ASL RECOGNITION USING LEAP MOTION AND HAND TRACKING MECHANISM

RANIA AHMED KADRY ABDELGAWAD BIRRY

College of Engineering and Technology, Computer Engineering Department
Arab Academy for Science, Technology and Maritime Transport Alexandria, Egypt
E-mail: rania_kadry2012@yahoo.com

Abstract- Sign Language is a widely used method of communication among the community of deaf mute people. It contains some series of body gestures, which enables a person to interact without the need of spoken words. Although the use of sign language is very popular among the deaf mute people but the other communities don't even try to learn it, this creates a gulf of communication and hence becomes a cause of the isolation of physically impaired people. A system is required to facilitate a way of communication between these two communities. This paper will demonstrate a method for recognition of American Sign Language (ASL) gestures using Leap Motion Camera Controller. Features extraction techniques are obtained from hand by using Leap Motion Camera in order to be entered to Artificial Neural Networks (ANN) classifier to develop a model to recognize hand gestures for both static and dynamic signs. Along with that, a robotic arm is built with Arduino Uno Micro-Controller and two servos motors to track the hand to keep it in the small viewing domain of the Leap Motion while performing gestures. The dataset obtained is 3600 images for ASL from which 3400 images were taken for 24 static letters and 10 numbers, in addition to 2 images for dynamic letters (Z, J). In this way, ten images were taken for every letter and number from 10 different signers to give the total of 3600 dataset images. Recognition of ASL images were applied by obtaining feature extraction to be entered as inputs to ANN classifier to yield an overall average success rate of 84.66% and a weighted average success rate of 83.11% accuracy in the recognition of ASL using Leap Motion.

Keywords- leap Motion, ASL, Image processing, feature extraction, DCT, ANN,Arduino Uno Micro-Controller, Servo motors, Robotic Arm.

I. INTRODUCTION

American Sign Language (ASL) is a visual/gestural language. It is a natural language, meaning that it has developed naturally over time by its users, Deaf people. ASL has all of the features of any language; that is, it is a rule governed system using symbols to represent meaning [1]-[2].One of the earliest written records of a sign language is from the fifth century BC, in Plato's Cratylus, where Socrates says: "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?".

The main problem nowadays faced by deaf people is to communicate with people who do not know sign language. While writing is an option, it is considered as a slow and inefficient way of communication. Therefore a viable option would be to hire a professional sign language translator. However, it would of course be much better and more viable to consider relying on technology using some kind of gesture translation software. In ASL, the symbols are specific hand movements and configure rations that are modified by facial expressions to convey meaning. These gestures or symbols are called signs. Contrary to common belief, ASL is not derived from any spoken language, nor is it a visual code representing English. It is a unique and distinct language, one that does not depend on speech or sound. ASL has its own grammar, sentence construction, idiomatic usage, slang, style, and regional variations-the characteristics that define any language [3]. In the United States there are several sign systems that should not be confused with American Sign Language (ASL). These systems are ways of putting the English language into a manual-visual form. Thus, they are called systems of Manually Coded English (MCEs) [4]. They are designed primarily for the purpose of teaching English to deaf children. This paper is concerned solely with ASL, which is the sign language most deaf people use when they are communicating among themselves [5] -[6]. The current phase of this research is focused on the development of an image-processing approach to the recognition of ASL by using Leap Motion which is a device that was released in 2013 to facilitate computer supported hand recognition. . Moreover, the data gathered by the device are relatively accurate and can be used in several classification methods [7].

Techniques are readily available in the form of image compression and feature extraction to be used for recognition by using ANN classifier. The system presented in this paper is born out of this necessity and its main focus is the graphical translation of American Sign Language in a fast, reliable and robust way to make communication easier than ever between different members of the society. This paper discusses the algorithms and techniques used to recognize ASL into English letters. It starts by introducing the proposed system and the Leap Motion. This is followed by the discussion of the differentfeature extraction (static and dynamic) gestures techniques used for ANN classification. In addition, an implementation of an actuating mechanism that consists of Arduino Uno Micro-Controller, two servos motors and robotic arm to

track the hand to keep it in the domain of the Leap Motion while performing gestures. Finally results and discussion are presented and supplemented with conclusions and proposals for future work.

