International Journal of Scientific Research & Engineering Trends
Volume 5, Issue 3, May-Jun-2019, ISSN (Online): 2395-566X

Sign Language Translator Using CNN Model

Diniya A J Arathi Remesh Dona Jose Jasmi Davis
Dept. of Computer Science and Engineering
Sahrdaya College of Engineering and Technology
Kodakara, Thrissur, Kerala, India

***Abstract*-As the web has developed it has become a place where people interact. They post opinions, modify and enhance each other's contributions and share information.The problem we are investigating is sign language recognition through supervised feature learning. With the technological trend in man-machine interfaces and the machine intelligence, exploiting these powers has become a challenge in many fields. In particular, it was observed that the body gesture-based interactions of human to human and human to machine are rapidly increasing, especially in the area of sign language interpretation. Statistics strongly suggest that the population of deaf and mute people is on the rise and there is a need to train more people the American Sign Language (ASL) to bridge the gap. Furthermore, the electronic devices such as TVs, PCs, PDAs Robots, cameras, etc. are advanced and built toread users gestures and respond to their commands. Therefore, it is of a great interest to try to conduct research in this area and propose efficient and effective solutions for man-machine gesture-based interaction. Here we propose a system for the automatic recognition of American Sign Language gestures. Aseries of image and vector processing operations are used in order to transform a visual hand gesture into a spoken letter, display the text and convert the received voice back into text providing a two-way communication.**

Keywords-CNN, ASL

I. INTRODUCTION

We the humans use language as well as gestures for communicating with each other. They are not only used in order to convey its meaning, but are also to explain about what cannot be said by just speaking. Research in automatic recognition of human communication has predominantly been focused on speech recognition and handwriting, while research in the automatic recognition of gestures has lagged behind. However, there has been a recent surge in the interest of the automatic recognition of human gestures. Some of the more structured forms of gestures are those that are performed in sign language.

Sign language is a specific area of human gesture communication and a full-fledged complex language that is used by various Deaf communities around the world. Unfortunately, there is a tremendous lack of non-Deaf people who have an in depth knowledge of sign language, which leads to the social isolation of the deaf community. This has brought forth motivation for the development of a computational system capable of automatically interpreting sign language.

Since sign language is a structured form of gesture communication, the development of such a system would also be beneficial in human-computer interaction, virtual reality, and robotics. Sign language plays an important role to communicate with themselves as well as with normal people in a non- verbal manner for dumb and deaf people. Gestures are the referred to as the initial method to pass the messages that are commonly conducted in a 3-D space called a sign space, along a combination of manual as well as non manual signal. Manual signal is usually used to make hand movements or finger posturing on the other hand non manual signal is used for an outside appearance like facial movements, movements that comes in mouth and body positions . Sign language is not standardized universally yet. Every country have invented their own SL, like the ASL while Germany has developed its Sign Language that is GSL. Each of the countries' sign languages differs within various regions in its country. So that, it can often be a challenge in order to find a global sign language interpretation system for using it universally.

II. PROBLEM DEFINITION

As the deaf and dumb community is usually deprived of communication using the normal language with other people, obviously they have to depend completely on the visual media communication like video calls. Most of us, the normal people don't have enough knowledge about the sign language, which gradually makes them isolated socially of the dumb community. This device can compensate the complete dependency of deaf people on the visual media communication like video calls. Even the sounds made by dumb people can be misinterpreted because their sign language is not completely interpreted by the normal people. This issue is considered and solved by our proposed system that converts the standard set of signs into the corresponding text or speech in the form of voice output. This way, we are able to negotiate the gap of communication between the two different people. Various aspects of our project will be looked in a critical viewpoint in order to analyze even the strengths and weaknesses in terms of feasibility. People who have hearing and speaking disabilities could be able to use the sign language as a mode of communication within themselves as well as with remaining world, but fortunately every one of the us knew the sign language and therefore, the end of the result is lacking from the communication as well as the isolation.

