

PERFORMANCE EVALUATION OF COMBINED CONSISTENCY-BASED SUBSET EVALUATION AND ARTIFICIAL NEURAL NETWORK FOR RECOGNITION OF DYNAMIC MALAYSIAN SIGN LANGUAGE

¹SUTARMAN, ²MAZLINA ABDUL MAJID, ³ENNY ITJE SELA

^{1,3}Faculty of Information Technology and Business
University Technology of Yogyakarta, Indonesia

²Faculty Computer System and Software Engineering
Universiti Malaysia Pahang, Malaysia

E-mail: ¹sutarman@uty.ac.id, ²mazlina@ump.edu.my, ³ennysela@uty.ac.id

ABSTRACT

Dynamic sign language recognition is important to intelligent human-computer interaction technology, but it is also very difficult to deal with, especially when the environment is quite complex. This paper proposes the use of consistency-based Subset Evaluation and Artificial Neural Network (ANN) in order to increase accuracy rate in the recognition of sign language. At the first stage, image acquisition data gets from the Kinect sensor, using skeletal data tracking with eight joint positions. The second stage, the skeletal feature extraction (data X, Y, and Z are taken the value relative to the torso and head; spherical coordinate a conversion process; segmentation of the frame to get the same number of dimensions). The third stage, the selection of feature data using the Consistency-based Subset Evaluation algorithms with the best first search method subset. Then last stage is the classification using Artificial Neural Networks, by variations of nodes in hidden layer. The data samples tested are 15 dynamic signs taken from the dynamic of Malaysian Sign Language (MySL). The results of the experimental show that our system can recognize with accuracy of 93.67%. The feature selection can contribute toward the improvement in the accuracy rate of sign language recognition data

Keywords: *Malaysian Sign Language, Sign Language Recognition, Artificial Neural Network, Kinect, Consistency-based Subset Evaluation*

1. INTRODUCTION

Sign language is fundamental alternative methods of communication between the deaf and the several dictionary words or a single letter has been determined to make this communication possible. Sign language used by deaf and mute difficult to understand by the general public they feel isolated by the surrounding environment. It is a combination of form and hand movement, orientation and movement within the body between arms, hands and facial expressions to express thought's speaker[1]. The conversation between two mute people is possible with the help of sign languages like the Malaysian Sign Language (MySL). However, it is difficult for them to communicate with people having no prior knowledge about sign language.

ASL consists of signs and gestures, with each having a particular meaning. By taking images of a person doing the Sign in real-time and converting

these visual images into text or audio, a mute person can communicate effectively with someone having little or no knowledge of sign language. Sign language recognition (SLR) aims to translate sign language into text so that the communication among deaf and hearing impaired to be convenient.

Researchers have used various methods for the sign language recognition such as [2]–[7] used Hidden Markov Models (HMM), [8], [9] used Artificial Neural Network (ANN). Recognition of dynamic Malaysian sign language using Artificial Neural Network has been done by [10]. The sample data using 15 Malaysian sign of sign language and the result are 80.56% accuracy. Feature selection as the classification process in SLR has proposed by [11]. They proposed Hill Climbing approach and Random Walk approach, to select the features. They claimed that both algorithms were easy-implemented but reasonable and efficient. Feature selection potential to be used in SLR. Currently,

there is no research work that used Feature Selection on Sign Language Recognition.

This paper is organized into several sections. Section 2 describes the related work, including sign language, sign language Malaysia, Consistency-based Subset Evaluation, The best first search method subset and Artificial Neural Network. Section 3 describes the methodology, including Dataset Collection, Image Acquisition, skeleton Feature Extraction, normalization of the data and the recognition rate. The experiments and results are described in Section 4, and the last section 5 contains of the conclusions in this work.

2. RELATED WORK

2.1 Sign Language

Sign languages are commonly developed in deaf communities, which can include interpreters and friends and families of deaf people as well as people who are deaf or hard of hearing themselves. Sign language are classified under several categories such as controlling gestures, manipulative gestures, conversational gestures, and communicative gestures [12]. There are two types of sign language static and dynamic sign language [13]. Static sign language is defined as the orientation and position of the hand in a certain time without any movement. Dynamic Sign language, including movements that involve parts from the body such as waving a hand motion while static includes the establishment as a single without motion like jamming a thumb and forefinger to form the "Ok" static pose symbol that represents the static movement.