II. REVIEW OF RELATED WORK

Sun, Zhang and Xu used Kinect to recognize signs supported by image processing. The method had an accuracy mean of 84.1 % using selected signs and gestures [8].Alexander, Tzipora and Nasir work includes classification of static data only gathered by the Leap Motion using trained classifiers [9].Chuan, Regina, and Guardino, used the palm-sized Leap Motion sensorto provide a much more portable and economical solution than Cyblerglove or Microsoft Kinect used in existing studies by apply k-nearest neighbor and support vector machine to classify the 26 letters of the English alphabet in American Sign Language using the derived features from the sensory data. The experiment result shows that the highest average classification rate of 72.78% and 79.83% was achieved by k-nearest neighbor and support vector machine respectively [10].

Marin et.al [Marin et al. 2015] works on hand gestures recognition using Leap Motion Controller and kinect devices. Ad-hoc features are built based on fingertips positions and orientations. These features are then fed into a multi-class SVM classifier to recognize their gestures. Depth features from the Kinect are also combined with features from Leap Motion Controller to improve the recognition performances. They only focus on static gestures rather than dynamic gestures [11]
Hand Motion Understanding system developed by [Cooper et al. 2011] utilize color-coded glove to track hand movement. The tracking system requires users to wear gloves which reduces the user experiences [12].

III. PROPOSED SYSTEM

The proposed system consists of two parts: the first part is the use of the Leap Motion to get the images of the signs from the signer, image processing is then applied to obtain feature extraction to be entered to ANN Classifier for recognition of dynamic and static signs and numbers.

The second part is the Actuating mechanism to keep the hand's palm position in the center of the camera vision by using two servo motor drivers. A summary for the proposed system diagram is shown in the figure 1.

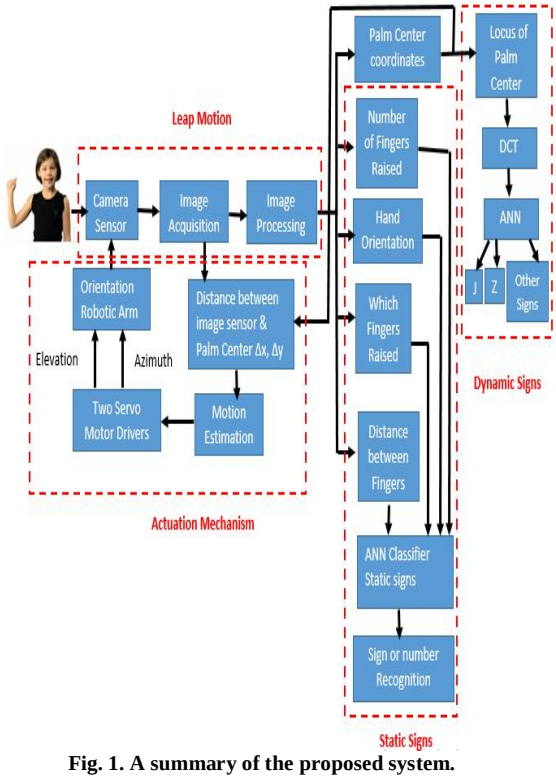


Fig. 1. A summary of the proposed system.

IV. LEAP MOTION

The Leap Motion is a hand tracker released in 2013 where an American company that manufactures and markets a computer hardware sensor device that supports hand and finger motions as input, analogous to a mouse, but requires no hand contact or touching. In 2016, the company released new software designed for hand tracking in virtual reality [13].
Leap motion is placed face up on a surface, the controller senses the area above it in a range of approximately 24 inches on the vertical axis. Leap Motion's precision is about 0.00039 inches, it operates in a close proximity at a rate of 200 frames per second , and it tracks up to 27 joints per arm, including hands, wrist and elbow joints [7]-[14]. Each recognized point is represented by an ordered triple (x, y, z) that specifies a point in 3-dimensional space as shown in figure 2. Leap Motion employs a right-handed Cartesian coordinate system; all axes intersect in the center of the device. The X-axis lies horizontally parallel to the long edge of the device, with a range of 24 inches at a maximum 150° angle. The Z-axis also lies horizontally parallel to the short edge of the device, reducing away from the user's body, i.e., closer to the person the higher the values, with a range of 24 inches at a maximum of 120° angle. The Y-axis is vertical, increasing upwards [15].

