A Hybrid Machine Learning Model for Hand Gesture Recognition

By

Jayed Talukder
ID: 183-25-705

This Report has been Prepared in Partial Fulfillment of the Requirements for the Degree of Master of Science in Computer Science and Engineering.

Supervised By

Md Zahid Hasan
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

August 2019

APPROVAL

This Project/thesis titled "A Hybrid Machine Learning Model for Hand Gesture Recognition", submitted by Jayed Talukder, ID No: 183-25-705 to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 26-08-2019.

BOARD OF EXAMINERS

for

Dr. Syed Akhter Hossain — **Chairman**
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Sheak Rashed Haider Noori — **Internal Examiner**
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Abdus Sattar — **Internal Examiner**
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Dr. Mohammad Shorif Uddin — **External Examiner**
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that, this thesis has been done by me under the supervision of **Md Zahid Hasan, Assistant Professor, Department of CSE,** Daffodil International University. I also declare that neither this thesis nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:

Zahid Hasan

Md Zahid Hasan
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Submitted by:

Jayed Talukder

Jayed Talukder
ID: 183-25-705
Department of CSE,
Daffodil International University

ACKNOWLEDGEMENT

My heartiest gratitude to the Almighty Allah for His divine blessing that made me capable of completing this project successfully.

I would like to thank my honorable teacher & thesis supervisor **Md Zahid Hasan, Assistant Professor, Department of CSE,** Daffodil International University for his endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior draft and correcting them at all stage have made it possible to complete this project.

I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

Sign language, a gesture based nonverbal technique of communication, is used by a significant number of populations throughout the world. Sign language recognition emerges as one of the most challenging exercises when the person interpreting it lacks the previous knowledge of signs. This paper proposes a method to recognize sign language using computer vision. Five features from a binary image of a hand shape, namely the area of the closed contour, the area of the convex hull, number of convexity defects, maximum depth of the defects, and the sum of the depths of the defects are extracted after preprocessing the image. In this study, for recognizing sign languages, the recognition task was accomplished by employing two classification algorithms: k-nearest neighbor (k-NN) and Support Vector Machine (SVM) individually. Then the system will decide the class of the gesture acquired from the result of a Logical AND operation from both of the classifiers.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background

A sign language uses hand gesture to express messages instead of audible sound pattern. The mute and deaf community solely relies on sign language for communication on a wide variety of daily needs. There are a lot of varieties of sign language, among which ASL is the most common one that is used across the globe.

American Sign Language is the sign language used in the country United States and parts of the country Canada. Other sign languages like British Sign Language (BSL) has significant difference from ASL. This research paper proposes the gesture recognition methodology based on ASL.

Recognizing sign language needs prior knowledge of hand gestures. To allow the mute and deaf community the seamless access to social opportunities and job markets, the other party must at least understand the message that are expressed in gestures. This is one of the primary motivations to focus the research attention to sign language recognition.

ASL, shown in Fig. 1.1, contains morphology, phonology, and syntax like any other languages used in the world. Similar to various sign languages, phonetics are composed of the shape and movement of the hands. These are accompanied by the body posture and facial expression which includes mouth [8].

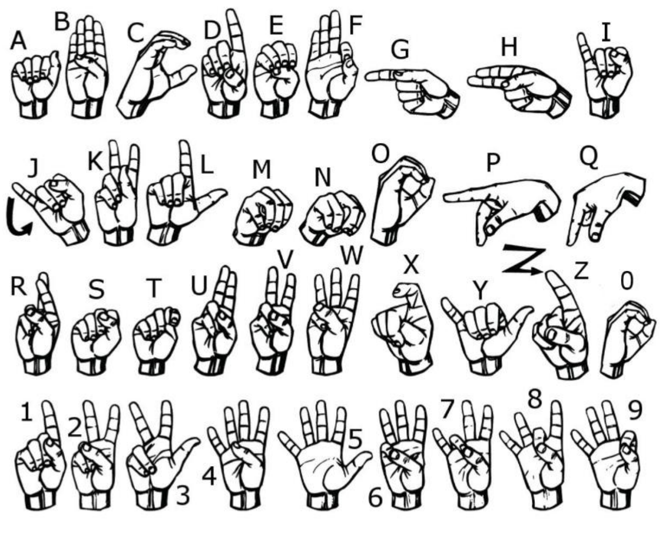

Fig. 1.1: American Sign Language

Sign languages are commonly misunderstood as being gestures without much meaning, but while iconicity plays a larger part in sign language than in oral language, signs are generally non-iconic.

