International Conference on Engineering Research & Applications (ICERA)
Istanbul (Turkey) May 17-18, 2017

An Adaptive Fingertips Detection Based on Skeletonization Technique for Invariant Hand Gestures

Md Jan Nordin[1], Ahmad Yahya Dawod[2] and Junaidi Abdullah[3]

[1]Centre for Artificial Intelligence Technology (CAIT), Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

[2]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia

[3]Assistive Technology SIG, Faculty of Computing and Informatics, Multimedia University (MMU), 63000 Cyberjaya, Selangor, Malaysia

***Abstract*:** *Finger identifications can be used to identify static hand poses letters and interprets pose sequences of gestures words terms. We present a new algorithm by using skeletonization and YCbCr color space for segmentation of various hand gestures followed by fingertips detection for each gesture. Fingertips detection is quite a challenging task for sign language recognition. Firstly, RGB color space is converted into YCbCr color space in the image. The binary image is done based on thresholding method that is by using range pre-defined for Cb and Cr value that represent the region value color. Hand gesture segmentation is completed by conducting the morphological process of a binary image obtained. Fingertips location are calculated by using skeletonization technique. The Euclidean distances and centroid are created between the coordinates of the pixels and the centroid belonging to the candidate fingertips are calculated to validate whether they represent actual fingertips. Experiments carried out was able to recognize 26 American Sign Language (ASL) alphabets and number samples of hand sign with an accuracy of 97.63%.*

Keywords: *Fingertips Detection; Centroid of hand; Segmentation; Skeletonization technique.*

1. Introduction

The recognition of hand gestures has become an important research topic because it introduces the potential use for human-computer interaction. Human hand gestures include a wide range of non-verbal communication. Examples range from the simple ones like pointing, to more complicated gestures like sign language. The accurate detection of hands in video or still images is still a challenging problem since the appearance of hands varies. Methods for detecting fingertips and recognising gestures are generally based on two stages. Segmentation of hand is the first stage and is followed by fingertips detection. When the hand is segmented, the areas of interest needed to find the fingertips are reduced. As a result, the detection process becomes quicker. However, most of the methods used for hand segmentation are not able to achieve good results because of factors like poor lighting conditions, cluttered background, and fast hand motion. Invalidation of the results of fingertip detection usually happens because of the poor performance of the hand segmentation method [1]. Hard-of-hearing or deaf people have utilised sign language as one of their several communication options. As a language, it makes use of signs made through hand motions as well as facial expressions and body postures. Identification of gesture language has a very high applicability. It also belongs to the areas being explored to help the integration of deaf people into the community. To detect gestures, researchers made use of specialised equipment (i.e. recognition techniques or gloves related to image processing and utilising computers and cameras).

https://doi.org/10.15242/DIRPUB.DIR0517046

Two trends are generally used to study the problem of hand gesture recognition. These trends correspond to two data types, namely the static and dynamic gestures. Static gestures are generally described through information gathered from hand shape and pose. On the other hand, the dynamic type is represented by hand motion. However, these two trends are closely related because the development of many of the methods used for dynamic gesture identification is by using the solutions of static gesture recognition. Therefore, the basis of study on general gesture recognition can be considered as the static gesture identification [2]. The detection of fingertips for sign language has proven to be a quite difficult task. Most of the research related to hand segmentation is based on skin colour detection. The main idea of this approach lies on the assumption that the colour of the skin is different from the colours of the other objects. Consequently, cluster formation in a few specific colour spaces results from its distribution. The detection of fingertips for sign language has proven to be a quite difficult task. Most of the research related to hand segmentation is based on skin colour detection. The main idea of this approach lies on the assumption that the colour of the skin is different from the colours of the other objects. Consequently, cluster formation in a few specific colour spaces results from its distribution.

2. Related work

Recently, hand gesture has been considered as one of the non-verbal approaches of communication that can be utilised in Human-Computer interaction. Additionally, this technique has even been used for human-robot interaction (HRI). Gathering of the images of hand can be done simply by using a camera that is attached to a robot or a computer. However, there are a few processing stages that have to be performed in order to get the gesture of hand meaning. Before starting the process of hand gesture recognition, extraction of some hand features like its finger directions, its centre, and the fingertip positions should be done.

