that can be used to implement an automated sign-language translator through the analysis of the methodologies and models used to make a working model of any sign-language translator from various sources. The purpose of this study is to explore various possible ways to implement Artificial Intelligence technology to improve the automated American Sign Language translator that is applicable. The authors have identified 22 different research papers within the period of the years 2015 - 2020. The analysis showed that every research studies picked have achieved respectable results, however, they are not perfect, since each research demonstrates its own unique strengths and weaknesses. There are some methods that might be suitable for our need to create an applicable Sign Language Translator, that is by using standard video camera for obtaining data, and either Convolutional Neural Network or Support Vector Machine can be used for the classification.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Computer Science and Computational Intelligence 2020

Keywords: computer vision; sign language recognition; cnn; hmm; image recognition; svm;

* Corresponding author. Tel.: 0858-8378-777-1.
E-mail address: novita.hanafiah@binus.ac.id

1877-0509 © 2021 The Authors. Published by Elsevier B.V.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Computer Science and Computational Intelligence 2020
10.1016/j.procs.2021.01.038

1. Introduction

Communication is mainly done verbally, however not all people can do so because of muteness or deafness. Deafness can be caused by genetics, complications at birth, infectious diseases, chronic ear infections, use of drugs, excessive noise exposure, and aging [1], while muteness can be caused by endotracheal intubation, tracheostomy, or damage to the vocal cords from disease or traumas [2]. Muteness in a way can also be an effect of deafness.

Currently, about 466 million people worldwide have hearing loss, 34 million among which are children, and by 2050, 900 million people are estimated to have hearing loss [1]. With such a staggering number of people suffering from hearing loss, roughly the same number of people will have lost the ability to speak. Despite many workarounds and prevention, those who are unfortunate with these disabilities primarily communicate using sign language.

The use of sign languages have dated back since the 5th century B.C. as stated by Socrates *"If we hadn't a voice or a tongue, and wanted to express things to one another, wouldn't we try to make signs by moving our hands, head, and the rest of our body, just as dumb people do at present?"* [3]. Since the 5th century B.C., there have been numerous versions of signed languages. However, the version focused on this project is American Sign Language (ASL).

Using sign language, the deaf and the mute can somehow communicate. Only some people understand sign language. Should the need for the deaf or mute to speak publicly arise, people usually employ the help of a translator. This paper aims to search for the available technique that would provide the best sign language translator. Which means it has to be efficient and accurate while being accessible for everyone.

2. Study Review

2.1. Planning the Review

The study involves reputable academic articles with correlation to certain keywords either in the title or content as suggested by John Dumay et al. [4]. The keywords are 'computer vision', 'sign language recognition', 'cnn', 'hmm', 'image recognition', and 'svm'. To analyze all the papers and extract essential information, several research questions are made [5]. These research questions which aim to ease the process of gathering and analyzing data are:

RQ1: What are the methods used to obtain the data for the SLR?
RQ2: What are the processes implemented for the SLR?
RQ3: What are the learning algorithms used for classifying in SLR?

Several electronic databases were used as the source of paper, journals, and theses. All of the electronic databases are certified and include only internationally approved academic documents; those are AAAI, BMVC (British Machine Vision Conference), Springer, IEEE Xplore, Ijsret, Semantics Scholar, ACM Digital Library, and Elsevier. To ensure the eligibility of the works selected, three selection and exclusion criteria are applied.

Table 1. Selection Criteria.

Selection Criteria	Exclusion Criteria
Internationally recognized academic paper.	Paper which was published prior to 2015.
Paper must cover SLR stages.	-
Paper must be relevant to the current research paper.	Web documents.

2.2 Conducting the Review

Analysis and review regarding the eligibility of the papers are carried out in this phase. The analysis and review are done based on the selection and exclusion criteria. In result, among the 40 papers previously correlated to the keywords, there are 45 papers eligible to be used as final studies as they met all the selection criteria (Fig. 1). Fig. 2

shows the publication trend. The rate of SLR-related publications fluctuates over the years. Even though there was a huge increase of publications in 2016, articles related to the keywords appeared less in the next years.

