ISSN 2393-9850
GEM — Georgian Educational Mine
Interdisciplinary International Peer-Reviewed Research Journal
St. George's College Aruvithura, Kerala, India

Volume 3, Issue 1, 2023
Pages 34–55
gem.sgcaruvithura.ac.in

ORIGINAL RESEARCH ARTICLE

Towards a Flexible and Cost-Effective System for Dynamic Gesture Prediction in Indian Sign Language

Jestin Joy[1,*], Sreeraj M[2]

[1]Department of Computer Applications, St. George's College Aruvithura, Kerala, India
[2]Department of Computer Science, Sree Ayyappa College, Eramallikkara, Kerala, India

[*]**Corresponding author:** Jestin Joy, jestinjoy@sgcaruvithura.ac.in

Received: 9 October 2023 · **Revised:** 20 November 2023 · **Accepted:** 15 December 2023 · **Published:** 18 December 2023

Keywords: Sign Language;Gesture Classification;Machine Learning;Depth Sensor

Abstract

Hand gesture is one of the most commonly used method for communication by the deaf people. It is mainly used to communicate among themselves. Though various sign language systems have been developed around the world, they are neither flexible nor cost-effective for the end users. Furthermore, lack Sign Language dynamic hand gesture datasets is also a challenge. This work presents a system for dynamic two handed Indian Sign Language gesture prediction. Six classification algorithms were evaluated on depth data. Results indicate that Support Vector Machines (SVM) and Deep Neural Networks (DNN) has reasonable accuracy for classifying Indian Sign Language dynamic signs.

1. INTRODUCTION

Sign language is widely used by individual with hearing impairment to communicate with each other conveniently using hand gestures. However hearing people find it very difficult to communicate with those with sign language users since interpreters are not readily available at all times. A robust recognition system without the use of costly hardware that can classify the gestures in real time will be much useful. Automatic Sign Language Recognition (ASLR) can be used to translate from sign language to spoken language. But this is a difficult task considering the fact that signs are made up of manual components and parallel modifications. Lack of understanding of Sign Language grammar is also a challenge.

According to World Health Organization (WHO)[WHO [2018]] estimates there are around 466 million with hearing disability. This study focusses on Indian Sign Language(ISL) and can easily be extended to other sign languages. India Census 2011[INDIA [2011]] estimates that there are around five million people with hearing disability. This accounts for 18.9% of the total disabled population in India. In India, number of schools that cater to the needs of deaf are very less. This results in higher unemployment rate for deaf population. Teaching in Sign Language is very helpful for the upliftment of deaf students. Officially recognizing sign languages, increasing the availability of sign interpreters and providing transcription in sign languages greatly improve accessibility. Unavailability of Indian Sign Language(ISL) dictionary, lack of understanding of ISL grammar and lack of availability of ISL interpreters badly affect the teaching learning process of the deaf students.

ALSR techniques are used for developing many helpful applications[Joy et al. [2019]]. Unlike traditional methods for capturing gestures, depth based sensors are favoured due to its insensitivity to illumination and colour changes, reliability in tracking movements and the ability to get 3D structural information. Though depth sensors are used for efficient ASLR, their acquisition is not that easy considering the high cost involved. This paper focuses on using low cost depth camera for ASLR. Use of depth enabled sensors[Mittal et al. [2019], Xiao et al. [2019], Quesada et al. [2017], Yang et al. [2019], Yu et al. [2020]] for ASLR is investigated extensively for ASLR. Techniques for different Sign Languages are investigated by many researchers. For this, different depth enabled sensors are investigated. Intel RealSense Depth Camera is selected for this work because of taking cost of acquisition into consideration. Most research[Guzsvinecz et al. [2019], Khalaf et al. [2019]] on dynamic sign recognition use Kinect, Leap Motion Controller(LMC), Intel RealSense and wearable sensors[Kudrinko et al. [2020]] for capturing depth information. Intel RealSense depth cameras are equipped with depth module and vision processor. The smaller form factor and Inertial Measurement Unit(IMU) makes it easy to integrate into other products. For calculating depth, a pair of identical cameras and infrared projector is used. Pair of cameras captures the scene and send it to vision processor, which calculates depth values for each pixel in the image. These are then processed to generate a depth frame. An infrared projector is also used to improve depth accuracy in scenes with low texture. Taking size into consideration, Leap Motion Controller(LMC) is similar to RealSense Depth Camera and is mainly used to track hands. LMC supports 120×150 degree field of view. LMC is equipped with a pair of 40x240-pixel near-infrared cameras. Kinect is a much efficient depth sensor as compared to Intel RealSense and LMC for tracking purpose. But size and cost are its disadvantages. RGB cameras, infrared projectors and Microsoft relased Software Development Kit (SDK) makes it a good choice for gesture recognition. Skeletal tracking facilities provided by Kinect is best[Guzsvinecz et al. [2019]] in its class. The use of WiFi[Zhang et al. [2020]] is also investigated by some authors.