As the surveys conducted by World Health Organization says that almost three hundered and sixty millions in the world are suffering from hearing diabilities and almost one lakh twenty thousand people are born with the speaking disability this year. As the estimates says, it almost covers around five and half percentage among the worlds population and around ninety percentage of them are eighteen plus. For providing help to the people with such disabilities, lots and lots of researches have been held every year and also so many solutions have been obtained worldwide till now but no complete success is been reported though.

People suffering from hearing as well as speaking disability have the sign language as the only way for the communication. Thus, it has became a major problem for the people who does not have the familiarity with the various gestures in their sign language. So, therefore it creates a barrier between the disabled and not for communication. Sign language does not utilize naturally made voices instead, it uses various sign patterns that are visually different from each other. By combining the different hand shapes, its movements also the facial expressions continuously, it has most possible information and also it describes the user's thoughts as well.

III.MOTIVATION

The deaf community is a very unique culture or group of humans who used to constantly find themselves battling the misinterpretation that being deaf or hearing impaired is a defect that must be corrected either by medical procedures or by technological or a medical equipment. The definition for the term, ”culture” defines that it is the development of an intellect by required training or enough education and also, we can see that the deaf culture has grown into just and only this. This is what that motivated us to develop a sign language interpreter that can help them grow and achieve their goals

IV. OBJECTIVES

One among the biggest challenges of sign language recognition is to find a technique that is efficient enough in order to grasp the language. There is approx around six thousand set of signs in ASL and many can be able to appear in various forms depending upon the subject, and object also. In most of the hand gesture interpretation systems, the complete gesture is modeled, and in sign language detection, where the language is somewhat large in size, it could not at all be feasible to model each one of the sign in sign language. It will be more feasible if it decreases the number of signs into a limited set of portion that is molded in order to create the complete set of signs in American Sign Language. These methods will enable each one of the sign to be combined uniquely like it was done in speech detection.

V. SYSTEM ARCHITECTURE

1. Existing System

1.1 Smart Hand Wearable Device - The hand wear- able holder[1] shown in the Fig. 1 has been manufactured. using some filaments having a great elasticity and flexibility. These flexible conducting wires enables the different joints and hinges that makes the device suitable to fit any hand size, small or big. 5 holders that fits each of our finger were also made by the same elastic filament that is been fixed at the very first joint to hold the sensors in each of the fingers.

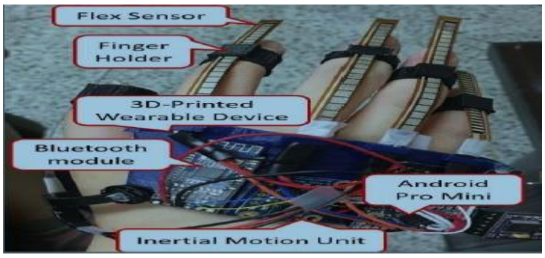

Fig. 1 3D printed wearable device.

Likewise those holders can fit any finger shapes or sizes of different people. Also a neutral state was also added along with 26 alphabets which has a gesture where all the fingers are opened broadly. Data are added in to the data set during the time of experimenting with those data using a desktop. Then it is saved in a text file for making it accessible. Desktop is not described in detail as it is only used as recording equipment.

The sensors that are used here are either 4.5 inch[31] or 2.2 inch[32] flex sensors. The smallest sensor for the pinky finger, while the largest for the remaining fingers. The flex sensor used for the thumb has a larger distance to the controller board in comparison with rest of the fingers so, a longest sensor is for the smallest.

1.2 Techno talk Sign Translator- It is an electric glove used by a deaf/dumb person in order to communicate normally with the normal people who are not able to under- stand the sign language. It gives the text as well as voice output on an LCD display or through a speaker. It includes 5 flex sensors that detects the fluctuations among the signs in a sign language.ASL is processed for conversion and interpretation can be taken using a camera or a sensor. Here,the flex sensor is made to lie on the hand glove that can be used to take the SL as in the diagram shown in the figure, Fig.2. Each sensors are fitted in every finger. It has the goal to translate ASL gestures. The purpose was to help users to communicate with people through a medium.