2.2 Feature Selection

According [14], Feature Selection is often used as a preprocessing step for machine learning, where a subset of the features available from the selected data for the application of learning algorithms. It is a process of choosing a subset of original features so that the feature space is optimally reduced according to a certain evaluation criterion. The best subset contains at least the number of dimensions that most contribute to accuracy, to remove the remaining dimensions is not an important [15]. Feature selection is the combination of the two operational components: evaluation of feature subsets; and searching for feature subsets. During the feature subset search, features are evaluated and assigned a value of evaluation measure; in heuristic search, this measure is evaluated to guide the search further or to terminate the search. When the feature selection process is completed, the evaluation

measure values are assessed and the best evaluated feature is selected [16]. There have been many algorithms of attributes selection developed, among others, is an algorithm that is based on the level of correlation of data, the level of consistency among the data, and based on set theory. Selection is done to select the attributes that are significant to the data class. If the data is not good or could not represent a class in the dataset, the process of induction method of classification would be more difficult, and may affect the performance of the classification. Selection of attributes subset-based is algorithms that perform an analysis of the subset of attributes that generated, either at random by genetic algorithms [17] or be greedy, or known as the Best First Search [18]. Attribute selection process was conducted using the software WEKA [19].

2.2.1 Consistency-based Subset Evaluation (CSE)

Success in the classification and data mining on a problem is determined by many factors. One factor is the quality to the data or information held. If the information held is irrelevant or contains redundancies, or data obtained containing high noise, then the model extraction process to the data is more difficult [20]. Therefore, the selection process required attributes, to sort out the good attributes and is not relevant. Already many attributes of selection algorithms are developed, among others, is an algorithm based upon the level of correlation of data, the level of consistency among the data, and based upon set theory. According to [21] there are several approaches to the use of the consistency class in attribute subset selection for the evaluation metrics. Namely to find the number of combinations of attributes, to divide the data subsets contain strong majority of one class. Usually, a search bias supports a subset of small features with high-grade consistency. Based consistency evaluator parts used by Liu and Setiono [21] consistency metric (Eq.1).

$$\text{Consistency}_s = 1 - \frac{\sum_{i=0}^J |D_i| - |M_i|}{N} \quad (1)$$

where s is an attribute subset, J is the number of distinct combinations of attribute values for s ,

$|D_i|$ is the number of occurrences of the i attribute value combination; $|M_i|$ is the cardinality of the majority class for the i^{th} attribute value combination

and N is the total number of instances in the data set.

2.2.2 The Best first search method subset

The aim of feature selection is to decide which attributes of a number of attributes that will be incorporated into the final section and are ignored. If there are n attributes at first, then it is likely the number of subsets is 2^n pieces. If n is large enough, then looking for as many as 2^n subset fruit becomes unreasonable. Therefore, a wide range of heuristic strategies is used to find some of the best subsets that make sense, among all possible subsets that can be formed from the data set. One of the heuristic strategies that are often applied to find the subset is hill climbing and best first [18].

In general, this subset search method is a method of searching a subset of attributes in the hill climbing technique plus the backtracking. A specified parameter to set how many nodes can be increased in a row, which is used to control the level of backtracking. The best first search starts with an empty set of features and generates all possible single feature expansions. The subset with the highest evaluation is selected and expanded by adding a single feature. If expanding a subset results in no improvement, the search drops back to the next-best unexpanded subset and continues from there.

The time of best first search will explore the entire space features section, so it is common to limit the number of expanded subset that produces no improvement. Best-first search is used in the final experiments as it gave slightly better results for some cases than hill climbing.

2.3 Artificial Neural Network

Artificial Neural Network (ANN) is a branch of Artificial Intelligence (AI) and has been accepted as a new computing technology in computer science fields. AI defined as intelligence exhibited by an artificial entity to solve complex problems, and such a system is generally assumed to be a computer or machine [22]. The ANN refers to a network or circuit of biological neurons. It is composed of neurons interconnected used to solve problems or concerns biological neural artificial intelligence. According to [23], ANN is defined as a structure of densely interconnected adaptive simple processing (artificial neurons or nodes) that can perform massively parallel computation for data processing and knowledge representation. Although ANNs drastic abstraction of biological counterparts, ideas ANNs not replicate the operation of biological systems but to make use of

what is known about the function of biological networks to solve complex problems. In recent years have many researchers which have used the ANN method for recognition in sign language, such as: [9], [24]–[29].