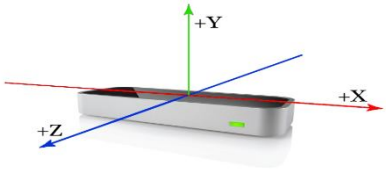


Fig. 2. X, Y, Z coordinates in Leap Motion

Leap Motion uses two high precision infrared cameras and three LEDs. Leap Motion was released with an API that allows access to hand's identifiers and approximate dimensions. It also allows access to joint's normalized direction vectors. These data can be gathered up to 100 times per second. Leap motion's API can be accessed using different programming languages[14].

4.1 Camera Sensors

The Leap Motion controller uses optical sensors and infrared light. The sensors are directed along the y-axis – upward when the controller is in its standard operating position – and have a field of view of about 150 degrees. The effective range of the Leap Motion Controller extends from approximately 25 to 600 millimeters above the device (1 inch to 2 feet). Detection and tracking work best when the controller has a clear, high-contrast view of an object's silhouette. TheLeap Motion software combines its sensor data with an internal model of the human hand to help cope with challenging tracking conditions [16].

4.2 Image Acquisition

Image acquisition in image processing can be defined as the action of retrieving an image from some source, usually a hardware-based source, so it can be passed through whatever processes need to occur afterward[17].

4.3 Image processing

Image-processing techniques such as analysis and detection of shape, texture, color, motion, optical flow, image enhancement, segmentation, and contour modeling, have also been found to be effective [18] . They are mainly targeted towards the detection of static signs, where no gesture is included. The output of image processing may be either an image or a set of characteristics or parameters related to the image [19].

V. FEATURE EXTRACTION

In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features (also named features vector). Transforming the input data into the set of features is called feature extraction. If the features extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm which over fits the training sample and generalizes poorly to new samples. Feature extraction is a general term for methods of constructing combinations of the variables to get around these problems while still describing the data with sufficient accuracy [20].

The following feature extraction techniques are mainly used for dynamic and static gestures:

5.1 Dynamic gesture feature

Which includes signs that are done by movement like "Z", "J", and other sign(good morning, goodbye….etc.). Those signs are implemented by finding:

Palm Centre Coordinates

Palm data consists of the unit direction vector of the palm, the position of the palm center, the velocity of the palm and the accuracy of the data. At the same time, the grab strength, the pinch strength, the sphere center and the sphere radius are obtained. The finger data consists of the direction of each finger, the length of each finger, the tip velocity and the position of joints for distal phalanges, intermediate phalanges, proximal phalanges and metacarpals [21] as shown in the figure 3.

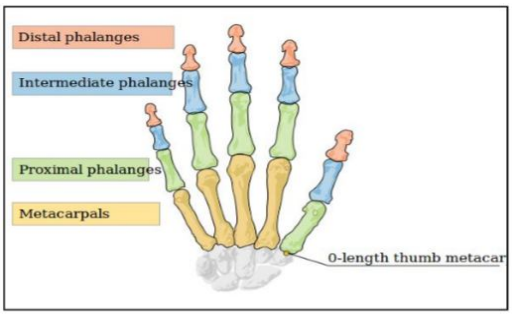


Fig. 3. Hand Bones

The Palm Position of the hand is relatively straightforward; represented in the x, y, and z coordinates. Locus of Palm center is obtained into 2D images to apply DCT for dynamic signs (Z, J, and other signs) recognition.

Discrete cosine transform (DCT)

A discrete cosine transform (DCT) expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies [22]. DCTs are important to numerous applications in science and engineering, from lossy compression of audio (e.g. MP3) and images (e.g. JPEG) (where small high frequency components can be discarded), to spectral methods for the numerical solution of partial differential equations. DCT is a Fourier-related transform similar to the discrete Fourier

transform (DFT), but using only real numbers, where in some variants the input and/or output data are shifted by half a sample [23]. The most common variant of discrete cosine transform is the type-II DCT, which is often called simply "the DCT", its inverse, the type-III DCT, is correspondingly often called simply "the inverse DCT" or "the IDCT" [18]. DCT will be used for the dynamic letter such as J, Z, and other dynamic signs where it will be applied on an image that represent the sign according to the palm position each position in the image represent the frequency of the palm position. The DCT tends to concentrate information, making it useful for image compression applications using Matlab toolkit [24]-[25]. After Applying DCT, zigzag scan is then applied on 8x8 part of the image. The main purpose of the zigzag scan is to group low frequency coefficients presented at the top of the vector. Although zigzag scan is under the title of Feature extraction, it simplifies the amount of resources required to describe a large set of data which reduces the complexity and computation power for the next stage which is Classification using ANN. Figure 4 shows a sample of 8x8 matrix representing the cropped part of the image after applying DCT and zigzag scan results (coefficients) that are stored in a vector.