2. Proposed System

2.1Sign Language- Sign languages are considered a ”manual language” as opposed to a spoken language using gestures and hand formations to convey thoughts. Like spoken languages, sign language uses its own grammar and rules and is not simply a translation of English into gestures. American Sign Language (ASL) is a complete, complex language that employs signs made by moving the hands combined with facial expressions and postures of the body. It is the primary language of many who are deaf and is one of several communication options used by people who are deaf or hard-of-hearing.

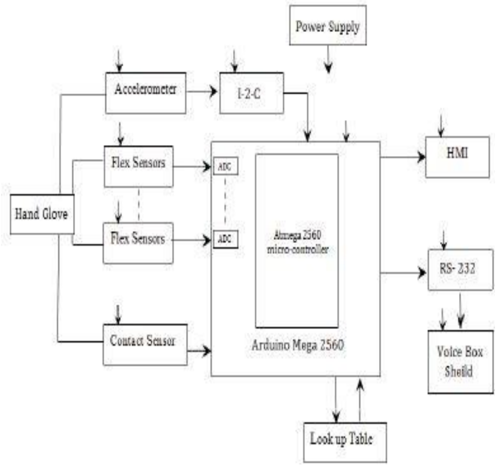

Fig. 2 Techno-Talk system.

The deaf community is a unique culture who constantly findsthemselves battling the misconception that being deaf or hearing impaired is a defect that must be corrected whether by medical procedures or technological devices. The definition of ”culture” states it is the development of the intellect through training or education, and we can see the deaf culture has grown into just this. Development of sign language as the first language of those who are deaf, deaf schools, and the continual improvement in accessibility into main stream culture indicates the deaf culture is strong and growing rapidly.

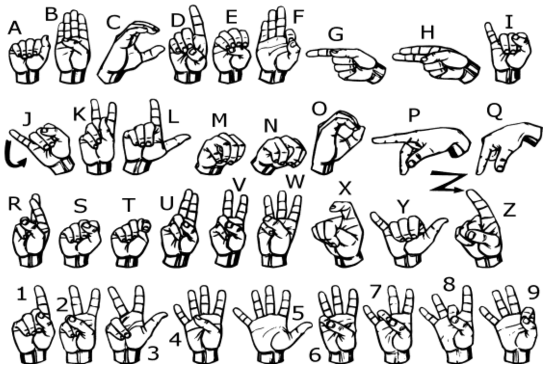

Fig. 3 American Sign Language.

Despite this strength in the deaf culture, American SignLanguages use amongst the hearing abled is limited. Communication between the hearing abled and hearing impaired or deaf commonly creates a barrier when one of the two parties is unable to communicate via sign language. Over- coming this barrier from the viewpoint of those who are unable to currently sign ASL, forms the basis of our project, a Vision-Based Hand Shape Identification implemented for developing the ASL Recognition system.

2.2 An Overview of SLR- One of the biggest challenges in sign language recognition (SLR) is to find a modeling technique that is powerful enough to capture the language, yet it can scale to large vocabulary size. There are approximately 6000 cataloged signs in American Sign Language, and many signs can appear in many different forms depending on subject, object, and numeric agreement. With the different forms in which the same sign can appear, the number of possible cases to consider increases to a degree much larger than 6,000.

In many hand gesture recognition systems the entire gesture is modeled, however, in sign language recognition, where the language is very large, it is not feasible to model each sign. It is more feasible to break down signs into a limited set of primitive parts, phonemes that can be combined to make up the entire set of signs in ASL. This procedure enables each of the phonemes to be modeled separately as is done in speech recognition.

Since the early 1960s research has been done on the linguistics of American Sign Language. Fortunately, one of the most significant findings of the properties of ASL, first proposed by William Stokoe, was that American Sign Language could be broken down into phonemes. Stokoe defined three types of phonemes: location where on the body the sign takes place, hand shape how the fingers are articulated, and movement how the hands move.