Understanding the adoption of assistive technologies is critical for design and development of new solutions. Studies[Kirkpatrick [2018]] have shown that adoption rates of assistive technology products are very low. High cost incurred in the acquisition of hardware based technology products is another problem. Assistive technology abandonment[Sears and Jacko [2007], Scherer [1996]] is also another problem facing assistive technology products.

Following contributions are made in this work

1. Development of dynamic sign data-set[1] with and without segmentation for Indian Sign Language (ISL)
2. Comparative analysis of different machine learning approaches for sign recognition
3. Study of classification performance on similar two handed dynamic signs
4. Evaluating the effectiveness of low cost hardware for Indian Sign Language recognition

[1] Joy, Jestin (2020), "*Indian Sign Language Depth data (Week days)*", Mendeley Data, V1, doi: 10.17632/smt3f5bkj8.1

2. RELATED WORKS

There are lots of research on Automatic Sign Language Recognition(ASLR) for different sign languages. Work ranges from recognizing static and dynamic signs to translating continuous sign utterances to spoken language text. This work focuses on recognising dynamic signs with focus on Indian Sign Language (ISL).

In literature, mainly Kinect[Kumar et al. [2018b], Raghuveera et al. [2020], Kumar et al. [2017]] based depth sensors are used for ISL recognition. This study considers signs with involvement of both hands. Better recognition accuracy is obtained using methods that involved more number of Infrared Sensors[Kishore et al. [2018], Kiran Kumar et al. [2018], Kumar et al. [2018a]]. But they are not easily adaptable to applications level settings in a low cost environment. Leap Motion Controller[Mittal et al. [2019]] is the most accessible sensor and studies have shown that they have reasonable recognition accuracy. Details regarding techniques explored for Indian Sign Language (ISL) is discussed in Table 1. Athira, Sruthi and Lijiya[Athira et al. [2019]] proposed an SVM classifier based system for recognising single handed static and dynamic gestures, double-handed static gestures from video.

Paper	Sensor	Technique	Signs	Results
Mittal et al. [2019]	Leap Motion Controller	LSTM	35 Signs	Accuracy-89.5%
Kishore et al. [2018]	8 IR cams and 1 RGB camera	Motionlets Matching With Adaptive Kernels	500 Signs	Accuracy-98.9% , Precision-98.3% , Recall-97.8%
Kiran Kumar et al. [2018]	8 IR cams and 1 RGB camera	Early estimation model	350 signs	Accuracy-97.38%
Kumar et al. [2018a]	8 IR cams and 1 RGB camera	Spatio Temporal Graph Kernels	500 signs	Accuracy-98.4%
Kumar et al. [2017]	Kinect and Leap Motion Controller	HMM+BLSTM-NN	50 signs	Single Hand Sign Accuracy-97.85%; Double Hand Sign Accuracy-94.55%
Kumar et al. [2018b]	Kinect	HMM	30 signs	Accuracy-83.77%
Raghuveera et al. [2020]	Kinect	SVM	140 signs	Accuracy-71.85%
Athira et al. [2019]	RGB	SVM	11 signs	Accuracy-89%

Table 1. ISL recognition techniques

Similar research exist for other sign languages also, with more work existing in American Sign Language(ASL). Details regarding techniques explored for other sign languages are discussed in Table 2.