2.3 Analysis

After collecting the paper, we analyze the eligible papers based on the research questions. Then, we compare the methods used to view the advantage and disadvantage. The following are our analysis of the reviewed papers.

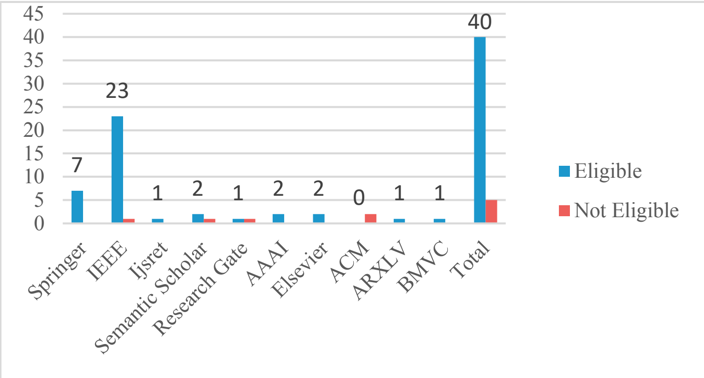

Fig. 1. Papers Eligibility Graph

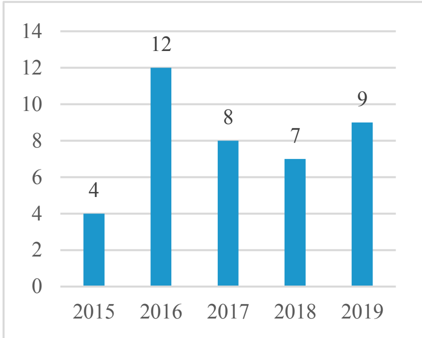

Fig. 2. Papers Years of Publication Graph

2.3.1 RQ1: What method are used to obtain the data for the Sign Language Recognition?

Classification problem falls into supervised learning category. Therefore, obtaining image data along with its label is crucial for training and testing purposes. The data acquisition needs to be consistent since tost classifier requires the image data to be in certain type and shape. Image data can be obtained in various ways. It can be taken manually or by available resources that someone provides on the internet. We have analyzed some of the common method used to obtain data for SLR and compared it in Table 2.

The most common method is using a standard video camera that will give a 2-dimensional image Joshi et al. [6], Jin et al. [7], and some others. This method is more common than the others because nowadays most people have a smartphone with built-in camera. However, the image quality depends on the camera so it's better to be preprocessed first. For example, a low-quality camera has too much noise on the captured image that might interfere with the feature extraction. Also, the resolution of the image depends on the camera so usually an image pre-processing needs to be applied so it can be consistent to the classifier.

Some researches use Kinect [8,9,10]. Kinect has sensors and a camera to capture colored images including the depths of the objects. These provide more detailed data that can improve the classification. The depths feature can help the segmentation by identifying the background and foreground objects. There's also a unique method by Abdelnasser et al. that utilizes wi-fi signal strength to detect in-air hand gestures around the device, called WiGest [11]. The changes in the returned signal becomes the input. This method is useful since the user does not have to carry an object and it's multi-directional. However, it is not easy to implement.

Data can also be obtained from datasets already created on the internet, which are ready to use and usually come in a large amount with good quality. For example, Jalal et al. use the Kaggle dataset which contains about 27000 sign language images [12]. With existing good quality datasets, the research can focus on other parts of the SLR. If the data collected is not enough, Data Augmentation can be done to acquire more data [8,13]. Data Augmentation augments an existing data to create a new different data. This is used to save time and to improve the classifier accuracy by preventing overfitting to some sign language while providing more invariant data to the classifier. In SLR that uses images as input, hand images are captured from different angles and distance, which cause invariance in scale and rotation. A common Data Augmentation example is the rotation, where the image angle is slightly rotated.

Table 2. Data Acquisition Method Comparison.