Gesture recognition is one of the demanding topics that many scholars have done their researches on. The ultimate aim of a gesture recognition system should be the ease of usability. That is why this paper proposes a concrete recognition method which takes image input from a simple web camera to identify different gestures.

While recognizing hand gesture, it is important to extract features that are identifiable not only in the visible image but also derivable from the characteristics of the contour. The error rate of recognizing the correct gesture drops significantly if a combination of more than one recognition method is used.

For detecting a hand gesture, the edge of the hand shape can be a fundamental feature. Discontinuity of bright pixels in a binary image are likely to resemble changes in the properties of the image. There are several highly popular algorithms among which the Canny Edge Detection algorithm is a preferable one in many cases.

This paper investigates on ASL alphabets and numbers using the proposed methodology. In summary, the major contributions in this study are as follows:

a) The features of several hand shapes were studied closely to short list the features that varied significantly. Unlike few proposed methods [16] where the number of features extracted from the contour is very limited, this study uses five important features for improved the accuracy.

b) Unlike many common approaches to recognize gestures based on a single method, this research proposes a hybrid of two methods that classify an image independently for minimizing error rate. Both of the classifications are taken into account for recognizing an ASL sign.

1.2 Motivation

The research is motivated because it has a wide range of application that is essentially beneficial in helping the communication between members of the mute and deaf community and the common people. Lack of prior knowledge of sign language is an obstruction that falters the engagement of the deaf community in many aspects of life. A deaf could easily perform alongside a person without any disability with a proper system of conveying messages.

Those who endure being hearing impaired and have limitation speaking should not be abandoned from communicating with the rest of their peers. Just because a portion of a population cannot hear or speak, does not necessarily mean that the way they interact should not advance along with the rest of the world. Sign language is the mean for the mute and deaf to express their feelings, contribute to a conversation, learn, and overall spend their lives as a natural human being would. As a necessary trend, many institutions are conducting seminars and classes for sign language into their course catalogs.

1.3 Objectives

The objective of the proposed method is to maximize the ease of sign language detection. Minimizing the error rate of recognition is another aim which lead to the proposal of using a hybrid model. If both of the sub systems working for the same goal produces the same result, the correctness is justified with confidence. In case of failure, one sub system can play the role of a fail-safe. The proposed method does not discourage people to learn sign language and use the system instead. Rather, it can be seen as a bridge to both of the communities who has less to no expertise in expressing via sign language. Quicker means of sign language detection will always have benefits both short term and long term. The ultimate goal is to engage people with least hassle.

1.4 Report Layout

This paper is divided into several sections. Chapter 1 presents the introductory information and objectives. Chapter 2 discusses the main literature review which includes the related work in the domain of sign language recognition. Chapter 3 describes the hybrid model. Chapter 4 focuses on future research scope. In that section, enhancement ideas along with existent limitations will be discussed. The very last section contains the references of published research articles, journal papers, wiki pages, articles, blogs, forum threads etc. which is used in this paper.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

A handful of sign language recognition system has been proposed and developed by many scholars till now. Primarily, the methods can be divided into two approaches: vision-based approach and hardware-based approach.

2.2 Hardware-based Approach

Hardware-based approach mostly include the use of gloves attached to sensors. This may cause discomfort and an unviable way to recognize gestures. In [1] the author has used a 7-sensor glove to extract features and a Neural Network to train the data. ANNs are used to recognize the sensor data that are retrieved from the glove being used. A limitation is, a glove is able to capture the shape of the hand but not the motion of rest of the parts of the body, e.g. arms, elbows, face, etc. so only postures are taken in their project. The most useful gesture that can be done by the four fingers is bending towards the person's palm and then going back to the starting position. A flex sensor resolves the amount of finger curvature, which is then used widely by both the researchers and the engineers, is the most common of these kinds of sensors. Flex sensors in the gloves was used in this method [2]. In this system two hand gloves is implemented to capture the hand gestures of a user. The data glove is equipped with flex sensors along the length of each finger. These sensors output a stream of data that varies with how much it is bent. The gesture is thus identified and the relevant information of the text can be detected.

Microsoft Kinect is a new technology containing voice and body recognition sensors which is usually used as controller for the XBOX gaming console. It gives users the capability to get engaged in games by using their voice and body movement and gestures, without the need of additional hardware or accessories to track their body movements. It has in built RGB video camera, a depth sensor that represents 3D visualization of the environment, a microphone for recognizing user's voice and a property software that does the work of human body recognition. Microsoft Kinect sensor were also used to acquire sensor readings as well by many researchers [15].