Bo Liao, Jing Li, Zhaojie Ju et al[Liao et al. [2018]] proposed a double channel (colour and depth) CNN based method for classifying ASL finger spelled signs. These signs lack enough dynamic and

Paper	Sensor	Technique	Sign Language and Signs	Results
Lu et al. [2016]	Leap Motion Controller	Hidden Conditional Neural Field(HCNF)	ASL- Leap Motion-Gesture3D-12 Signs;HandicraftGesture-10 Signs	Accuracy: 89.5%(LeapMotion-Gesture3D); 95.0%(Handicraft-Gesture)
Avola et al. [2018]	Leap Motion Controller	RNN	ASL-12 Dynamic + 18 Static	Accuracy-96.41%, Precision-96.64%, Recall-96.41%
Huang et al. [2018]	Kinect	Keyframe-Centered Clips+LSTM	CSL, 310 Signs	Accuracy-91.18%
Yu et al. [2019]	sEMG	Deep Belief Net	CSL, 150 Signs	Accuracy-95.1%
Sun et al. [2013]	Kinect	Discriminative Exemplar Coding(DEC)	ASL, 73 Signs	Accuracy-86.8%
Aly et al. [2019]	Kinect	PCANet	ASL, 24 Signs	Accuracy-88.7%
Liao et al. [2018]	RealSense	CNN	ASL, 24 Signs	Accuracy-99.4%

Table 2. ISL recognition techniques

two handed gestures making it easy[Joy et al. [2019]] for classifiers. Y. Zhang et al.[Zhang et al. [2020]] proposed a fusion based method for detecting static+dynamic hand gestures from strain sensors. For 16 static and 8 dynamic gesture class accuracy of 88.75% was achieved by the best method(SVM+DTW). Dynamic gestures considered in their study mainly involved movement of hands with ample difference between each gesture as compared to signs compared in this study. W. Aly, S. Aly and S. Almotairi[Aly et al. [2019]] proposed a CNN+SVM based method for classifying ASL static signs from Kinect depth images. An accuracy of 88.7% is achieved for this method.

Our study has shown that research on ISL recognition using low cost sensors is not investigated thoroughly. We aim to fill this gap. Details of which follows in the following sections.

3. STUDY DESIGN

Depth data from sensor is captured and both segmented and raw versions are stored for processing. These are then fed to dimensionality reduction algorithm to reduce the features. This is done for segmented and raw data. This is then fed to the classification algorithms and evaluated.

3.1 Data preprocessing

Signs captured are stored as a sequence of frames where $s_i = \{f_j : 1 \leq j < 25\}$ where f is the frames having the size $424X240$ corresponding to each sign $s_i \varepsilon S$ ie, $S = \{s_i : 1 \leq i < 7\}$. Each frame is considered as a very long 1D vector by $Fv = fv_i \frown fv_{i+1} : 1 < i < 25$ frame pixels column by column , i.e.$fv_i : 424X240 = 101760$ and $Fv_i : 424X240X25 = 2544000$.

Depth data

$$D = \begin{bmatrix} [d_1^1]_{240X424} & [d_2^1]_{240X424} & .... & [d_{25}^1]_{240X424} \\ [d_1^2]_{240X424} & [d_2^2]_{240X424} & .... & [d_{25}^2]_{240X424} \\ .... & & & \\ .... & & & \\ [d_1^{160}]_{240X424} & [d_2^{160}]_{240X424} & .... & [d_{25}^{160}]_{240X424} \end{bmatrix}$$

is converted into one-dimensional data using striding window indexing method and thus the position of any depth data can be considered as a linear combination of corresponding indices, as follows. Each block consists of

$$shape = (d_0, d_1, \dots d_{N-1})$$

where, N, is, the, dimension and stride is defined as

$$(s_0, s_1, \dots, s_{N-1})$$

and each stride S_i is calculated based on the window size W, stride step St and it is slided $((E - W)//St) + 1$ times where E is the number of elements of data and $//$ is the integer division with index=$(I_0, I_1, \dots, I_{N-1})$. So each position can be calculated as position=$\sum_{l=0}^{N-1} s_l I_l$ and if item size=α then we can represent the data into either column-major as $s_l^{column} = \Pi_{j=0}^{l-1} \alpha d_j$ and row-major as $s_l^{row} = \Pi_{j=l+1}^{N-1} \alpha d_j$. From this blocks are created with respect to non-overlapping views of the data. Calculating the local mean of elements in each block of size by a factor and thus we can down sample the entire depth data into a reduced size of

$$D_{red\_size} = \begin{bmatrix} [d_1^1]_{60X142} & [d_2^1]_{60X142} & \cdots & [d_{25}^1]_{60X142} \\ [d_1^2]_{60X142} & [d_2^2]_{60X142} & \cdots & [d_{25}^2]_{60X142} \\ \cdots & & & \\ \cdots & & & \\ [d_0^{160}]_{60X142} & [d_2^{160}]_{60X142} & \cdots & [d_{25}^{160}]_{60X142} \end{bmatrix}$$

A total of 160 observation for each s_i dataset with total of size $NXM : N = 160, M = 2544000$ is used.