2.3Model Description-We created our system asmodules. Each and every module had it's own specific task in hand such as:

- Adding a gesture
- Loading all the testing and image labels
- Setting the hand histogram
- Displaying the current gestures
- CNN model initialization and training
- Recognizing and converting gesture to speech

2.4Gesture Recognition- In recent years, there has been a tremendous amount of research on hand gesture recognition. Some of the earlier gesture recognition systems attempted to identify gestures using glove-based devices that would measure the position and joint angles of the hand. However, these devices are very cumbersome and usually have many cables connected to a computer. This has brought forth the motivation of using non intrusive, vision-based approaches for recognizing gestures. Vision- based approaches involve using one or more video cameras to capture a person gesturing and using computer vision techniques to interpret each particular gesture.

2.5 Preprocessing Image-Operations at the lowest level of abstraction with images have the common name of preprocessing, both input and output are intensity images.Improvement of the image data that suppresses unwilling distortions or enhances some image features important for further process are the aim of preprocessing ,pre-processing methods here since similar techniques are used to classifying the geometric transformation of images(rotation, scaling, translation).

Image pre-processing is one of the most important steps in gesture recognition. By pre-processing, we can rectify some minor errors, important features can be extracted easily. Basically, a pre-processed image is more easy to train and more accurate. If we pass an image without pre-processing, it gets generalized. The steps are :

- Flip the image.
- Convert from RGB to HSV.
- Calculate Back Projection based on the stored Histogram
- Filter using an Elliptical Structuring Element.
- Gaussian and Median Blur.
- Threshold using a combination of Binary and Otsu method.
- Merge threshold 3 times.
- Convert to Grayscale.
- Find the contours.

2.6 Training Model- Convolution is a process where the network tries to label the input signal by referring to what it has learned in the past. The first layers that receive an input signal are called convolution filters.

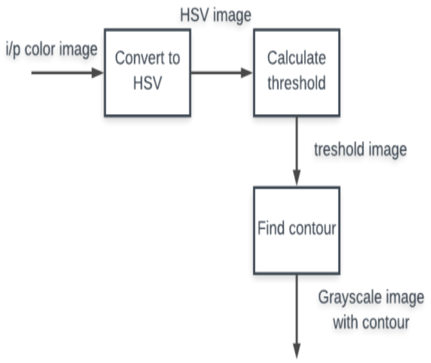

Fig. 4 Preprocessing.

Convolution is a process where the network tries to label the input signal by referring to what it has learned in the past. If the input signal looks like previous images it has seen before, the label which it had seen will be mixed to, or convolved with the input signal. The resulting output signal is then passed on to the next layer. Convolution has the property of being translational invariant. This means that each convolution filter represents a feature of interest and the CNN algorithm learns which features comprise the resulting reference. Thus, even if a blood clot appears in a different orientation, the CNN algorithm would still be able to recognize it.

Sub sampling is the inputs from the convolution layer can be smoothened to reduce the sensitivity of the filters to noise and variations. This is called subsampling and can be achieved by taking averages or taking the maximum over a sample of the signal. Examples of sub- sampling methods (for image signals) include reducing the size of the image, or reducing the color contrast across red, green, blue (RGB) channels.Activation is the activation layer controls how the signal ows from one layer to the next, emulating how neurons are fired in our brain. Output signals which are strongly associated with past references would activate more neurons, enabling signals to be propagated more efficiently for identification.

2.7 Convolutional Neural Network(Cnn)- Convolution neural networks (ConvNets or CNNs) are widely used tools for deep learning. They are specifically suitable for images as inputs, although they are also used for other applications such as text, signals, and other con- tinuous responses. They differ from other types of neural networks in a few ways: Convolution neural networks are inspired from the biological structure of a visual cortex, which contains arrangements of simple and complex cells. These cells are found to activate based on the subregions of a visual field. These sub-regions are called receptive fields. Inspired from the findings of this study, the neurons in a convolution layer connect to the sub-regions of the layers before that layer instead of being fully-connected as in other types of neural networks. The neurons are unresponsive to the areas outside of these sub-regions in the image. These sub-regions might overlap, hence the neurons of a ConvNet produce spatially-correlated outcomes, whereas in other types of neural networks, the neurons do not share any connections and produce independent outcomes. In addition, in a neural network with fully-connected neurons, the number of parameters (weights) can increase quickly as the size of the input increases. A convolutional neural network reduces the number of parameters with the reduced number of connections, shared weights, and down sampling.