2.2.1 Limitations

Constantly wearing a glove does not ease the user experience. Hence, vision-based approaches seem feasible in terms of flexibility and accuracy. Moreover, dynamic gesture recognition is difficult while using sensor data from gloves. If the hardware does not support wireless transmission, more parts need to be included for displaying the result or making sound wave for the result.

2.2 Vision-based Approach

In [3], the recognition system is based on TSL (Hue, Saturation, Luminance) adaptive area detection. It proposes a system that uses HMD (head-mounted display). Afterwards, the distance signatures are acquired from hand shape contours. Ultimately, the finger points are labeled by the information of signature. One other approach [4] uses the Eigen value weighted Euclidean distance to classify gestures. The detection happens in four steps in this methodology: the first step includes skin filtering. Afterwards, feature extraction and classification are done with an intermediate step of hand cropping. Skin segmentation and figure recognition are vital sub problems in gesture recognition which has various solutions like using geometric features of hand [5] and genetic algorithm [6]. Many computer vision applications those deals with the recognition of human body along with their activities often require the segmentation of skin regions as a pre-processing step. However, the detected regions may not necessarily correspond to skin. That is another challenge to keep in mind whenever working with skin segmentation. Edge information is also a viable feature [7] for skin segmentation from colored images. Contour sequence [8] was also used as a feature to detect ASL after image segmentation and morphological filtering. After completing edge detection, upon getting the boundary of the hand gestures from the database, Localized Contour Sequence (LCS) technique is applied for classification process. Upon applying the LCS on the gathered dataset, it tracks the border of the blob that is essentially the contour in a rotative direction and the contour pixels are counted sequentially.

2.2.1 Limitation

Recognition rate my drop significantly in different environments (ie. low light, distance from the camera etc.). Rate of correctly detecting skin deteriorates drastically in different lighting environments. Skin color detection is one of the primary steps to detecting gestures. Failure to do so may lead to production of imperfect result. Variable distance from the camera may also trigger the increase of error rate simply because the trained data set does not match in many cases.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Outline

Fig. 3.1 illustrates a flow diagram of the proposed methodology. The method is described in below:

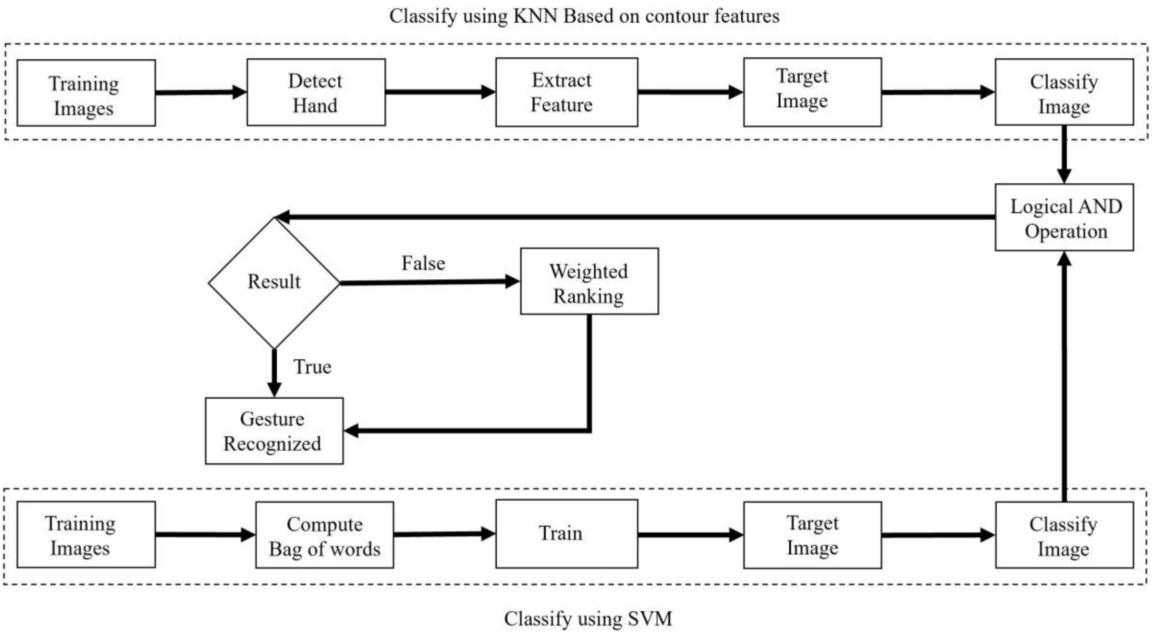

Fig. 3.1: Research Methodology

3.2 Input Acquisition

Training Images were captured from a web camera of 10 mega pixels. 30 varieties of a single gesture were captured. Images were resized to 500 X 500 images.