We consider the transpose of data set of each s_i includes N observations, each of the variables M normally $n >> m$. To reduce the dimensionality of the dataset of each s_i, so that only L variables are necessary to represent each observation., $1 \leq L < M$.The empirical mean is subtracted from each column of of each s_i. The mean square error ε^2 is calculated for the approximation of higher dimensional space X (of dimension M) as follows

$$\varepsilon^2 = \frac{1}{N} \sum_{n=1}^{N} |\mathbf{x}_n|^2 - \sum_{i=1}^{L} \mathbf{b}_i^T cov(x) \mathbf{b}_i,$$

where $\mathbf{b}_i, i = 1, ..., L$ are basis vector of the linear space of dimension L and if $\sum_{i=1}^{L} \mathbf{b}_i^T cov(x) \mathbf{b}_i$ is maximum then ε^2 will minimum and

$$cov(x) = (\frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_n \mathbf{x}_n^T)$$

can be guaranteed to have real eigen-values. These values may be sorted in decreasing order and consider the basis vectors as corresponding associated eigen-vectors. Complete dimensions are achieved by omitting the smallest values of eigen values and vectors and its mean square is calculated as

$$\varepsilon^2 = trace(cov(x)) - \sum_{i=1}^{L} \lambda_i = \sum_{i=L+1}^{M} \lambda_i.$$

where $trace(cov(x))$ is the sum of the diagonal elements of the cov matrix and equals the sum of all eigenvalues.

3.2 Classification Algorithms

Six classification algorithms are evaluated on the dataset collected.

Support Vector Machine (SVM) is a supervised learning method that is effective in high dimensional spaces. The particular approach implemented here employs “one-versus-one” approach where, $\frac{n\_classes*(n\_classes-1)}{2}$ classifiers are constructed. Each of these classifiers trains data from two classes.

K-NN is a non parametric method used for classification and outputs k nearest neighbours. Based on the user defined value of k, a sample is s classified by assigning the sign class which is most frequent among the nearest k training samples.

A Deep Neural Network (DNN) is a an artificial neural network (ANN), which uses multiple hidden layers between the input and output. DNN is very useful in classification tasks. Figure 1 describes the DNN architecture used in this work. DNN computations are made up of matrix based computations which makes it easy run on GPU’s. DNNs are feedforward networks and can model complex non-linear relationships. The network in this paper uses softmax activation function in output layer and Rectified Linear Unit (ReLU) activation in other hidden layers. Softmax classifier is mainly used in multi-class classification problem.

For a feature vector $Z^{[L]}$, the probability of each class is given by the Softmax classifier as

$$a^{[L]} = \frac{e^{Z[L]}}{\Sigma_{j=1}^{n} e^{Z[L]}}$$

where n is the number of classes. And

$$Z^{[L]} = W^{[L]}a^{[L-1]} + b^{[L]}$$

where W is weight vector and b is bias vector

The softmax function $a^{[L]}$ returns a probability value for each class in the range [0,1]

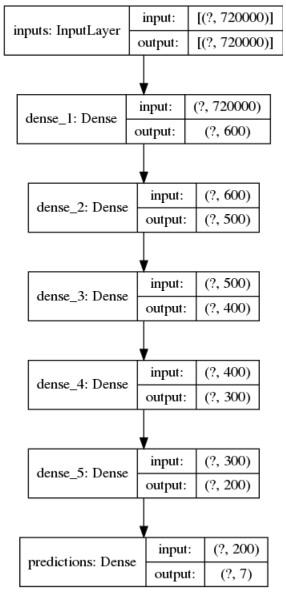

Figure 1. DNN architecture

Decision tree classifier is a non-parametric supervised learning method that classifies based on the simple decision rules inferred from the sample.

Logistic regression model models probability of output in terms of input and can be used as a classifier by classifying inputs with probability greater than the cut-off as one class, below the cut-off as the other. In multi-class classification case, the algorithm uses the one-vs-rest (OvR) scheme.