Jmaa et al. [3] presented a hand segmentation technique that is based on skin colour and which utilises the YCbCr colour space. To detect the skin, extraction of the colour space's luminance component was done through the use of YCbCr. However, the method has proven to be unreliable because finding a skin colour model to represent people of various ethnicities and be applicable under different illumination conditions is difficult. A simpler system was proposed by [4]. This system uses the skin colour segmentation and artificial neural network. But, these methods extract features that do not have enough hand shape information. This information is an important factor in achieving accurate recognition. Therefore, the generalisation level for these methods is not particularly high.

Cooper et al. [5] suggested utilising a set of new feature pixels to locate the position of the fingertips. The emphasis of this method lies in locating the fingertip position in an image frame that consist of both the hands and the face. These feature pixels, also known as distance based feature pixels, are extracted using a method that is based on the distance transformation of the connected component labelling image. This makes the detection of the binary skin colour quicker and easier. As a result, hand tracking and hand gesture recognition can be performed with more accuracy and ease. Furthermore, these processing stages do not involve any complex calculations, allowing one to achieve real-time performance for hand tracking system or the hand gesture recognition [6]. A presentation given by [7] regarding the computer vision system for the identification of six American Sign Language (ASL) hand signs utilises colour information and combined hand geometry. This system has an overall accuracy rate of 89%.

3. Proposed Methodology

This research study's proposed methodology has four key stages. These four stages are RGB Image, Segmentation, Morphology, and Detection of Fingertips.

3.1. RGB image

The first stage implicated the acquisition of the hand gestures for Sign Language alphabets and numbers. A camera was used to capture these images. The obtained images may consist of changing light condition and cluttered background conditions. In addition to the static American Sign Language (ASL), gathering of the normal static hand gesture images was done based on a Logitech USB 2.0 webcam. The resolution was set at 320 x 240 pixels. Generally, the next steps will be less complicated if there is a high contrast between the hand and the background as well as if image is taken with a simple background. Several open data sources were used to gather some of the images being utilised [8]. Several of the existing online datasets at "http://www.lifeprint.com/asl101/pages-layout/handshapes.htm" became sources for the American Sign Language (ASL) images. A few images of signs i.e. (0 to10) in numbers and Alphabet (A to Z) of the American Sign Language were used. This is shown in Fig. 1.

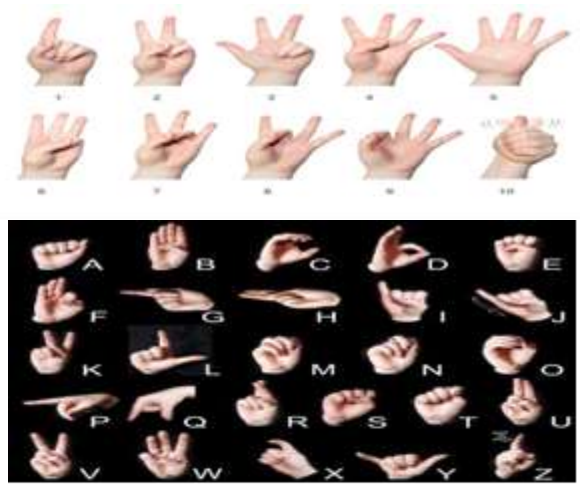


Fig. 1: Images of numbers and `Alphabets from the dataset.

3.2. Segmentation

In the proposed method, conversion RGB of images to YCbCr colour space is done using "Equation (1)" to "Equation (3)", where R, G and B are used to represent Red, Green and Blue pixel values, respectively. Y represents the luma component, and Cb and Cr represent the blue-difference and red-difference of the chroma components, respectively. Simple thresholding was done to segment the hand for those with simpler background images. Hand segmentation for complex backgrounds and hand gestures that involve people from various ethnicities was done using the free-form skin colour model described in [9]. Equation (4) shows the condition of the thresholding process. A binary image is produced, where the hand is represented by the white pixels (value of 1) and the background is signified by the black pixels (value of 0).

$$Y = 0.299R + 0.587G + 0.114B \qquad (1)$$

$$Cb = (B - Y) * 0.564 + 128 \qquad (2)$$

$$Cr = (R - Y) * 0.713 + 128 \qquad (3)$$

$$SkinColor(x, y) = \begin{cases} 1 & if\ (72 \le cb \le 134) \cap (133 \le cr \le 190) \\ 0 & Otherwise \end{cases} \qquad (4)$$

The thresholding of the algorithm is stated below:

```
for i = 1: a
  forj = 1: b
   if(Cb(p,q) > 90 && Cr(p,q) < 135 && Cb(p,q) > 140 && Cr(p,q) < 190)
    ImageHand(i,j) = 1;
   Else
    ImageHand(i,j) = 0;
    end
  end
end
```

In the previous algorithm, the YCbCr colour space used the Cb and Cr pixel values are represented by the Cb and Cr matrices. The resulting binary image is represented by the ImageHand matrix. Here, determination of the threshold values by using the positive results that they have supplied experimentally. The hand segmentation result is shown in Fig. 2.