3. THE METHODOLOGY

There are four stages in Malaysian Sign Language Recognition System:

- Data acquisition Sign Language of kinect using MatLab 2013a
- The skeletal feature extraction is taking special characteristics of the data kinect.
- Applying Consistency-based Subset Evaluation (CSE) in feature selection algorithms and the best first search method subset.
- Combine Consistency-based Subset Evaluation and Artificial Neural Networks for dynamic Malaysian Sign Language, by doing variations of nodes in hidden layer and use number of joint 8

3.1 Dataset Collection

List of the dynamic Malaysian Sign Language as shown in Table 1, recorded with five different people. Subjects were asked to repeat five times for the same signs. Samples of data on the form of X, Y, and Z from data obtained from Kinect. Total sample data 375 in Malaysia Sign Language. As a reference to the sign 15, we use video performed by an expert in sign language, she is a teacher of school extraordinary.

Table 1: The list of 15 Dynamic Malaysian Sign Language [10]

No	Malaysian Sign language	No	Malaysian Sign language
1	Apa kabar	9	Mengapa
2	Assalamu'alaikum	10	Pagi
3	Bahasa isyarat	11	Petang
4	Belajar	12	Saya
5	Cikgu	13	Selamat
6	Gembira	14	Siapa
7	Kamu	15	Terima kasih
8	Makan		

3.2 Image Acquisition

Generally, image acquisition in image processing can be defined as a way to retrieve a picture or video from a source, usually a hardware-based source, so it can be passed through to

whatever process needs to occur afterward. In this project, the image is retrieved using kinect sensor with the use of skeletal-based algorithm. The next stage is the pre-processing. There are few pre-processing operations applied such as subtraction, edge detection, and normalization to improve the segmented hand image [30]. Pre-processing in this study is based on skeleton tracking the joints to produce 3D coordinates X, Y, Z.

Skeletal Tracking allows Kinect to follow the actions of the people. The joints of the tracked users in space can be located, and their movements can be followed. The skeletal tracking recognizes users in standing or sitting position, and facing the sensor of the camera. In the default mode for range of the kinect can detect people who stand between 0.8 meters and 4.0 meters. The practical range is 1.2 meter to 3.5 meter, which allows for the hand movement of the user [31]. To track users, first enable the skeletal tracking in Kinect. The information about the tracked users is provided in the form of an array of skeleton objects present within the frame. The skeletons in a frame can be in a state of 'tracked' or 'position only'. The problems within the sign language recognition system are different physical characteristics of the users such as body size (children and adults) and the position of the camera sensor. Hence, the process of normalization is needed to cater the users of different sizes, whilst the camera sensor positions are detected by the system. In this project, each frame of the joint list tracked reduced from 20 to eight as show in Figure 1 which are the corresponding positions of the joints and the notation that will be used.

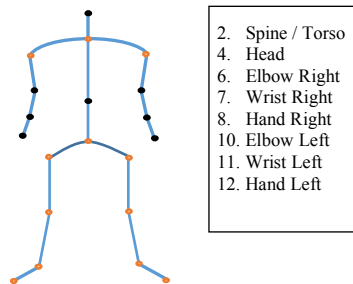


Figure 1: Used joints

3.3 Feature Extraction

3.3.1 Invariant to user's positions (Torso)

The normalization process consider the position of the user. For the position of the user is within range as the data can be stored according to its position. As shown in Fig 2., the slight variations in the depth can cause variations in the value of X and Y. The distance between the joint,

and the other can vary depending on the user's position.

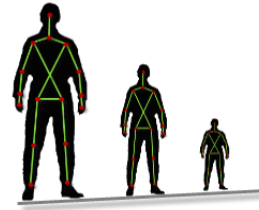


Figure 2: Invariant to user's positions

The normalization of all coordinates connected with Torso position. This position remains constant along the sign frames is true and will be used to create a position-invariant system. By using the Cartesian coordinates X, Y, and Z and the Spherical coordinates to consider the torso as the origin stored. A spherical coordinate system is a coordinate system of three-dimensional space in which the position from the point is determined by three dimensions: the radial distance from the point of origin fixed, the polar angle measured from the zenith fixed, and the azimuth angle of the orthogonal projection on the reference plane that passes through the point of origin to zenith, measured from a fixed reference direction. [32], [33]. Fig. 3(a) show three numbers or values and correspondence of the three a value in the system shown in Fig. 3(b).