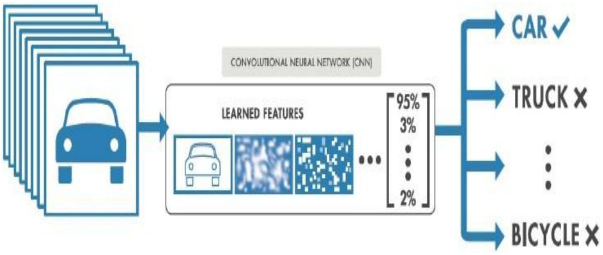

Fig. 5 Convolutional Neural Networks[43]

A ConvNet consists of multiple layers, such as convolutional layers, max- pooling or average-pooling layers, and fully-connected layers. Convolutional neural networks (or convnets for short) are used in situations where data can be expressed as a map wherein the proximity between two data points indicates how related they are. An image is such a map, which is why you so often hear of convnets in the context of image analysis. If you take an image and randomlyrearrange all of its pixels, it is no longer recognizable. The relative position of the pixels to one another, that is, the order, is significant (Fig. 7).
CNNs contains three components:

- Convolutional layers, which apply a specified number of convolution filters to the image. For each subregion, the layer performs a set of mathematical operations to produce a single value in the output feature map. Convolutional layers then typically apply a ReLU activation function to the output to introduce non-linearities into the model.
- Pooling layers, which down sample the image data extracted by the convolutional layers to reduce the dimensionality of the feature map in order to decrease processing time. A commonly used pooling algorithm is max pooling, which extracts sub-regions of the feature map (e.g., 2x2-pixel tiles), keeps their maximum value, and discards all other values.
- Dense (fully connected) layers, which perform classification on the features extracted by the convolutional layers and down sampled by the pooling layers. In a dense layer, every node in the layer is connected to every node in the preceding layer.

Typically, a CNN is composed of a stack of convolutional modules that perform feature extraction. Each module con- sists of a convolutional layer followed by a pooling layer. The last convolutional module is followed by one or more dense layers that perform classification. The final dense layer in a CNN contains a single node for each target class in the model (all the possible classes the model may predict), with a softmax activation function to generate a value between 01 for each node (the sum of all these softmax values is equal to 1). It can be interpreted as the softmax values for a given image as relative measurements of how likely it is that the image falls into each target class.

2.8 Identification Of Letter-Prediction refers to the output of an algorithm after it has been trained on a historical dataset and applied to new data when youre trying to forecast the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be. In some cases, it really does mean that you are predicting a future outcome, such as when youre using machine learning to determine the next best action in a marketing campaign.

For example, you may want to predict which of five (or even more) marketing channels will have the highest return on investment based on historical customer behavior, so thatyou can optimize your marketing budget by focusing on the most effective channels.On input of an gesture image, preprocessing is done on the image as described below Fig. 8. It is then converted to an array and fed into the trained model. This will now give the output of the predicted class and its accuracy score.

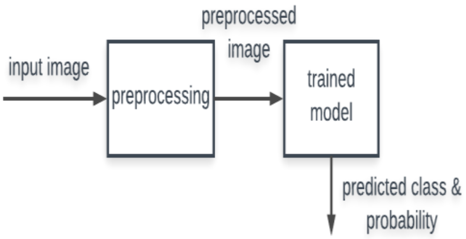

Fig. 6 Classifications.

2.9 Text To Speech- A Text-To-Speech (TTS) synthesizer is a computer based system that should be able to read any text aloud, when it is directly introduced in the computer by an operator. It is more suitable to define

Text-To-Speech or speech synthesis as an automatic production of speech.

VI. RESULTS & DISCUSSION

Our system develops an interfaces that enables deaf and dumb community to have a normal way of communicating with others. Sign language which is the most natural way of communicating with each other among the deaf and dumb. Sign language includes a set of signs as same as that of the spoken language words.It is now possible for us to add our own gestures in to the asl gestures in various lighting conditions, which can make another variety of inputs under different shapes, sizes as well as the colours of different hands, even in dark or medium or even bright lighting.