3.2 Preprocessing

The captured images for training were processed before extracting features. This includes skin color detection, generating a binary image based on threshold, applying Erosion and Dilation for noise removal, and finding the biggest blob that essentially is the hand. Fig. 3.2 illustrates the steps of preprocessing.

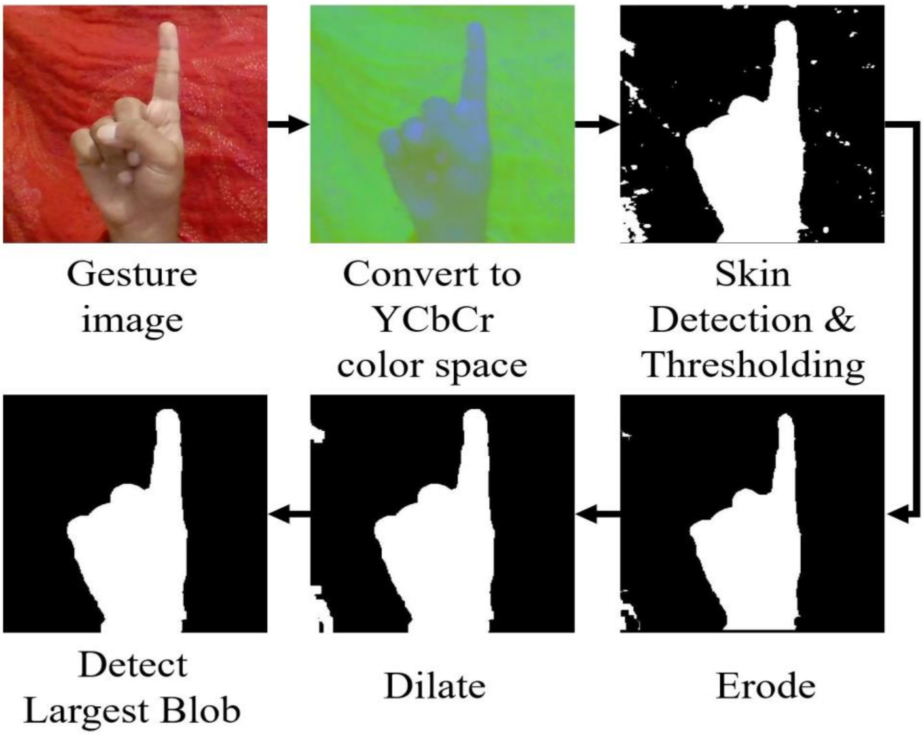

Fig. 3.2: Preprocessing of Images

3.2.1 Skin Detection

The captured RGB images were converted into YCbCr images. YCbCr, that is well known as YCBCR or Y′CBCR, is a color space that is different from HSL or RGB color space and is used much widely in video and digital photography systems. Y′ is the luma component. The CB represents the blue difference where the CR represents the red difference. Y′ (with prime) is detected from Y which is luminance. It is basically the intensity of the light that is is non-linearly encoded using the gamma correction.

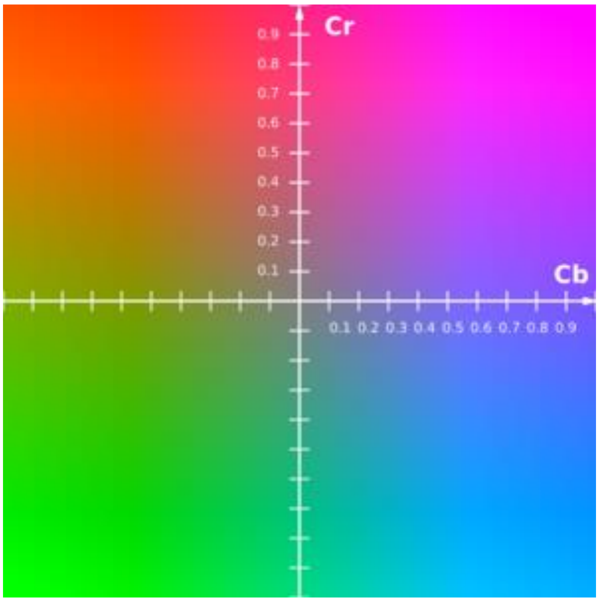

Fig. 3.3: YCbCr Colorspace

Rather than being a color space that is absolute, Y′CbCr is a versatile way of encoding the Red Green Blue information. The original color displayed however relies on the actual RGB values that were used to generate the signal. Hence, a value expressed as Y′CbCr can be predictable only if the chromacity comes from RGB color space. It has the following equation to RGB:

$$K_B = 0.114$$

$$K_R = 0.299$$

From the above constants and formulas, the following can be derived for ITU-R BT.601. Analog YPbPr from analog R'G'B' is derived as follows:

$$\begin{array}{rrrr} Y' = & 0.299 \cdot R' + & 0.587 \cdot G' + & 0.114 \cdot B' \\ P_B = - & 0.168736 \cdot R' - & 0.331264 \cdot G' + & 0.5 \cdot B' \\ P_R = & 0.5 \cdot R' - & 0.418688 \cdot G' - & 0.081312 \cdot B' \end{array}$$

Digital Y′CbCr (8 bits per sample) can be derived from analog R'G'B' as follows:

$$\begin{array}{rrrrr} Y' = & 16 + & (65.481 \cdot R' + & 128.553 \cdot G' + & 24.966 \cdot B') \\ C_B = & 128 + & (-37.797 \cdot R' - & 74.203 \cdot G' + & 112.0 \cdot B') \\ C_R = & 128 + & (112.0 \cdot R' - & 93.786 \cdot G' - & 18.214 \cdot B') \end{array}$$

The range that is received can differ starting from 16 to a maximum of 235; the lower that starts from 0 upto are known as footroom, while the values above 236 are known as the headroom. Alternatively, the Y′CbCr that has a digital form, is derived from digital R'dG'dB'd (8 bits per sample) with respect to the following equations:

$$\begin{aligned} Y' &= 16 + \frac{65.738 \cdot R'_D}{256} + \frac{129.057 \cdot G'_D}{256} + \frac{25.064 \cdot B'_D}{256} \\ C_B &= 128 + \frac{-37.945 \cdot R'_D}{256} - \frac{74.494 \cdot G'_D}{256} + \frac{112.439 \cdot B'_D}{256} \\ C_R &= 128 + \frac{112.439 \cdot R'_D}{256} - \frac{94.154 \cdot G'_D}{256} - \frac{18.285 \cdot B'_D}{256} \end{aligned}$$

Based on a threshold, they were converted to a binary image that contains noise.

3.2.2 Erosion and Dilation

The binary images may contain noise pixels. Two vital operations of morphological image processing are used: Erosion and Dilation.

3.2.2.1 Erosion

Erosion is used for noise removal from an image. Erosion shrinks image objects resulting in removing noise. This operator takes two inputs: the image on which the erosion should be applied, and a small set of coordinate points known as kernel. Eroding the binary image is necessary since the background may contain objects similar to the skin color.

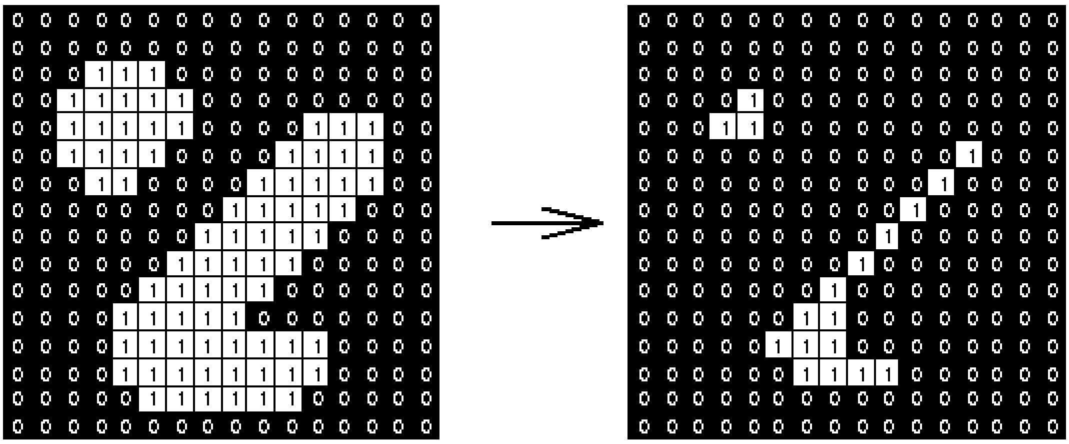

Fig. 3.4: Erosion

The fundamental idea in binary morphology is to skim through an image with a previously defined structure, determining decision on how it fits or vise versa, misses the shapes in the target photo. This "probe" that was done is known widely as the structuring element, which is also a form of image that is in Binary. Fig. 3.4 shows the basic erosion proves.