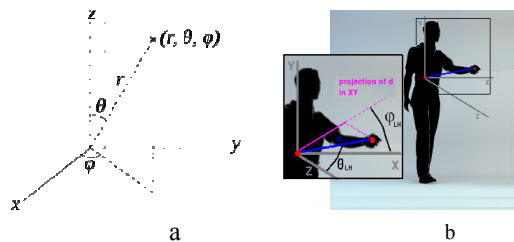


Figure 3: Use of the spherical coordinates [34]

Radial distance r is expressed by d, and d is defined as the vector between Torso and concerned joints (theta and phi) are the angle that describes the direction of the vector 3D.

3.3.2 Invariant to user's size (Head)

In the sign language translator, the system must be able to translate the receipt of the user either tall or short, so that the translator can produce output as the right word in every case. The distance between one joint to another, changing significantly depending on the size of the user, as shown in Fig. 4. After the user's position is normalized, every joint and two angles theta and phi which describes the orientation of the distance.

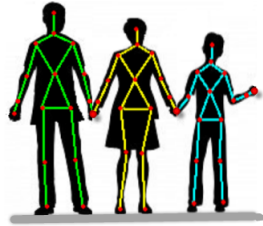


Figure 4: Normalization Required For The User Sizes

Normalization of relative distance d by a factor determined by the distance of the head and joints Torso (d_{HD}) is shown in Fig. 5. This factor shows the size of the users and all distances D that can be normalized in accordance with this value.

The set of distances $D = \{d_{EL}, d_{ER}, d_{HL}, d_{HR}, d_{WL}, d_{WR}\}$, the distance D_{norm} normalization is obtained as follows Eq. (2).

$$\sum_{i=1}^n D_{norm}(i) = \frac{D(i)}{d_{HD}} \quad (2)$$

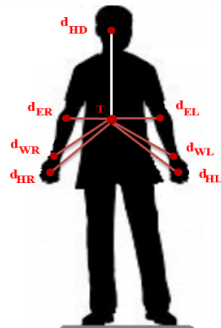


Figure 5: Set Of Distances D Sizes

where n is the number distance D , and d_{HD} is the distance Head-Torso (as in Fig. 5. - line of white). The angles θ and ϕ do not need to be normalized for expressing direction and the direction remains the same after normalization.

3.4 The recognition rate

We evaluate the performance of systems based upon the ability to classify the samples to the correct class. We use this metric to resolve the so-called level of recognition. Recognition rate is defined as the ratio number of samples classified with the correct number of samples, as follows Eq. (3).

$$\text{Recognition rate} = \frac{\text{Number of correctly Classified Samp}}{\text{Total number of samples}} \quad (3)$$

4. EXPERIMENTS AND RESULTS

All experiments were performed by Computer specs: Intel (R) Core (TM) i3-3240 CPU @ 3.4 GHz and 4.00 GB of memory, the operating system Windows 8. The software used MATLAB 2013a.

The feature extraction, starting with making the relative distance from the data of X, Y, and Z to the Spine and head. The process of segmentation of the frame for every sign we use statistical functions mean. To determine the best number of groups we did experiment with the number of groups 12, 13 and 14. The results from the experiment where the best group number is 14. The experiment used the number of groups 14.

The network parameters such as learning rate (μ), the momentum constant (α) and the number of epochs determined by experimentation. The Neural Network is trained and tested with different kinds of learning rate, momentum coefficients and hidden layer, as shown in Table 2.

Table 2: The parameter of Neural Network

Parameter	Value
Input layer	: 672 cells
Hidden layer	: 25, 50, 75, 100 --- 500
Output layer	: 15 class
Learning rate	: 0.04; 0.05; 0.06; 0.07
Momentum coefficients	: 0.06; 0.07; 0.08

The evaluation using the Spherical (X, Y, and Z) of the eighth joints used, Consistency-based Subset Evaluation algorithms and Artificial Neural Network classifier. The experimental results is the best of the variation of nodes in hidden layer 375 and the best average recognition rate 93.67%. Table 3 show result experiments of Spherical (X, Y, and Z) and using Artificial Neural Network classifier with the Spherical (X, Y, and Z) and CSE and using Artificial Neural Network classifier.