Reference	Method	Advantage	Disadvantage
6 7 14 15 16 17 18 19 20 21 22 23 24 13	Standard Video Camera	Easy to use; 2D Image data that is easily readable.	Different hardware requires pre-process to the image; Noise.
8,9,10	Kinect	Accurate data; Image and depth provide more detail to the data; Easier segmentation process.	Less Expensive; Sensitive to external infrared source.
11	WiGest	Multi-directional	Complex implementation.
12,25,26,27,22,13	Web Obtained Dataset	Ready to use; Comes in a large amount No input hardware needed; Saves time and resources.	-
8,13	Data Augmentation	Provide more data; Prevent overfitting	-

2.3.2 *RQ2: What are the processes that is implemented for the Sign Language Recognition?*

From the reviewed papers, there are several processes that is applied after obtaining the data. These processes are taken to help the classifier perform better by enhancing the data or reducing some unnecessary data. There are 3 processes identified amongst all the SLR research, that is the Pre-processing, Segmentation, and Feature Extraction.

2.3.2.1 *Pre-processing*

Pre-processing is the stage where the data is changed after obtained. It's also to help the classifier to perform better by reducing some bad data that may cause an inaccuracy. We have analyzed some of the common pre-processing methods that are used from the reviewed papers and compare it in Table 3. Some of the well-known techniques are Gaussian Filter [14,23] and Median Filter [10,21] that will reduce unwanted noise in a 2-dimensional image to improve edge detection on the segmentation stage. There's also a method where the image size is cropped or resized, so the classifier doesn't have to compute too much and have consistent data format, for example like converting an HD image to a 28x28 image.

Table 3. Pre-processing Method Comparison.

Reference	Method	Advantage	Disadvantage
14 23	Gaussian Filter	Reduce noise; Smoothen the image.	Reduce details.
10,21	Median Filter	Filter out the noise; Preserve sharp features.	No Error Propagation.
10,26	Image Cropping	Consistent input.	-
20	Bootstrapping	Avoid loss of information; Simple implementation.	Time consuming.

2.3.2.2 *Segmentation*

For SLR, the most important part is the hand gesture which is acknowledged by the hand movements. Therefore, image segmentation is applied to remove unwanted data like the background and other objects as the input, that might interfere with the classifier calculation. Image segmentation works by limiting the region of the data, so the classifier will only look at the Region of Interest (ROI). For this stage, the researcher must understand what part of image is important for the classifier. Analysis of the segmentation methods used in the papers are listed in Table 4.

One of the methods in segmentation is Gray scaling [6,7,17]. It turns RGB/Colored Image to a grayscale image. Sign Language does not care for the skin of the person, so researchers usually applied gray scaling first before going further. This will make the classifier avoid taking color as consideration for the calculation. And the important part is that gray scaling also helps identifying the background and foreground.

The most common method in segmentation is Thresholding [17,18,21,10,23,24]. Usually applied after the image is gray scaled, thresholding turns the image into binary form. Since gray scaling turns an image black and white, thresholding is applied to identify the background and foreground where black represents background and white represents foreground. A threshold value will be chosen by the researcher and that value will determine which color will be the background and which will be the foreground. Joshi et al. use Otsu Algorithm, an algorithm that automatically determines a good threshold value, so the researcher does not need to find a good threshold value [6].

Another method is Skin Segmentation [16,18,22,24]. This method is popular because of the simplicity. It takes a predetermined color range or color histogram of human skin so the image will only display the selected color. This way, the classifier won't receive any unwanted information such as the background or objects. Researcher need to determine the color space to use first. The downside is that sometimes objects with similar color to the skin (e.g. other body parts or face) will also get detected. Skin segmentation is sensitive to illumination.

Furthermore, a Morphological filter or operation is usually applied to the binary image [6,14,18,23]. Morphological filter help reducing an error either from the foreground or the background, effectively increasing the region of interest to fit the object. Moreover, there's a Canny edge detection method which is used to extract edges from the image [6,7,11,18]. This method is good to map out the hand's edges and determine the Region of Interest.

Some also use Background Subtraction [18,23]. Background Subtraction requires video input with the static objects like the backgrounds removed. Since the signer hand will move around, this method is used for an easy segmentation process. However, if there are many moving objects like cars, television, etc., it will also take it as the foreground.

Table 4. Segmentation Method Comparison.