(a) (b)

Fig. 2: Hand Segmentation (a) Before Segmentation
(b) After Segmentation.

3.3. Morphology

Pre-processing of image is a common name for image operations that deal with the lowest level of abstraction for image data enrichment. The process involves reduction or removal of undesired noises and enrichment of some image features to carry out further analysis and processing at the later stage. However, it does not add content to the image information. The hand segmentation binary image goes through structural processes: ‘opening’ followed by ‘dilation’. The opening operation involves a series of functions, i.e., removal of miniature connected components, breaking down of narrow isthmuses and smoothening of contours. In the dilation operation, a disk-shaped structuring element is used to structure the contour and increase the size of the skin region, which later aids in performance enhancement of the region labelling operation. All these help create a fine morphed image. This study involves employing dilation and erosion with 3×3 structuring elements. This was intended to eliminate small background objects and make the hand conspicuous from the background, as shown in Fig. 4. Let $N \in x^2$, i.e. the 2D space of the (a,b), and let $M \in x^2$ be the SE that controls the morphological operations’ structure. Now, for any binary image I (where $N \subseteq I$), dilation and erosion can be defined as in Equations (5) and (6), respectively [10]. Considering Dilation Operator is given by the sets of N and M in X^2, the dilation of N by M is defined as:

$$N \oplus M = \{X \mid (\hat{M})_x \cap N \neq \phi\} \tag{5}$$

Similarly, considering Erosion Operator is given by the sets of N and M in X^2, the erosion of N by structuring element M can be defined as:

$$N \ominus M = \{X \mid (M)_x \subseteq N\} \tag{6}$$

The erosion of N by structuring element M is defined by the set of all points x, such that M, translated by x, is contained in N.

(a) (b) (c)

Fig. 3 (a) Source image (b) Before Dilation and Erosion,
(c) After morphological operation

However, a few background objects that have almost identical colour to that of the skin continue to exist even after performing morphological operations. To solve this problem, a blob detection method is applied. In this, a unique label is allocated to each blob to make it distinct from other blobs. The same labels are also assigned to all of the pixels within a spatially connected blob of 1s. First, the blob with the highest number of pixels is selected as it represents the hand, and then all other blobs are removed for the study of researchers.

3.4. Coordinate Hand

The fingertips are detected by skeleton algorithm, and the palm center is located by finding the point that has the maximum distance to the closest region boundary. Assume through measurement by five fingers to average person stretches and the angle five fingers are between $20^\circ - 180^\circ$. In this work begin by straight line between the palm centre coordinate and the fingertips. It consider the segmented hand region, its centroid or centre given by $D(x_C, y_C)$ is the centroid of the hand image is calculated. In this research work, the silhouette is calculated the absolute approach α_d and the relative approach β_d. The five fingertip coordinates based on $S_1(m_1, t_1)$,…, $S_d(m_d, t_d)$ using $d \in [1,2,3,4,5], d$ is the fingertips number. Define Gesture 1 by single fingertips and more than one finger can angles calculate [11].

The absolute approach it calculate through the palm centre (centroid) coordinates $D(x_c, y_c)$ and the fingertips coordinates $S_d(m_d, t_d)$. If $m_d - x_c < 0$, finger is in the first or second quadrant of the palm follow equation (7)

$$\alpha_d = \frac{180^\circ}{\Pi} \arccos \frac{(t_d - y_c)}{\sqrt{(m_d - x_c)^2 + (t_d - y_c)^2}} \tag{7}$$

Or $m_d - x_c > 0$, the fingertips is in third or fourth quadrant of the palm follow equation (8):

$$\alpha_d = 360^\circ \frac{180^\circ}{\Pi} \arccos \frac{(t_d - y_c)}{\sqrt{(m_d - x_c)^2 + (t_r - y_c)^2}} \tag{8}$$

Else $m_d - x_c = 0$ and $t_d - y_c > 0, \quad \alpha_d = 0^\circ$ the finger is to the Right of the palm