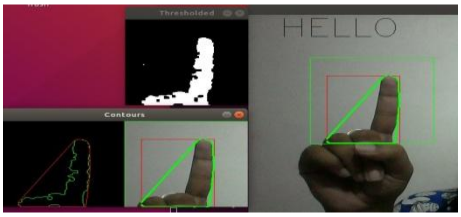

Fig.7 Hand Recognition.

VII. CONCLUSION

A machine learning domain based sign language gesture interpretation system has been developed that has the capability to identify a total of forty four hand gesturesincluding the letters from A to Z and the numbers from 1 to 10 and also some commonly used words/phrases like Love,Like, Remember, Best of Luck etc. A series of images are used in order to convert each of the hand gesture into text as well as voice output and display long paragraphs and convert the received voice back into text providing a two-way communication. The system was trained using a data set of around one thousand two hundered gestures in the form of images which was split into training images as well as test the images. The preprocessing was done using OpenCV and the hand histogram was captured and serialized using pickle python library. Our system has an accuracy of ninety five percentage in predicting the output.

VIII. FUTURE WORK

- Inclusion of all possible two-hand gestures could be a feature that can be added in the future version of the project.
- Integration to long paragraphs of communication can be developed on a later stage.
- Including automatic hand recognition without the setup of a Histogram region can be a mechanism to be thought of on the later versions.

REFERENCES

[1] B. G. Lee, and S. M. Lee, (2017), "Smart Wearable Hand Device for Sign Language Interpretation System with Sensors Fusion", IEEE Sensors.

[2] J. Wang and T. Zhang,(2014) "An ARM-based embedded gesture recognition system using a data glove", presented at the 26th Chinese Control and Decision Conf., Changsa, China.

[3] A. Z. Shukor, M. F. Miskon, M. H. Jamaluddin, F. A. Ibrahim,M. F. Asyraf and M. B. Bahar, (2015) "A new data glove approach for Malaysian sign language detection", Procedia Comput. Sci., vol. 76, no. 1, pp. 60-67.

[4] N. Sriram and M. Nithiyanandham, (2013) "A hand gesture recog- nition based communication system for silent speakers", presented at the Int. Conf. Human Comput. Interact..

[5] S. V. Matiwade and M. R. Dixit,(2016) "Electronic support system for deaf and dumb to interpret sign language of communication", Int.J. Innov. Research Sci. Eng. Technol., vol. 5, no. 5, pp. 8683-8689.

[6] S. Goyal, I. Sharma and S. Sharma, (2013), "Sign language recognition system for deaf and dumb people", Int. J. Eng. R. Technol., vol. 2, no. 4, pp. 382-387.

[7] S. P. More and A. Sattar,,(2016), "Hand gesture recognition system using image processing", presented at Int. C onf. Electrical, Electronics, Opt. Techniq..

[8] K. Murakami and H. Taguchi,(1991), "Gesture recognition using recurrent neural network", in Proc. SIGCHI Conf. Human Factors Comput. Syst..

[9] P. R. V. Chowdary, M. N. Babu, T. V. Subbareddy, B. M. Reddy and V. Elamaran,(2015) "Image processing algorithms for gesture recognition using matlab", presented at Int. Conf. Adv. Comm. Control Comput. Technol..

[10] T. Khan and A. H. Pathan,(2015), "Hand gesture recognition based on digital image processing using matlab", Int. J. Sci. Eng. R., vol. 6, no. 9, pp. 338-346.

[11] J. Siby, H. Kader and J. Jose,(2015), "Hand gesture recognition",Int. J. Innov. Technol. R., vol. 3, no. 2, pp. 1946-1949.

[12] J. L. Lamberti and F. Camastra,(2011), "Real-time hand gesture recognition using a color glove ", presented at Int. Conf. Image Analy. Process., Ravenna, Italy .

[13] Y. Iwai, K. Watanabe, Y. Yagi and M. Yachida,(1996), "Gesture recognition using colored gloves", Proc. 13th Int. Conf. Pattern Recog., Vienna, Austria.