Reference	Method	Advantage	Disadvantage
6,7,17	Gray scaling	Low computation; Simple implementation.	-
17,18,21,10,23,24	Thresholding	Low computation; Fast performance.	-
6	Otsu Algorithm	Automatic threshold value.	-
6,14,18,23	Morphological Filter	Effective in growing the Region of Interests.	Features might have less detailed.
6,7,11,18	Canny Edge Detection	Effective against noisy and various environment.	High computation; Time consuming.
7	Seeded Region Growing	Fast image segmentation; Robust and effective for growing the Region of Interests.	Sensitive to noise; Slow.
14	Sobel Edge	Simple implementation.	Increased noise.
16,18,22,24	Skin Segmentation	Simplicity of the implementation.	Sensitive to illumination.
18,23	Background Subtraction	Low computation.	Depend on frame rate and object speed; Sensitive to illumination.
18	Viola and Jones Algorithm	Removal of face.	Sensitive to lighting condition and invariant rotation.

2.3.2.3 Feature Extraction

Feature Extraction is the part where relevant information from the data is taken and enhanced. It cuts out the redundant information from the region of interest and start taking the features to be used for the classifier calculation. For SLR, that features might vary depends on what the researcher thinks can be taken for recognizing gestures. For example, Ahmed et al. [24] take the center of mass while Kumar et al. [20] take the fingertips position and direction to help determine the gestures. The feature extraction methods from the SLR papers are listed in Table 5.

The most common feature extraction is by using the convolution layers of Convolutional Neural Network (CNN) like the ones used in the paper of Jalal et al. [12], Huang et al [19], and some others. The convolution layers can extract important features from the image. This feature is not just taken, but also enhanced by the MaxPooling layer and increase the performance of the calculation. Which is why CNN is the most commonly used because of its robustness.

Another method, Principal Component Analysis (PCA), is used to reduce the features/data dimension, transforming the variable to be uncorrelated. Rao et al. [14] use PCA to increase the classification performance and reduce overfitting. Before applying the PCA, Discrete Cosine Transform is used to compress the image. There's also PCANet [9], a Principle Component Analysis Network that is computationally efficient for extracting features.

The Speeded Up Robust Features (SURF) is used by Jin et al. [7]. The method was based on Scale-Invariant Feature Transform (SIFT). Like SIFT, SURF can detect local features of a data to find its interest points and it is more computationally efficient. This method is robust against scaling, rotation, occlusion, and variation for finding features. Which is why this method is good for image classification so that the classifier can handle different sign rotations.

Table 5. Feature Extraction Method Comparison.

Reference	Method	Advantage	Disadvantage
14	Discrete Cosine Transform	Fast algorithm.	Require additional steps.
14	Principle Component Analysis	Improve performance; Reduce overfitting.	Require data standardization.
7	Speeded Up Robust Features	Robust against invariant data; Efficient and faster compared to SIFT.	Can sometimes false matching.
9	PCANet	Efficient computation.	Space complexity.
11	Discrete Wavelet Transform	Easy to filter noise for signal.	Complex implementation.

2.3.3 RQ3: What are the learning algorithms used for classifying in Sign Language Recognition?

Different methods used for the classification process are shown in Table 6. The most common method is the Convolutional Neural Network (CNN) because of its accuracy. An example of CNN is the work of Cayamcela et al. [15] where the accuracy is 99.39%. Unlike Artificial Neural Network (ANN), CNN adds extra convolutional layers for feature extraction. After that, it determines on which node that data belongs to and keep learning from the result of the training. The more layers it has, the better accuracy it gets but it also gets computationally expensive.

Table 6. Feature Extraction Method Comparison.

Reference	Method	Advantage	Disadvantage
6	Cross-Correlation Coefficient	Low computation.	Too simple for classification. Does not learn.
7,9	Support Vector Machine	Memory efficient. Effective for classification.	Low performance when noise is in the data.
17,27	Artificial Neural Network	Able to learn. Robust fault-tolerant network.	Slow convergence speed.
8,20,26	Hidden Markov Model	Efficient learning algorithm.	High computation. Need many training data.
12,19,25,21,22,10,23	Convolutional Neural Network	Highly accurate for image classification. Can work well even without segmentation or pre-processing.	High computation. Need strong hardware.
15,16,26,13	Convolutional Neural Network with Transfer Learning	Highly accurate for image classification. Saves time since it is pre-trained.	High computation. Need strong hardware. Preprocessing needed to fit the network.