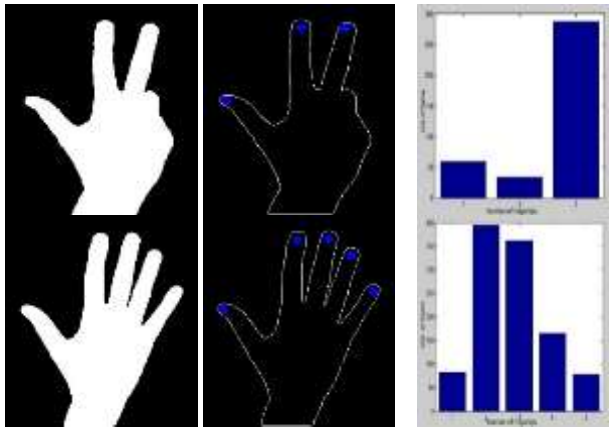
Or $m_d - x_c = 0$ and $t_d - y_c < 0, \ \alpha_d = 180^\circ$ the finger is to the Left of the palm

The relative Approach is calculate by $\beta_1 = |\alpha_2 - \alpha_1|$

$$\beta_d = |\alpha_1 - \alpha_d|$$

3.5. Detection of Fingertips

Skeletonisation is utilized to detect fingers in image of hand. In this process, until a skeleton of size one pixel is obtained, the hand blob is ‘thinned’ continuously. More details on the algorithm proposed by Dawod et al. can be seen in [12]. The line detection process makes use of edge detection approaches such as the first step. The edge detection approach is employed to detect boundaries of complex object. In this, possible edge points are designed such that they correspond to the portions in an image with difference or changes in the brightness levels. The best edge operators is Sobel operator with 3x3 mask that employs a median filter to smoothen the hand image for recognition (as displayed in Fig. 4). Hand pose is then identified by analyzing different binary frames involving fingertips detection that corresponds to the histogram of each known fingertips [13].



(a) (b) (c)

Fig. 4 (a) Binarization images. (b) Fingertips detection.
(c) The corresponding Histogram for various fingertips frames gestures.

4. Experimental Results

The fingertips skeleton image might consist of some noise pixels and these are eliminated by regional connectivity analysis. The outcome of fingertips skeleton is depicted in Fig. 5(c). From Fig. 6, it is evident that the position of the fingertip should be an endpoint on the finger skeleton for fingertip detection.

(a) (b) (c)

Fig. 5: (a) Original image. (b) Segmentation of hand.
(c) Skeleton of fingertips.

To identify whether the endpoints denoted the fingertips or the wrist, the research scholars carried out the following steps. The first step involved the examination of the endpoint's convex hull. Next, the distance between the neighbouring endpoints was observed. In the inference, the endpoints that were far apart or had the longest distance between them were marked as the wrist, whereas the remaining endpoints were marked as fingertips.

(a) (b)

Fig. 6 (a) image of Skeleton (b) convex hull.

A PC consisting of Intel Core i7, 2450 M, 8G, 1.97 GHz processor was employed for this experiment. The MATLAB software with Windows 8.1 operating system was employed. The dataset utilized comprised one single-handed gesture image with lighting conditions and various backgrounds. The resolution of image used was 320×240. The Static Hand Posture Database http://www.lifeprint.com/asl101/pages-signs/n/numbers.htm and http://www.idiap.ch/resource/gestures were used to collect the dataset. The centroid or centre (x_c ,y_c) of the hand image is calculated from the segmented hand region. In this study, the following formulae are employed to calculate the silhouette:

x_i is the x-coordinate of each edge pixel of the image as depicted in Equation (9).

$$X_C = \frac{\sum x_i}{M} \tag{9}$$

y_i is the y-coordinate of each edge pixel of the image as depicted in Equation (10).

$$Y_C = \frac{\sum y_i}{M} \tag{10}$$

Where x_i and y_i represent the x and y coordinates of the i-th pixel in the hand region, M is the total number of pixels in the hand region and s denotes the total number of edge points. Finally, Δx and Δy centroid of the hand image are calculated [14]. Fig. 7 and Fig. 8 depict the outcomes of the fingertips detection and the calculated centroid coupled with the identified fingertips.