3.2.2.2 Dilation

Dilation is used to fill in the holes created by the erosion process. It adds pixels to image boundaries and It thickens an object controlled by the structuring element. It connects disconnected pixels as well.

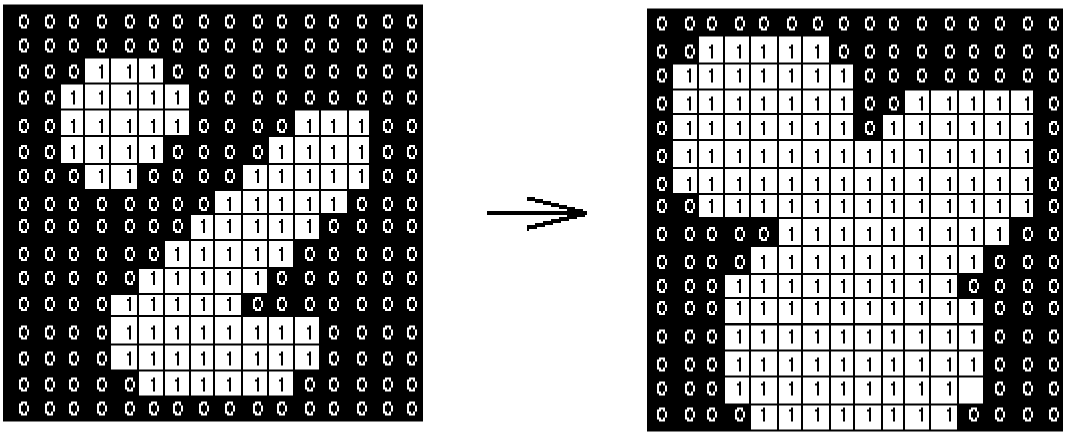

Fig. 3.5: Dilation

It is used for bridging gaps of character which may be caused by poor resolution. Fig. 3.5 shows the basic dilation process.

3.2.3 Detect Largest Blob

After the Dilation, the largest blob is considered as the hand. The morphological operation already removed the noise from the binary image. Hence, the largest blob is the obvious object representing the hand.

3.3 Feature Extraction

Fig. 3.6 shows the features that are identified from a sample hand gesture. Table 3.1 shows four random samples of the extracted features for four signs of the ASL. The five selected features for this research were extracted in the following manner:

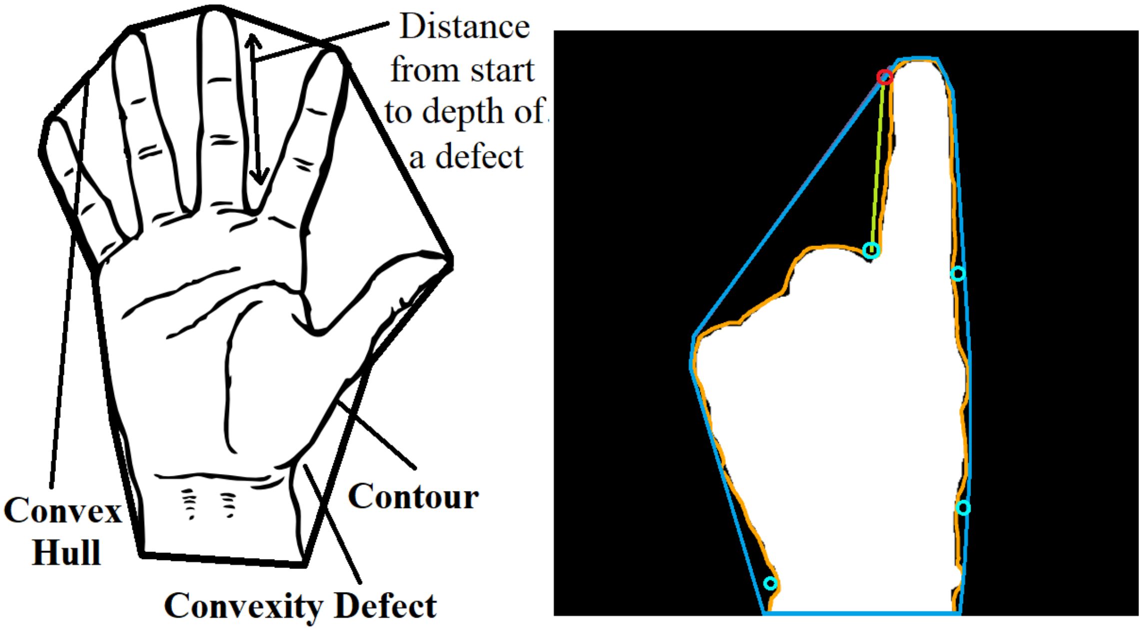

Fig. 3.6: Extracted Features

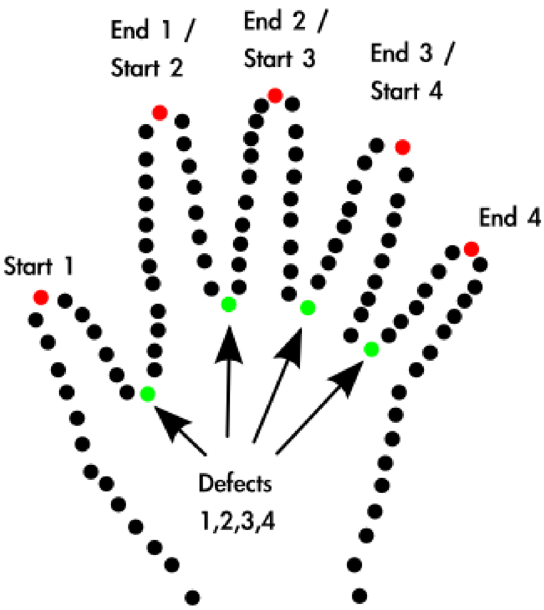

Fig. 3.7: Convexity Defects

3.3.1 Obtain the Contour Area

Contour tracing refers to a method that is applied to digital images in hope of extracting the blob boundary. The contour from the binary image is retrieved using the algorithm developed by Satoshi Suzuki [10]. The area of the contour is then computed using the Green's theorem [11]. Green's theorem shows the relationship between a double integral over the plane region and a line around a closed curve.

3.3.2 Obtain the area of the convex hull

The convex envelope or convex hull of a set of X amount of points in the Euclidean space refers to the minimal convex set that contains X, shown in Fig. 3.8. The point set of the contour is retrieved from the extracted contour. Using this, the convex hull of a two dimensional point set is determined using the algorithm developed by Sklansky' [12]. Followed by, the area of the convex hull is computed using the Green's theorem [11].

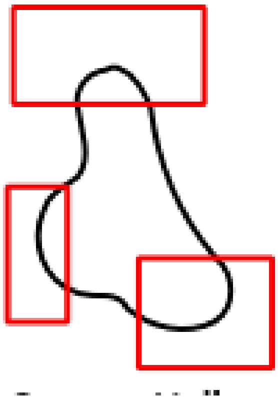

Fig. 3.8: Convex Hull

3.3.3 Obtain the number of convexity defects

Convexity defect is a depth in an object that is essentially a contour or a blob, segmented out from an image shown in Fig. 3.9. That ultimately means an area that may not be placed on the object but the outer boundary has that item. By using the contour and the convex hull, the convexity defects are found using the OpenCV image library [13].

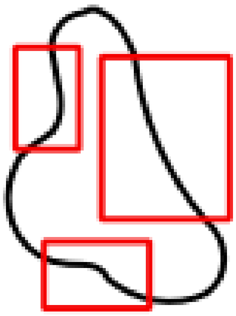

Fig. 3.9: Convexity Defects

3.3.4 Obtain the maximum depth of the defects

The maximum Euclidian distance between the point of the outer line where the defect begins and the longest distance from the convex hull farthest point inside the defect is measured.

3.3.5 Obtain the sum of the depths of the defects

The sum of the Euclidian distances between the point of the line where the convexity defect begins and the farthest from the convex hull point is measured.

3.3.6 Extracted Features from Binary Image

Table 3.1 shows a random set of data for four of the gestures. These data will later be used for matching gesture data using k-NN.

Table 3.1: Extracted Feature of a Sub set

Sign	Features				
	Contour Area	Convex Hull Area	Max Depth of Defect	Sum of depths of defects	Convexity Defect Count
B	69.51	73.77	10.14	36.92	7
D	63.24	75.97	15.05	48.33	6
Y	54.56	72.63	24.79	54.81	3
5	90.96	139.02	23.34	93.8	6

3.4 Edge Detection

A number of mathematical methods are combined in the process of an Edge Detection which has an aim of specifying points in a digital image at which the image brightness changes suddenly or has discontinuities in Fig. 3.10.