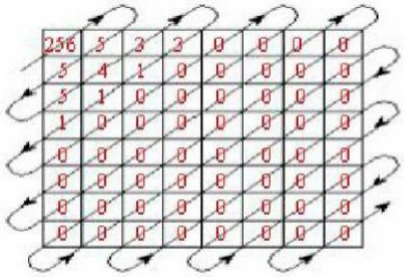


Fig. 4. Sample of 8x8 Zigzag scan method.

5.2 Static gesture features

For the detection of static letters and numbers the following feature extraction is obtained: Number of fingers raised, Hand orientationwhere extracting the tip position of all fingers from the hand and its direction in the form of 3D vectors as shown in figure 5. Which fingers raised and Distance between fingers (An array that contains the distances between each finger and all the other fingers of the hand) [26].



Fig. 5: Palm Fingers tip position and direction

4. 6. CLASSIFICATION USING ANN

Artificial neural networks have been widely used in sign language recognition research [27]. This paper presents recognition for ASL using ANN classifier. For dynamic signs, neural network is trained for the position recognition of a sign language with few letters and signs in the beginning. The neural network is a feed-forward two layers perceptron type with the backpropagation learning method. It requires 36 neurons and output layers equal to the number of gestures that are used at the training process. For the two dynamic signs (Z,J) ten images were taken for every letter from 10 different signers to give the total of 200 images dataset, where 70% for training (140 images) and 30% for testing (60 images). The gestures used in the training process are characterized by the feature extraction of the locus of the palm center after that applying DCT and zigzag scan to obtain a vector with coefficients to be entered to ANN.For static letters and numbers which are 34 signs, ten images were taken for every letter and number from 10 different signers to give the total of 3400 images dataset.

The resulted static feature extraction parameters are entered to ANN classifier forletter and number recognition. The user has the choice to select between letters and numbers.The neural network used for static symbols is a feed-forward three layers perceptron type with the backpropagation learning method. It requires 60 neurons in the hidden layer in order to accommodate for all symbols, where 70% training (2380 images) and 30% (1020 images) for testing.

5. 7. ACTUATION MECHANISM

This paper demonstrates how to use serial communication in order to control servos with leap motion. The goal of that is to keep the hand in the domain of the leap motion while performing the signs. To obtain this,Arduino Uno Micro-Controller is used with two motor servos drivers, and robotic arm.

7.1 Arduino Uno Micro-Controller

The Arduino Uno is a micro-controller board based on the ATmega328P as shown in figuer 6. It is used in this paper to control the servo motors to track the hand's motion. The Leap-Motion sends its vectors of (X, Y and Z) axises through the NODE.JS (an open-source, cross-platform runtime environment for developing server-side Web applications), the Arduino then receives the values and match it to the corresponding angles which the motors should move with to keep the hand's palm position in the centre of the camera vision.



Fig. 6.Arduino Uno Board

7.2 Servo Motors

The two servo motors are connected to the analog pins to control them with Pulse Width Modulation technique as in figure 7. Each motor is taking a distinct signal from the Arduino. This signal depends on the reading of the corresponding axis given by the Leap-Motion Controller. One for the X-axis and the other is for Z-axis of the camera (Leap-Motion).

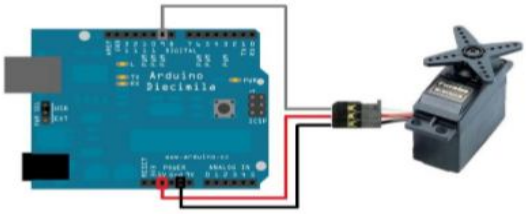


Fig. 7. Wiring servo-motor to the Arduino.

7.3 Robotic Arm

This describes the outer design and the micro-controller used in this research to control the hardware moving parts as follows:

1. Arduino Shield

It is a PCB made to minimize the amount of connecting wires and hence reduce the size of the final research project to be easily used and transport. It is directly connected to the Arduino's pins and then connecting the Servo Motors' wires tightly through it to replace the large bread board.