[14] C. Preetham, G. Ramakrishnan, S. Kumar and A. Tamse,(2013), "Hand talk- implementation of a gesture recognizing glove", pre- sented at 2013 Texas Instruments India Educators Conf., Bangalore, India.

[15] K Patil, G. Pendharkar and G. N. Gaikwad,(2014), “American sign language detection”, , Int. J. Sci. R. Pub., vol. 4, no. 11, pp. 1-6.

[16] J. Kim, N. D. Thang and T. Kim,(2009), “3-D hand motion tracking and gesture recognition using a data glove”, presented at IEEE Int. Symp. Industrial Elec. 2009, Seoul, South Korea.

[17] D. Lu, Y. Yu and H. Liu,(2016), “Gesture recognition using data glove: an extreme learning machine method”, in Proc. 2016 IEEE Int. Conf. Robotics and Biomimetics, Qingdao, China.

[18] J. Lm, D. Lee, B. Kim, I. Cho and J. Ryou,(2010), “Recognizing hand gestures using wrist shapes”, presented at 2010 Digest Technical Papers Int. Conf. Consumer Elec., Las Vegas, USA.

[19] R. Xie, X. Sun, X. Xia and J. Cao,(2015), “Similarity matching- based extensible hand gesture recognition”, IEEE Sensors J., vol. 15, no. 6, pp. 3475-3483.

[20] Y. L. Hsu, C. L. Chu, Y. J. Tsai and J. S. Wang,(2015), “An inertial pen with dynamic time warping recognizer for handwriting and gesture recognition”, IEEE Sensors J., vol. 15, no. 1, pp. 154-163.

[21] L. Yin, M. Dong, Y. Duan, W. Deng, K. Zhao and J. Guo,(2014), “A high-performance training-free approach for hand gesture recog- nition with accelerometer”, Mult. Tools App., vol. 72, no. 1, pp. 843- 864.

[22] J. Galka, M. Masior, M. Zaborski, K. Barczewska,(2016), “Inertial motion sensing glove for sign language gesture acquisition and recognition”, Sensors J., vol. 16, no. 16, pp. 6310-6316.

[23] X. Cai, T. Guo, X. Wu and H. Sun,(2015), “Gesture recognition method based on wireless data glove with sensors”, Sensor Letters, vol. 13, no. 2, pp. 134-137.

[24] K. Liu, C. Chen, R. Jafari and N. Kehtarnavaz,(2014), “Fusion of inertial and depth sensor data for robust hand gesture recognition”, IEEE Sensors J., vol. 14, no. 6, pp. 1898-1903.

[25] K. W. Kim, M. S. Lee, B. R. Soon, M. H. Ryu and J. N. Kim,(2016), “Recognition of sign language with an inertial sensor- based data glove, Technol. Health Care”, vol. 24, no. 1, pp. 223-230.

[26] L. Sousa, J. M. F. Rodrigues, J. Monteiro, P. J. S. Cardoso and R. Lam,(2016), “GyGSLA: a portable glove system for learning sign language alphabet ”, presented at Int. Conf. Universal Access in Human-Comput. Interact., Toronto, ON, Canada.

[27] N. Caporusso, L. Biasi, G. Cinquepalmi, G. F. Trotta and A. Brunetti,(2017), “A wearable device supporting multiple touch- and gesture-based languages for the deaf-blind ”, presented at Int. Conf. Appl. Human Factors and Ergonomics, LA, CA, USA.

[28] Z. Yu, X. Chen, Q. Li, X. Zhang and P. Zhou,(2014), “”, A hand gesture recognition framework and wearable gesture-based interaction prototype for mobile devices.

[29] J. Wu, Z. Tian, L. Sun, L. Estevez and R. Jafari,(2015), “Real-time American sign language recognition using wrist-worn motion and surface emg sensors”, presented at IEEE 12th Int. Conf. Wearable Implantable Body Sensor Net., Cambridge, USA.

[30] J. Wu, L. Sun and R. Jafari,(2016), “A wearable system for recognizing American sign language in real-time using imu and surface emg sensors”, IEEE J. Biomedic. Health Info., vol. 20, no. 5, pp. 1281-1290.