Naive Bayes is a probabilistic classifier and probability of given sample a falls into class c_i is given by

$$p(x = a \mid c_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(a-\mu_i)^2}{2\sigma_i^2}}$$

where μ_i is the mean of values in the sample associated with class c_i and σ_i is the variance of values in the sample associated with class c_i. The parameters μ_i and σ_i are estimated using maximum likelihood.

3.3 Dataset

Since no data set exists for Indian Sign Language (ISL), we created one with seven signs. It covers seven signs which corresponds to days of a week. Three different signers are used to record the signs. Each sign is made up of 25 frames and is of the size 424x240. Signs released by Indian Sign Language and Research Center (ISLRTC) is used as a reference for creating the signs. These signs are released[2] for public use, since datasets like these are not available for Indian Sign Language (ISL). Segmented versions of these signs are also used in the experiments. Threshold based segmentation is used for segmenting the captured frames. Figure 2 shows a sample captured frame with and without segmentation. Every sample is tagged with its gesture name and is stored as a single array of

(a) Frame without Segmentation **(b)** Frame with Segmentation

Figure 2. Sample captured frame

depth values corresponding to the depth image of the gesture. Dataset is collected in uncontrolled environments and each subject possesses variation in orientation, speed and size. Weekdays are selected as signs for study because these signs are two handed, dynamic and similar. These features makes it a good candidate for dynamic gesture classification study.

4. EXPERIMENTAL EVALUATION

This section aims to evaluate the effects of different machine learning algorithms for ISL recognition. Six classification algorithms are used in this study. Table 4 shows recognition results after applying different algorithms for segmented and not segmented versions.

ISL signs are captured using Intel®RealSense™ Depth Camera D435i. Intel RealSense Depth Camera D435i comes under D4300 family of depth camera by Intel and uses stereo vision to calculate depth. It supports resolution of upto 1280x720 for stereo depth and 1920x1080 for RGB. It also supports Depth Diagonal Field of View over 90 degree. 25 frames are captured for each sign. Each frame is of size 424x240 and its dimension.

Dataset is divided into training and test set. 70% of the data is used for training and 30% is used for testing. Accuracy, precision, recall and F1 score metrics are reported for each algorithm.

[2] Joy, Jestin (2020), "*Indian Sign Language Depth data (Week days)*", Mendeley Data, V1, doi: 10.17632/smt3f5bkj8.1

Since we are dealing with a multi-class classification problem, tp and fp for each class is calculated and its mean is taken. Since this does not take class imbalance into account, weighted average is reported.

All the experiments are done on a system with 32GiB RAM, Intel 3.7GHz Core i7-8700K CPU with NVIDIA TITAN Xp GPU. Titan Xp GPU is based on Pascal architecture and has 3840 CUDA cores and is equipped with 12 GB GDDR5X memory.

Algorithm	Data Set	Accuracy	Precision	Recall	F1 Score
SVM	Segmented	0.99	0.99	0.99	0.99
	Not Segmented	0.96	0.96	0.96	0.96
DNN	Segmented	0.99	0.99	0.99	0.99
	Not Segmented	0.98	0.98	0.98	0.98
Naive Bayes	Segmented	0.68	0.72	0.68	0.67
	Not Segmented	0.55	0.60	0.55	0.56
Logistic Regression	Segmented	0.98	0.98	0.98	0.98
	Not Segmented	0.88	0.88	0.89	0.88
KNN	Segmented	0.99	0.99	0.99	0.99
	Not Segmented	0.92	0.93	0.92	0.92
Decision Tree	Segmented	0.78	0.78	0.78	0.78
	Not Segmented	0.89	0.89	0.89	0.89

Table 3. ISL sign recognition results - Using downscaling

Algorithm	Data Set	Accuracy	Precision	Recall	F1 Score
SVM	Segmented	0.80	0.80	0.82	0.81
	Not Segmented	0.89	0.89	0.89	0.89
DNN	Segmented	0.77	0.79	0.78	0.78
	Not Segmented	0.89	0.90	0.89	0.89
Naive Bayes	Segmented	0.73	0.75	0.75	0.73
	Not Segmented	0.73	0.77	0.75	0.74
Logistic Regression	Segmented	0.68	0.69	0.71	0.69
	Not Segmented	0.80	0.81	0.81	0.79
KNN	Segmented	0.73	0.79	0.76	0.74
	Not Segmented	0.79	0.81	0.79	0.78
Decision Tree	Segmented	0.60	0.62	0.62	0.61
	Not Segmented	0.69	0.71	0.70	0.70