Most of the previous CNN needs a lot of testing to get the layers right. Some apply CNN with Transfer Learning, so it does not have to worry about the layer and only focus on the training and predicting part. Cayamcela et al. [15] and Shahriar et al. [16] AlexNet from Google for their research, and it has good performance on classifying hand gestures.

Jin et al. [7] and Aly et al. [9] use Support Vector Machine (SVM), a method that is well-known for its ability to solve margin maximization optimization problems. The solution creates a Decision Boundary that will help greatly in generalizing linear classifiers. It's also better than some linear classifiers and non-linear classifiers.

Another method is the Hidden Markov Model (HMM), maximizing the posterior probability of where the data belong to while minimizing the error of prior probability. Guo et al. [8] use a HMM based on Gaussian Mixture Model, named GMM-HMM. Their research proves that the model can improve the accuracy compared to normal ones. Joshi et al. [6] use the Cross-Correlation Coefficient. This method measures the similarity of two signals in different time-shifted functions. It proves to be good for dynamic gestures which is important in sign language. And it's not performance heavy so it can be implemented on normal hardware. There's also a Minimum Distance Classifier. Rao et al. [14] use Minimum Distance Classifier with Mahalanobis Distance for an efficient classification. While it's not good at handling different hand shapes or sizes, it is the cheapest method than the others so it can be implemented on many hardware.

3. Discussion

This section gives an overview of the methods used and the result from the previous studies on SLR. Based on the studies, there are 5 common steps of SLR (shown in Fig. 3): data acquisition, pre-processing, segmentation, feature extraction, and classification. Each stage play an important role to help the performance of later stages.

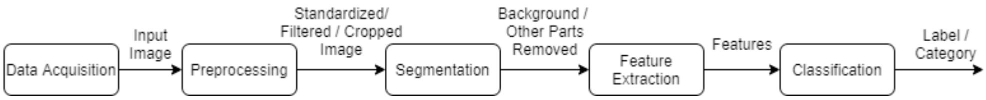

Fig. 3. Steps of Sign Language Recognition

The techniques used for the 5 stages and the accuracy from the previous SLR works is presented in Table 7. While each works may have a different type of gesture recognition for example like static or dynamic gestures recognition, we choose to merge all of them in one table and only focus on the methods used.

To obtain data, the most popular method is using a camera. Others use Kinect, or existing datasets from various sources. Some researches apply preprocessing like cropping, grayscaling, gaussian smoothing, while others do not.

Table 7. Papers Techniques and Result Overview.

Ref	Data Aqcuisition	Pre-processing	Segmentation	Feature Extraction	Classification	Accuracy
6	Standard Video Camera	Grayscaling, Morphological Filtering	Otsu Algorithm	Edge Detection	Cross-Correlation Coefficient	94%
7	Standard Video Camera	Grayscaling	Canny Edge, Seeded Region Growing	SURF	K-Means, Bag of Features, SVM	97.13%
14	Standard Video Camera	Gaussian Filter	Sobel Edge, Morphological Gradient	Discrete Cosine Transform, PCA	Min. Distance Classifier	90.58%
12	Kaggle ASL Dataset	-	-	CNN	CNN	99%
15	Standard Video Camera	-	-	CNN	CNN with Transfer Learning	99.39%
16	Standard Video Camera	-	Skin Segmentation	CNN	CNN with Transfer Learning	94.70%
8	Kinect and Data Augmentation	-	-	-	GMM-HMM	-
17	Standard Video Camera	Grayscaling	Thresholding	Contour-based Technique	ANN	95%
9	Kinect	-	-	PCANet	SVM	88.70%
11	WiGest	-	Special Preamble	Discrete Wavelet Transform, Edge Detection	Match String Pattern with Gesture Template	96%
18	Standard Video Camera	-	Background Subtraction, Thresholding, Canny Edge, Skin Segmentation, Morphology Operator	Contour Extraction	Heuristic Assumption	86.66%
19	Standard Video Camera	-	-	CNN	CNN	82.70%