2. Pan Tilt

The Pan tilt is a metal used for connecting the servo motors together perpendicular on each other, and help in easy moving the motor on its axis.



Fig. 8. Metal Pan Tilt holding two servo motors.

3. Wooden Box

The outer base of the design is made from wood. It is a box containing the Arduino and its shield with the connecting wires. It is holding the X axis Servo Motor directly on the top. The rest of the wires and USB wires (Arduino and Leap-Motion) are then coming out through a small circle at the back of the box as in figure 9.



Fig. 9. The wooden box.

4. Metal Frame Holder

The Pan tilt which is stuck to the wooden box from the servo side, is connecting its edge with a Metal frame holder from the other side. It is used for holding the Leap-motion at the end as in figure 10.

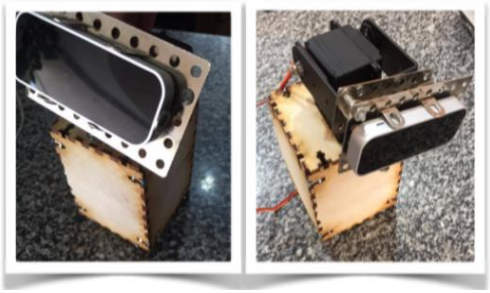


Fig. 10. Metal frame holder holding the Leap-motion onto the Pan Tilt on the wooden box.

VI. RESULTS AND DISSCION

The results of the presented paper work were done by using Leap Motion to obtain a dataset of 3600 images for ASL from different signers.Feature extraction using DCT for dynamic letters (200 images) were obtained followed by ANN classifier to give an average success rate of 87% accuracy. The rest of the dataset which are 3400 images are the static signs for both letters and numbers, using the following feature extraction: Number of fingers raised, Hand orientation, which fingers raised and Distance between fingers to be entered as parameters to ANN classifier to give average success rate of 82% and 85% accuracy for static letters and numbers recognition respectively. The overall average success rate is 84.66% accuracy for the recognition of all signs as shown in Table 1. The average weighted success rate turned out to be 83.11%. This paper is also supported with, a robotic arm thatis built usingArduino Uno Micro-Controller and two servos motors to track the hand in order to keep it in the small viewing domain of the Leap Motion while performing gestures.

Table. 1. Results of the presented paper

Static letters		Static numbers		Dynamic letters		Total Dataset	Total average success rate
Dataset	Average Success rate	Dataset	Average Success rate	Dataset	Average Success rate	3600	84.66%
2400	82%	1000	85%	200	87%		

A comparison is done between our dataset of 3600 images using leap motion and ANN classifier and other related work using different dataset with different classification techniques. Many factors are considered such as: the number of dataset used, static or dynamic recognition, and some papers didn't apply recognition for numbers or dynamic letters, and different types of classifiers are used to give different overall average success rate in each case. This is shown in Table 2.

Table 2. Comparison between the results of the proposed paper and other related work using different classification techniques.

Classifier	Static letters	Average Success rate of static letters	Numbers	Average Success rate of numbers	Dynamic letters	Average Success rate of Dynamic letters
Using Kinect supported by image processing [8].	26	84.1 %	---	---	---	---
Leap Motion using trained classifiers [9].	Not mentioned	98%	---	---	---	---
Leap Motion using k-nearest neighbor classifier [10].	26	72.78%	---	---	---	---
Leap Motion using SVM [10].	26	79.83%	---	---	---	---
Leap Motion using SVM [11].	10 letters (10 letters x 10 repeated times x 14 signers to obtain 1400 data samples)	80.9%	---	---	---	---
Kinect features using SVM [11]	10 letters (10 letters x 10 repeated times x 14 signers to obtain 1400 data samples)	89.7%	---	---	---	---
Leap Motion + Kinect features using SVM [11]	10 letters (10 letters x 10 repeated times x 14 signers to obtain 1400 data samples)	91.3%	---	---	---	---
The proposed system using Leap Motion , ANN, and the actuation mechanism	24letters (24 letters x 10 images x 10 different signers to obtain 2400 dataset)	82%	10 numbers(10 numbers x 10 images x 10 different signers to obtain 1000 dataset)	85%	2 letters(2 letters x 10 images x 10 different signers to obtain 200	87%

CONCLUSIONS

The paper's main target is to obtain a system to facilitate a way of communication for deaf people. This was done by getting ASL gestures using Leap Motion Camera Controller and actuation mechanism followed by features extraction techniques for both static and dynamic gestures. Those obtained features parameters are entered to ANN classifier to develop a model to recognize hand gestures for both static and dynamic signs. A robotic arm was built with Arduino Uno Micro-Controller and two servos motors to track the hand to keep it in the small viewing domain of the Leap Motion while performing gestures. The next challenge would then be to translate the characters into acoustic sounds and to upgrade this work so it can produce whole sentences with correct spelling and grammar.Moreover, implementing this work on other different languages signs like Arabic for example, and applying different classification techniques to get higher success rate accuracy.