Table 4. ISL sign recognition results - Using PCA

Results show that among the six classifications algorithms considered SVM and DNN algorithms are proved to be better at classifying signs from depth data. Decision tree classifier was the worst performer in terms of the metrics considered. Also dimensionality reduction on segmented samples resulted in the reduction in accuracy as compared to samples which are not segmented when PCA is used for dimensionality reduction. Among the signs considered, all the algorithms could easily classify signs corresponding to Wednesday and Thursday. For most classification algorithms the signs corresponding to Monday, Friday and Sunday are shown to have less classification accuracy. Among the signs considered Sunday is the one having the most movement.

5. CONCLUSION AND FUTURE SCOPE

This study examined the effect of machine learning algorithms for Indian Sign Language (ISL) classification from depth data. Low cost depth sensor used for data acquisition making it a good choice for practical applications. Results obtained shows that SVM and DNN based methods have good accuracy for classifying dynamic gestures which include both hands for the signer. The number of signs used for recognition is limited to seven. The effect of machine learning algorithms on more signs is worth exploring. Increase in the number of signs is challenging from the implementation point of view. More dimensionality reduction techniques should also be investigated for this purpose. In the present study only depth data is used for classification. Sensor used in this study can easily capture RGBD information. Classification based on RGBD data is also worth looking at. In sign language, manual components like facial expressions are important. Multi stream recognition techniques are worth exploring because of this.

REFERENCES

W. Aly, S. Aly, and S. Almotairi. User-independent american sign language alphabet recognition based on depth image and pcanet features. *IEEE Access*, 7:123138–123150, 2019.

P.K. Athira, C.J. Sruthi, and A. Lijiya. A signer independent sign language recognition with co-articulation elimination from live videos: An indian scenario. *Journal of King Saud University - Computer and Information Sciences*, 2019. ISSN 1319-1578. doi: https://doi.org/10.1016/j.jksuci.2019.05.002. URL http://www.sciencedirect.com/science/article/pii/S131915781831228X.

Danilo Avola, Marco Bernardi, Luigi Cinque, Gian Luca Foresti, and Cristiano Massaroni. Exploiting recurrent neural networks and leap motion controller for the recognition of sign language and semaphoric hand gestures. *IEEE Transactions on Multimedia*, 21(1):234–245, 2018.

Tibor Guzsvinecz, Veronika Szucs, and Cecilia Sik-Lanyi. Suitability of the kinect sensor and leap motion controller—a literature review. *Sensors*, 19(5):1072, 2019.

S. Huang, C. Mao, J. Tao, and Z. Ye. A novel chinese sign language recognition method based on keyframe-centered clips. *IEEE Signal Processing Letters*, 25(3):442–446, 2018.

POMPI INDIA. *Census of India 2011 Provisional Population Totals*. Office of the Registrar General and Census Commissioner New Delhi, 2011.

J. Joy, K. Balakrishnan, and M. Sreeraj. Signquiz: A quiz based tool for learning fingerspelled signs in indian sign language using aslr. *IEEE Access*, 7:28363–28371, 2019.

Ahmed S. Khalaf, Sultan A. Alharthi, Igor Dolgov, and Z O. Toups. A comparative study of hand gesture recognition devices in the context of game design. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*, ISS '19, page 397–402, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450368919. doi: 10.1145/3343055.3360758. URL https://doi.org/10.1145/3343055.3360758.

E. Kiran Kumar, P.V.V. Kishore, D. Anil Kumar, and M. Teja Kiran Kumar. Early estimation model for 3d-discrete indian sign language recognition using graph matching. *Journal of King Saud University - Computer and Information Sciences*, 2018. ISSN 1319-1578. doi: https://doi.org/10.1016/j.jksuci.2018.06.008. URL http://www.sciencedirect.com/science/article/pii/S1319157818303719.

Keith Kirkpatrick. Technology for the deaf, 2018.

P. V. V. Kishore, D. A. Kumar, A. S. C. S. Sastry, and E. K. Kumar. Motionlets matching with adaptive kernels for 3-d indian sign language recognition. *IEEE Sensors Journal*, 18(8):3327–3337, 2018.

K. Kudrinko, E. Flavin, X. Zhu, and Q. Li. Wearable sensor-based sign language recognition: A comprehensive review. *IEEE Reviews in Biomedical Engineering*, pages 1–1, 2020.