25	RWTH-PHOENIX-Weather 2014 Continuous SLR Dataset	-	-	CNN	CNN	-
20	Standard Video Camera	Bootstrap resampling	-	Fingertip Position and Direction	Coupled-HMM	90.80%
21	Standard Video Camera	Median filter	Thresholding	CNN	CNN	96.20%
26	RWTH-PHOENIX-Weather 2012, RWTH-PHOENIX-Weather 2014, SIGNUM single signer Datasets	Image Cropping	-	CNN	CNN with Transfer Learning, HMM	-
27	Indian Sign Language Dataset	-	-	-	ANN, Genetic Algorithm, Particle Swarm Optimization	99.96%
22	Massey University Gesture Dataset 216, Standard Video Camera	-	Skin Segmentation, Convex Hull Algorithm	CNN	CNN	98.05%
10	Kinect	Image Cropping, Median Filter	Thresholding	CNN	CNN	91.70%
23	Standard Video Camera	Gaussian Filter	Background Subtraction, Thresholding, Morphological Transformation	Contour Extraction, CNN	CNN	99.80%
24	Standard Video Camera	-	Skin Segmentation, Thresholding	Centre of Mass, Distance Measures	Dynamic Time Warping	90%
13	ASL Dataset from University of Surrey, Standard Video Camera, Data Augmentation	-	-	CNN	CNN with Transfer Learning	98%

All the studies performed relatively well. Among the plethora of different studies selected in Table 7, the most popular feature extraction and classification method is Convolution Neural Network (CNN) including the hybrid methods that involve CNN like CNN-HMM method [26]. Other studies use ANN [17], HMM [26] [16], etc. for classification. For feature extraction, aside from CNN, some researchers use SURF [7], Principal Component Analysis [9] [14], etc.

4. Conclusion

In this study, we have reviewed 22 research articles related with Sign Language Recognition (SLR). From the mentioned researches, the SLR process can be divided into 5 common steps: acquisition, pre-processing, segmentation, feature extraction, and classification. Standard video camera is most commonly used for data acquisition to get hand images from various persons, angles, lightings backgrounds, and sizes. Other researchers prefer to use datasets obtained from the internet. Other techniques include Kinect, WiGest, and data augmentation.

CNN is the most widely used classifier, which accounts for 11 out of 22 mentioned researches. CNN is popular because of its accuracy, which can reach 90% or more for SLR tasks. We noticed some drawbacks of several researches. Some models with the highest accuracy only allow static image input. In ASL, some alphabets require hand movement, and therefore, need a real-time, live translator. Overall, these study researches have been insightful since it shows different approaches for sign language translation.