Table 3: The Comparison Between Spherical +ANN And Spherical+ CSE +ANN

Method	Accuracy (%)	Epoch	Time (Sec)	Number Of Layer	Node Hidden	Learning Rate (LR)	Momentum Coefficients (MC)
Spherical +CFE +ANN	93.67	691	8.706	275		0.07	0.08
Spherical + ANN[10]	80.54	238	11.034	150		0.06	0.07

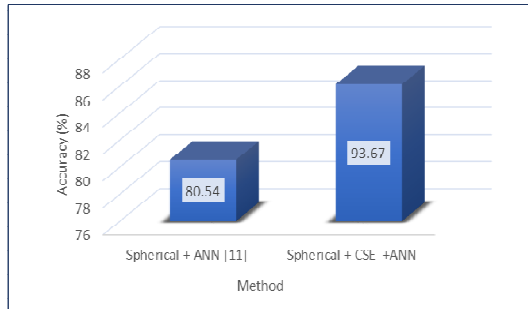


Figure 6: Comparative Diagram Of Result Experiment

5. CONCLUSION

In general has shown that Kinect and depth of the camera are suitable for the recognition of sign language. The level of accuracy obtained in this study can be caused by several factors such as the data capturing process, the position of the kinect and speed of participants in demonstrating sign language. Therefore for future development some of the factors which can lead to a better accuracy rate can be considered. Experiments indicated that our system was able to recognize 15 dynamic Malaysian Sign Language with 93.67% accuracy (as show in Fig. 6). The results of this study show improvement on previous research, conducted by [10]. The use of feature selection can contribute towards the improvement of the accuracy rate of sign language recognition data. The combination of Consistency with ANN can contribute to improve accuracy. It could be developed by combining the Consistency with other classification methods like Artificial Neural Network (ANN), Hidden Markov Model (HMM, Dynamic Time Wrapping (DTW), Support Vector Machine (SVM). We will develop this research in the future using different methods of feature selection such as Correlation-based Feature Subset Selection Evaluation and Correlation-based Attributes of Feature Selection.

REFERENCES:

- [1] P. Mekala, Y. Gao, J. Fan, A. Davari, and C. Eng, "Real-time Sign Language Recognition based on Neural Network Architecture," Institute of Electrical and Electronics Engineers, pp. 197–201, 2011.
- [2] Y. Quan and P. Jinye, "Application of improved sign language recognition and synthesis technology in IB," in *IEEE Conference on Industrial Electronics and Applications*, 2008, pp. 1629–1634, IEEE.
- [3] I. N. Sandjaja and N. Marcos, "Sign Language Number Recognition," Fifth International Joint Conference on INC, IMS and IDC, pp. 1503–1508, 2009.
- [4] F. Wong, G. Sainarayanan, W. M. Abdullah, A. Chekima, F. E. Jupirin, and Y. F. A. Gaus, "Software-based Malaysian sign language recognition," *Advances in Intelligent Systems and Computing*, vol. 182 AISC, pp. 297–306, 2013.
- [5] Y. F. A. Gaus, F. Wong, R. Chin, R. R. Porle, and A. Chekima, "Analysis and Performance Comparison of the Feature Vectors in Recognition of Malaysian Sign Language," *International Journal of Computer and Electrical Engineering*, vol. 5, no. 1, pp. 65–68, 2013.
- [6] T. T. Swee, S. Salleh, A. K. Ariff, C. Ting, and S. K. Seng, "Malay Sign Language Gesture Recognition System," in *International Conference on Intelligent and Advanced Systems*, 2007, pp. 982–985, IEEE.
- [7] S. Auephanwiriyakul, S. Phitakwinai, W. Suttapak, P. Chanda, and N. Theera-Umpon, "Thai sign language translation using scale invariant feature transform and hidden markov models," *Pattern Recognition Letters*, vol. 34, no. 11, pp. 1291–1298, 2013.