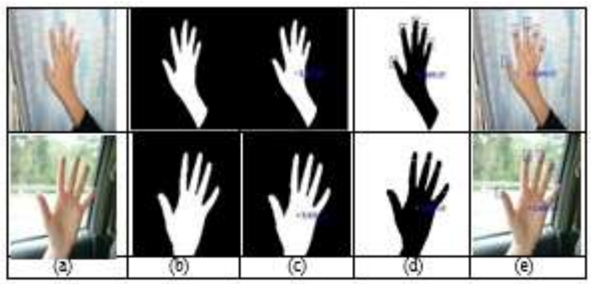
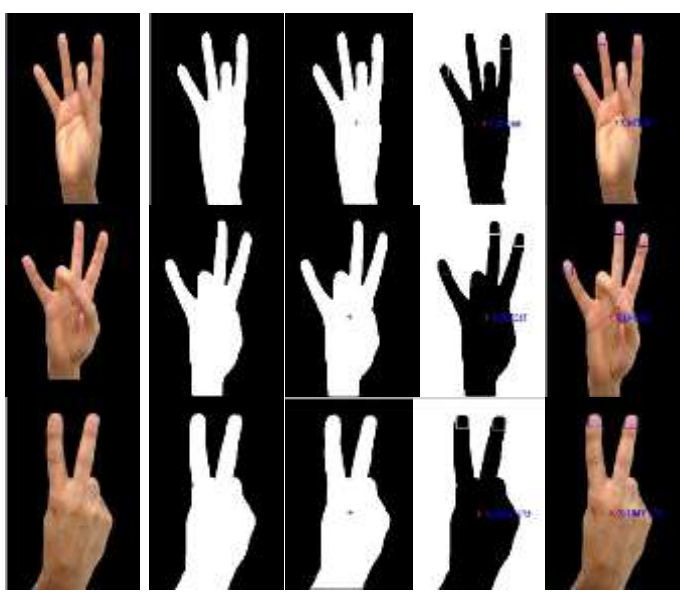


Fig. 7: Fingertips detection for outdoor images and complex background.

The detection of the fingertips for the hand image that has a cluttered background have been shown in Fig. 9. It is clear from the outcome that the detection of the fingertips for all images with a cluttered background is successful when all fingers are open. The fingertips detection for the datasets of American Sign Language (ASL) alphabet is depicted in Fig. 10. The detection of the fingertips of all the images was possible by our method. However, our method could not detect fingertips that were bended into the palm. We identify this as a limitation of our method and consider this as the course of our future work.



(a) (b) (c) (d) (e)

Fig. 8 (a) Source image (b) Segmented of hand (c) detection of Centroid
(d) The binary image of Fingertips Detection, (e) Original image of fingertips detection

The confusion matrix for binary classification is categorised into four, where F_P denotes false positive, F_N represents false negative, T_P represents true positive and T_N represents true negative counts. The x-axis constitutes the false positive rate (FPR) and the y-axis constitutes the true positive rate (TPR). This means recall is plotted on the x-axis and precision is plotted on the y-axis. While the recall denotes the true positive rate, the precision denotes the positive predictive. The recall is considered the same as TPR. Our recommended method is compared with other works, which is given in Table 1. When compared to the other methods, we attained a good accuracy level of 97.63%, precision of 92.46% and recall of 95.33%. Equations (11)-(14) demonstrate the method to calculate accuracy (ACC).

$$\text{Recall } TPR = \frac{T_P}{T_P + F_N} \tag{11}$$

$$\text{Specificity } FPR = \frac{F_P}{F_P + T_N} \tag{12}$$

$$\text{Precision} = \frac{T_P}{T_P + F_P} \tag{13}$$

$$ACC = \frac{T_P + T_N}{T_N + F_P + T_P + F_N} \tag{14}$$

TABLE 1: Method Comparison

Method	Precision	Recall	Accuracy
[15]	89.41%	97.72%	95.56%
[16]	91.37%	85.30%	97.44%
[17]	93.56%	96.69%	99.00%
Proposed	92.46%	95.33%	97.63%

5. Conclusions and Future work

Based on skeletonization and some morphological operation, this study proposed a novel technique for the detection of fingertip positions. The new method is highly efficient in terms of identifying static hand gestures. This approach can be used for human-computer interaction and sign language recognition areas as well. This research used the free-form skin colour model demonstrated in [12] coupled with morphological filtering to achieve a high level of accuracy in hand detection. From the experiments, it is evident that the proposed method is capable of detecting all the open fingertips from images that have a cluttered background and from the ASL alphabet and number dataset that executes different gestures. As the future course of work, we aim at enhancing our new method and making it capable of recognising dynamic hand gestures in long-term studies. Also, this work can be further investigated to find ways to overcome the technical limitations for carrying out the recognition of the static gestures.