REFERENCES

[1] Sign American Language, "Dictionary of American History", The Gate Group Inc, 2003.

[2] D. C. Baynton, "American Culture and the Campaign against Sign Language", 1996.

[3]R. E Johnson, S. K Liddell,"Toward a Phonetic Representation of Signs: Sequentialityand Contrast", 2010.

[4]L. Fant and B. Bernstein Fant, "American Sign Language Phrase Book", Third Edition.

[5] K. Mulrooney, "American Sign Language Demystified", McGraw Hill, 1st edition, 2009.

[6] Thad Eugene Starner, "Visual Recognition of American Sign Language Using Hidden Markov Models", 1995.

[7] L. Quesada1, G. López2, and L.A. Guerrero, "Sign Language Recognition Using Leap Motion", Springer International Publishing Switzerland, 2015, LNCS 9454, pp. 277–288, 2015.

[8] Sun, C., Zhang, T., Xu, C., "Latent support vector machine modeling for sign language recognition with Kinect". ACM Trans., Technol. 6(2), 1–20, 2015.

[9] A. Chan, T. Halevi, N. Memon, "Leap Motion For Authentication Via Hand Geometry and Gestures", Human Aspects of Information Security, Privacy, and Trust"pp 13-22, 2015.

[10]Chuan, C., Regina, E., Guardino, C."American Sign Language recognition using leap motion sensor". International Conference on Machine Learning and Applications, pp. 541–544. IEEE Press, New York, 2014.

[11] Marin, G., Dominio, F., &Zanuttigh, P." Hand gesturerecognitionwithjointlycalibratedleapmotionanddepth sensor". Multimedia Tools and Applications, 1–25, 2015.

[12] Cooper, H., Holt, B.,& Bowden, R. 2011. "Sign language recognition in Visual Analysis of Humans: Looking at People", ch. 27, 539 – 562, 2011.

[13] https://pediaview.com/openpedia/Leap_Motion]

[14] Leap motion. http://www.leapmotion.com

[15] Guna, J., Jakus, G., Pogačnik, M., Tomažič, S., Sodnik, J., "An analysis of the precision and reliability of the leap motion sensor and its suitability for static and dynamic tracking", Sensors 14(2), 3702–3720, 2014.

[16] https://robotics.shanghaitech.edu.cn/node/115]

[17]http://www.wisegeek.com/what-is-image-acquisition-in-image-processing.htm]

[18] Umbaugh, Scott E ." Digital image processing and analysis : human and computer vision applications with CVIP tools", second edition, Boca Raton, FL: CRC Press. ISBN 97814-3980-2052, 2010.

[19]"Image processing". Available at: http://en.wikipedia.org/wiki/Image_processing

[20] John D. Cook, "Three algorithms for converting color to grayscale". Available at: http://www.johndcook.com/blog/2009/08/24/algorithms-convert-color-grayscale/

AddThis Sharing Buttons

Share to Facebook2Share to TwitterShare to EmailShare to LinkedInShare to More2

[21] M. Funasaka, Y. Ishikawa, M. Takata, & K. Joe, "Sign Language Recognition using Leap Motion Controller", Nara Women's University, Japan.

[22] Wallace G. K, "Overview of the JPEG still Image Compression standard" SPIE, vol. 1244, pp. 220-233, 1990.

[23] Zhu O., AlwanA., "An Efficient and Scalable 2D DCT-Based Feature Coding Scheme For Remote Speech Recognition", IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 1, 2001.

[24] Matlab Student version, The Math Works kit 2015.

[25] The Math Works, Available at: www.mathworks.com

[26]L.Shao,"Hand movement and gesture recognition using Leap Motion Controller", Stanford EE 267.

[27] K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks", pp. 237-242, 1991.

★★★