D. Anil Kumar, A.S.C.S. Sastry, P.V.V. Kishore, and E. Kiran Kumar. 3d sign language recognition using spatio temporal graph kernels. *Journal of King Saud University - Computer and Information Sciences*, 2018a. ISSN 1319-1578. doi: https://doi.org/10.1016/j.jksuci.2018.11.008. URL http://www.sciencedirect.com/science/article/pii/S1319157818307304.

Pradeep Kumar, Himaanshu Gauba, Partha Pratim Roy, and Debi Prosad Dogra. A multimodal framework for sensor based sign language recognition. *Neurocomputing*, 259:21 – 38, 2017. ISSN 0925-2312. doi: https://doi.org/10.1016/j.neucom.2016.08.132. URL http://www.sciencedirect.com/science/article/pii/S092523121730262X. Multimodal Media Data Understanding and Analytics.

Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, and Debi Prosad Dogra. A position and rotation invariant framework for sign language recognition (slr) using kinect. *Multimedia Tools and Applications*, 77(7):8823–8846, 2018b.

B. Liao, J. Li, Z. Ju, and G. Ouyang. Hand gesture recognition with generalized hough transform and dc-cnn using realsense. In *2018 Eighth International Conference on Information Science and Technology (ICIST)*, pages 84–90, 2018.

Wei Lu, Zheng Tong, and Jinghui Chu. Dynamic hand gesture recognition with leap motion controller. *IEEE Signal Processing Letters*, 23(9):1188–1192, 2016.

A. Mittal, P. Kumar, P. P. Roy, R. Balasubramanian, and B. B. Chaudhuri. A modified lstm model for continuous sign language recognition using leap motion. *IEEE Sensors Journal*, 19(16):7056–7063, 2019.

Luis Quesada, Gustavo López, and Luis Guerrero. Automatic recognition of the american sign language fingerspelling alphabet to assist people living with speech or hearing impairments. *Journal of Ambient Intelligence and Humanized Computing*, 8(4):625–635, 2017.

T Raghuveera, R Deepthi, R Mangalashri, and R Akshaya. A depth-based indian sign language recognition using microsoft kinect. *Sādhanā*, 45(1):34, 2020.

Marcia J. Scherer. Outcomes of assistive technology use on quality of life. *Disability and Rehabilitation*, 18(9):439–448, 1996. doi: 10.3109/09638289609165907. URL https://doi.org/10.3109/09638289609165907. PMID: 8877302.

Andrew Sears and Julie A Jacko. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications*. CRC press, 2007.

Chao Sun, Tianzhu Zhang, Bing-Kun Bao, Changsheng Xu, and Tao Mei. Discriminative exemplar coding for sign language recognition with kinect. *IEEE Transactions on Cybernetics*, 43(5): 1418–1428, 2013.

WHO. Deafness and hearing loss, 2018. URL http://www.who.int/news-room/fact-sheets/detail/deafness-and-hearing-loss.

Q. Xiao, M. Qin, P. Guo, and Y. Zhao. Multimodal fusion based on lstm and a couple conditional hidden markov model for chinese sign language recognition. *IEEE Access*, 7:112258–112268, 2019.

L. Yang, Y. Zhu, and T. Li. Towards computer-aided sign language recognition technique: A directional review. In *2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, volume 1, pages 721–725, 2019.

Y. Yu, X. Chen, S. Cao, X. Zhang, and X. Chen. Exploration of chinese sign language recognition using wearable sensors based on deep belief net. *IEEE Journal of Biomedical and Health Informatics*, 24 (5):1310–1320, 2020.

Yi Yu, Xiang Chen, Shuai Cao, Xu Zhang, and Xun Chen. Exploration of chinese sign language recognition using wearable sensors based on deep belief net. *IEEE Journal of Biomedical and Health Informatics*, 24(5):1310–1320, 2019.

Lei Zhang, Yixiang Zhang, and Xiaolong Zheng. Wisign: Ubiquitous american sign language recognition using commercial wi-fi devices. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(3):1–24, 2020.

Y. Zhang, Y. Huang, X. Sun, Y. Zhao, X. Guo, P. Liu, C. Liu, and Y. Zhang. Static and dynamic human arm/hand gesture capturing and recognition via multiinformation fusion of flexible strain sensors. *IEEE Sensors Journal*, 20(12):6450–6459, 2020.