6. Acknowledgements

The researchers wish to thank Universiti Kebangsaan Malaysia (UKM) and Ministry of Education, Malaysia for supporting this work by research grant DIP-2014-018 and DIP-2016-018.

7. References

[1] T.-N. Nguyen, H.-H. Huynh, and J. Meunier, "Static Hand Gesture Recognition Using Artificial Neural Network," Journal of Image and Graphics, vol. 1, pp. 34-38., 2013.
https://doi.org/10.12720/joig.1.1.34-38

[2] T.-N. Nguyen, D.-H. Vo, H.-H. Huynh, and J. Meunier, "Geometry-based static hand gesture recognition using support vector machine," in Control Automation Robotics & Vision (ICARCV), 2014 13th International Conference on, , pp. 769-774, 2014.

https://doi.org/10.1109/icarcv.2014.7064401

[3] A. B. Jmaa, W. Mahdi, Y. B. Jemaa, and A. B. Hamadou, "Hand localization and fingers features extraction: Application to digit recognition in sign language," in Intelligent Data Engineering and Automated Learning-IDEAL 2009, ed: Springer, pp. 151-159, 2009.

https://doi.org/10.1007/978-3-642-04394-9_19

[4] R.-L. Vieriu, I. Mironica, and B.-T. Goras, "Background invariant static hand gesture recognition based on Hidden Markov Models," in Signals, Circuits and Systems (ISSCS), 2013 International Symposium on, , pp. 1-4, 2013.

https://doi.org/10.1109/isscs.2013.6651245

[5] H. Cooper, E.-J. Ong, N. Pugeault, and R. Bowden, "Sign language recognition using sub-units," The Journal of Machine Learning Research, vol. 13, pp. 2205-2231, 2012.

[6] L. Dung and M. Mizukawa, "Fast hand feature extraction based on connected component labeling, distance transform and hough transform," Journal of Robotics and Mechatronics, vol. 21, p. 762-738, 2009.

[7] P. J. Martina, P. Nagarajan, and P. Karthikeyan, "Hand Gesture Recognition Based Real-time Command System," International Journal of Computer Science and Mobile Computing, pp. 295-297, 2013.

[8] http://www.lifeprint.com/asl101/pages-layout/numbersdiscussion.htm

[9] A. Y. Dawod, M. J. Nordin, and J. Abdullah,. "Static Hand Gestures Fingertips Detection Based on Segmanted Images" Journal of Computer Science. Vol 11, Issue. 12, pp. 1090-1098, 2016.

[10] M. M. Hasan and P. K. Mishra, "Real time fingers and palm locating using dynamic circle templates," International Journal of Computer Applications, vol. 41, 2012, pp. 33-43.

[11] Yu, Bo, YongQiang Chen, YingShu Huang, and Chenjie Xia. "Static hand gesture recognition algorithm based on finger angle characteristics." In Control Conference (CCC), 33rd Chinese, pp. 8239-8242, 2014.

https://doi.org/10.1109/chicc.2014.6896380

[12] A. Y. Dawod, J. Abdullah, and M. J. Alam, "A new method for hand segmentation using free-form skin colour model," in Advanced computer theory and engineering (ICACTE), 3rd international conference on, pp. V2-562-V2-566, 2010.

[13] H. Hasan and S. Abdul-Kareem, "Static hand gesture recognition using neural networks," Artificial Intelligfence Review, vol. 41, pp. 147-181, 2014.

https://doi.org/10.1007/s10462-011-9303-1

[14] M. Bhuyan, D. R. Neog, and M. K. Kar, "Fingertip detection for hand pose recognition," International Journal on Computer Science and Engineering, vol. 4, pp. 501-511, 2012.

[15] A. Mittal, A. Zisserman, and P. H. Torr, "Hand detection using multiple proposals," in BMVC, pp. 1-11, 2011

https://doi.org/10.5244/c.25.75.

[16] A. Tsagaris, and S. Manitsaris, "Colour space comparison for skin detection in finger gesture recognition". International Journal of Advances in Engineering & Technology, vol 6(4), pp. 1431-1437. 2013.

[17] Z.-h. Chen, J.-T. Kim, J. Liang, J. Zhang, and Y.-B. Yuan, "Real-Time Hand Gesture Recognition Using Finger Segmentation," The Scientific World Journal, pp. 1-9. 2014

https://doi.org/10.1155/2014/267872