References

1. Deafness and hearing loss. [Online].; 2020. Available from: https://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.
2. Tietze K. Clinical Skills for Pharmacists 3rd Edition; 2011.
3. Ohna SE. Open your eyes: deaf studies talking. Scandinavian Journal of Disability Research. 2010 Jun; 12(2): p. 141-146.
4. Dumay J, Bernardi C, Guthrie J, Demartini P. Integrated reporting: A structured literature review. Accounting Forum. 2016 Sep 22; 40(3): p. 166-185.
5. Ávila-Pesántez D, Rivera LA, Alban MS. Approaches for Serious Game Design: A Systematic Literature Review. [Online].; 2017. Available from: http://www.asee.org/documents/papers-and-publications/papers/CoEd_Journal-2017/Jul-Sep/AVILA_PESÁNTEZ.pdf.
6. Joshi A, Sierra H, Arzuaga E. American sign language translation using edge detection and cross correlation. 2017 IEEE Colombian Conference on Communications and Computing, COLCOM 2017 - Proceedings. 2017.
7. Jin CM, Omar Z, Jaward MH. A mobile application of American sign language translation via image processing algorithms. Proceedings - 2016 IEEE Region 10 Symposium, TENSYMP 2016. 2016;: p. 104-109.
8. Guo D, Zhou W, Wang M, Li H. Sign language recognition based on adaptive HMMS with data augmentation. Proceedings - International Conference on Image Processing, ICIP. 2016; 2016-Augus: p. 2876-2880.
9. Aly W, Aly S, Almotairi S. User-independent american sign language alphabet recognition based on depth image and PCANet features. IEEE Access. 2019; 7: p. 123138-123150.
10. Pigou L, Dieleman S, Kindermans PJ, Schrauwen B. Sign Language Recognition Using Convolutional Neural Networks. In Pigou L, Dieleman S, Kindermans PJ, Schrauwen B..; 2015. p. 572-578.
11. Abdelnasser H, Harras KA, Youssef M. WiGest demo: A ubiquitous WiFi-based gesture recognition system. Proceedings - IEEE INFOCOM. 2015; 2015-Augus: p. 17-18.
12. Jalal MA, Chen R, Moore RK, Mihaylova L. American Sign Language Posture Understanding with Deep Neural Networks. 2018 21st International Conference on Information Fusion, FUSION 2018. 2018;: p. 573-579.
13. Brandon Garcia , Sigberto Alarcon Viesca. Real-time American Sign Language Recognition with Convolutional Neural Networks. [Online]. Available from: http://cs231n.stanford.edu/reports/2016/pdfs/214_Report.pdf.
14. Rao GA, Kishore PVV. Sign language recognition system simulated for video captured with smart phone front camera. International Journal of Electrical and Computer Engineering. 2016; 6(5): p. 2176-2187.
15. Cayamcela MEM, Lim W. Fine-tuning a pre-trained Convolutional Neural Network Model to translate American Sign Language in Real-time. 2019 International Conference on Computing, Networking and Communications, ICNC 2019. 2019;: p. 100-104.
16. Shahriar S, Siddiquee A, Islam T, Ghosh A, Chakraborty R, Khan AI, et al. Real-Time American Sign Language Recognition Using Skin Segmentation and Image Category Classification with Convolutional Neural Network and Deep Learning. IEEE Region 10 Annual International Conference, Proceedings/TENCON. 2019; 2018-Octob(October): p. 1168-1171.
17. Thongtawee A, Pinsanoh O, Kitjaidure Y. A Novel Feature Extraction for American Sign Language Recognition Using Webcam. BMEiCON 2018 - 11th Biomedical Engineering International Conference. 2019;: p. 1-5.
18. Yeo HS, Lee BG, Lim H. Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools and Applications. 2015; 74(8): p. 2687-2715.
19. Huang J, Zhou W, Zhang Q, Li H, Li W. Video-based sign language recognition without temporal segmentation. 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. 2018;: p. 2257-2264.
20. Kumar P, Gauba H, Roy PP, Dogra DP. Coupled HMM-based multi-sensor data fusion for sign language recognition. Pattern Recognition Letters. 2017; 86: p. 1-8.
21. Flores CJL, Cutipa AEG, Enciso RL. Application of convolutional neural networks for static hand gestures recognition under different invariant features. Proceedings of the 2017 IEEE 24th International Congress on Electronics, Electrical Engineering and Computing, INTERCON 2017. 2017;: p. 5-8.
22. Taskiran M, Killioglu M, Kahraman N. A Real-Time System for Recognition of American Sign Language by using Deep Learning. 2018 41st International Conference on Telecommunications and Signal Processing, TSP 2018. 2018;: p. 1-5.
23. Xu P. A Real-time Hand Gesture Recognition and Human-Computer Interaction System. 2017 Apr 24.
24. Ahmed W, Chanda K, Mitra S. Vision based Hand Gesture Recognition using Dynamic Time Warping for Indian Sign Language. In 2016 International Conference on Information Science (ICIS); 2016: IEEE. p. 120-125.
25. Cui R, Liu H, Zhang C. Recurrent convolutional neural networks for continuous sign language recognition by staged optimization. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017. 2017; 2017-Janua: p. 1610-1618.
26. Koller O, Zargaran S, Ney H, Bowden R. Deep sign: Hybrid CNN-HMM for continuous sign language recognition. British Machine Vision Conference 2016, BMVC 2016. 2016; 2016-Septe: p. 136.1-136.12.
27. Hore S, Chatterjee S, Santhi V, Dey N, Ashour AS, Balas VE, et al. Optimized Neural Networks. 2017;: p. 139-151.