- [8] Y. F. A. Gaus and F. Wong, "Hidden Markov Model-Based Gesture Recognition with Overlapping Hand-Head/Hand-Hand Estimated Using Kalman Filter," 2012 Third International Conference on Intelligent Systems Modelling and Simulation, pp. 262–267, 2012.
- [9] A. Sakhalkar, D. Sanghavi, K. Ajmera, M. Joshi, S. Patil, and Y. Rawani, "Gesture Recognition Of American Sign Language : A Neural Network Based Approach," International Journal Of Science, Engineering And Technology, vol. 2, no. 4, pp. 76–83, 2014.
- [10] Sutarman, M. A. Majid, J. M. Zain, and A. Hermawan, "Recognition of Malaysian Sign Language Using Skeleton Data with Neural Network," in *International Conference on Science in Information Technology (ICSITech)*, 2015, pp. 231–236, IEEE Explore.
- [11] J. Xu and J. Yang, "Sign Language Classification Using Search Algorithms," 2006.
- [12] D. Kelly, "Computational Models for the Automatic Learning and Recognition of Irish Sign Language," National University of Ireland, Maynooth, 2010.
- [13] S. Athavale and M. Deshmukh, "Dynamic Hand Gesture Recognition for Human Computer interaction; A Comparative Study," International Journal of Engineering Research and General Science, vol. 2, no. 2, pp. 38–55, 2014.
- [14] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data : A Fast Correlation-Based Filter Solution," in *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, 2003.
- [15] Isabelle Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157–1182, 2003.
- [16] P. Scully and R. Jensen, "Investigating Rough Set Feature Selection for Gene Expression Analysis," the Aberystwyth University, 2011.
- [17] D. E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning." Addison Wesley, Reading, Massachusetts, 1989.
- [18] M. a Hall, "Feature Selection for Discrete and Numeric Class Machine Learning," in *Seventeenth International conference on Machine Learning*, 2000, pp. 1–16.
- [19] M. Hall, H. National, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA Data Mining Software : An Update," SIGKDD Explorations, vol. 11, no. 1, pp. 10–18, 2009.
- [20] M. A. Hall and G. Holmes, "Benchmarking attribute selection techniques for data mining," IEEE Transactions On Knowledge and Data Engineering, vol. 15, no. 6, pp. 1437–1447, 2003.
- [21] M. Dash and H. Liu, "Consistency-based search in feature selection," Artificial Intelligence, vol. 151, no. 1–2, pp. 155–176, 2003.
- [22] K. Kumar and G. S. M. Thakur, "Advanced Applications of Neural Networks and Artificial Intelligence: A Review," International Journal of Information Technology and Computer Science, vol. 4, no. 6, pp. 57–68, 2012.
- [23] R. J. Schalkoff, "Artificial Neural Networks Hardcover." McGraw-Hill Companies, 1997.
- [24] G. Bailador, G. Tröster, and D. Roggen, "Real time gesture recognition using Continuous Time Recurrent Neural Networks," in *International Conference on Body Area Networks.*, 2007, pp. 454–876, ICST.
- [25] R. Akmeliawati, M. P.-L. Ooi, and Ye Chow Kuang, "Real-time Malaysian sign language translation using colour segmentation and neural network," in *Instrumentation and Measurement Technology Conference*, 2007, no. May, pp. 1–6, IEEE.
- [26] M. Maraqa, "Recognition of Arabic Sign Language (ArSL) Using Recurrent Neural Networks," Journal of Intelligent Learning Systems and Applications, vol. 4, no. 1, pp. 41–52, 2012.
- [27] M. P. Paulraj, S. Yaacob, H. Desa, and W. M. R. W. A. Majid, "Gesture recognition system for Kod Tangan Bahasa Melayu (KTBM) using neural network," in *2009 5th International Colloquium on Signal Processing & Its Applications*, 2009, pp. 19–22, IEEE.
- [28] A. Karami, B. Zanj, and A. K. Sarkaleh, "Persian sign language (PSL) recognition using wavelet transform and neural networks," Expert Systems with Applications, vol. 38, no. 3, pp. 2661–2667, Mar. 2011.
- [29] Y. F. Admasu and K. Raimond, "Ethiopian sign language recognition using Artificial Neural Network," in *Intelligent Systems Design and Applications (ISDA)*, 2010, pp. 995–1000, IEEE.

- [30] M. M. Hasan and P. K. Misra, “Brightness Factor Matching For Gesture Recognition System Using Scaled Normalization,” *International Journal of Computer Science & Information Technology (IJCSIT)*, vol. 3, no. 2, pp. 35–46, 2011.
- [31] M. Corporation, “Kinect for Windows - Human Interface Guidelines v1.7.” Microsoft Corporation, 2013.
- [32] M. Hazewinkel, “Spherical coordinates,” *Encyclopedia of Mathematics*. Springer, 2001.
- [33] E. W. Weisstein, “Spherical Coordinates,” *MathWorld - A Wolfram Web Resource*, 2015. [Online]. Available: <http://mathworld.wolfram.com/SphericalCoordinates.html>. [Accessed: 23-May-2015].
- [34] D. M. Capilla, “Sign Language Translator using Microsoft Kinect XBOX 360TM,” University of Tennessee, Knoxville - USA, 2012.