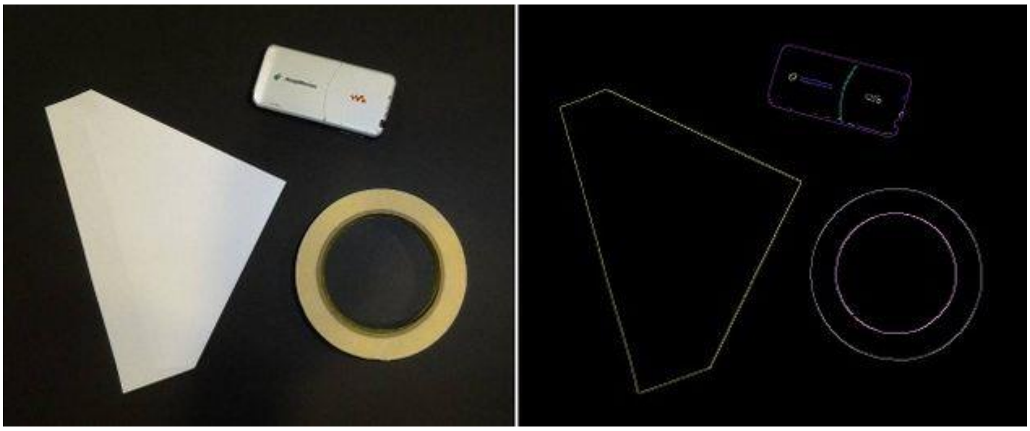

Fig. 3.10: Edge Detection

The edge of the contour was detected using the Canny Edge Detection algorithm. Fig. 3.10 illustrates the before and after images of this process. It proved to provide better results in marking the edges closely as possible to the actual edges of the contour. The edge detection process follows the given steps:

3.4.1 Apply Gaussian Filter

In a Gaussian filter, the image is blurred by a Gaussian function. It is a commonly used effect in graphical and visualization software, usually to discard image noise and to get a less detailed image. Typically, the input signal is modified by convolution with a Gaussian function using a Gaussian filter.

3.4.2 Gradient Calculation

This step identifies the intensity of the edge and direction by identifying the gradient of that image. Edges congruous with a change of pixels' sharpness. To detect a change of the intensity of the pixel, which corresponds to the edge, one of the ways is to apply filters that visualizes the change in intensity from bi directions: horizontal (x) and vertical (y).

3.4.3 Non-Maximum Suppression

Non-maximum suppression (NMS) is an important processing step that is performed in various applications that are backed by computer vision. In the area of detecting an object or a blob, it is used to transform a smooth response map that triggers a single bounding-box for every detected object. To thin out the edges, this step is essential.

3.4.4 Double Threshold

This step works for identifying 3 kinds of pixels: non-relevant, strong, and weak. The high threshold identifies the strong pixels where the low threshold is used to detect the irrelevant pixels.

3.4.5 Edge Tracking by Hysteresis

Among the weak edge pixels, some may be extracted from the true edge, others are purely noise. In this step, an edge that has pixels caused from true edges will be connected to a nearby pixeled edge that is strong.

Fig. 3.11: Edge Tracking

3.5 Training Phase

In this phase, training is performed using the extracted features for two sub systems separately.

3.5.1 Train contour features using k-NN

The k-nearest neighbors algorithm (k-NN) is a method widely utilized for classification. All computation is adjourned until a query is made. The training phase only includes storing the feature vectors and labels of the class. Hence, the extracted features along with their relevant classes are fed to the system.

3.5.2 Train the computed Bag of Visual Words using SVM

The commonly used bag-of-words model (BoW model), which is used for document classification, can be applied in the process of image classification by SVM. To train images for this model, these given steps are used.

3.5.2.1 Detect Feature and Extract Descriptors

These two steps can be completed by using a feature extractor algorithm. The Speeded up Robust Features (SURF) feature detector and descriptor was used to extract features from the training images. Since each image is considered as a document, and the words in the documents are parts of the images, the images needed to be divided. Breaking the image to a higher number caused poor performance since the images are the edge representation of the hands. 12 was determined as the number of words for the process.

3.5.2.2 Generate Visual Dictionary

For each image, frequency histogram is generated from the words and the frequency of the words in the image. These histograms are the Bag of Visual Words or codebook. Fig. 3.12 displays part of a prototype application used to extract feature using SURF.

3.5.2.3 Train

The extracted features are trained using SVM with the Gaussian kernel. Fig. 3.12 shows a prototype application used to extract feature.

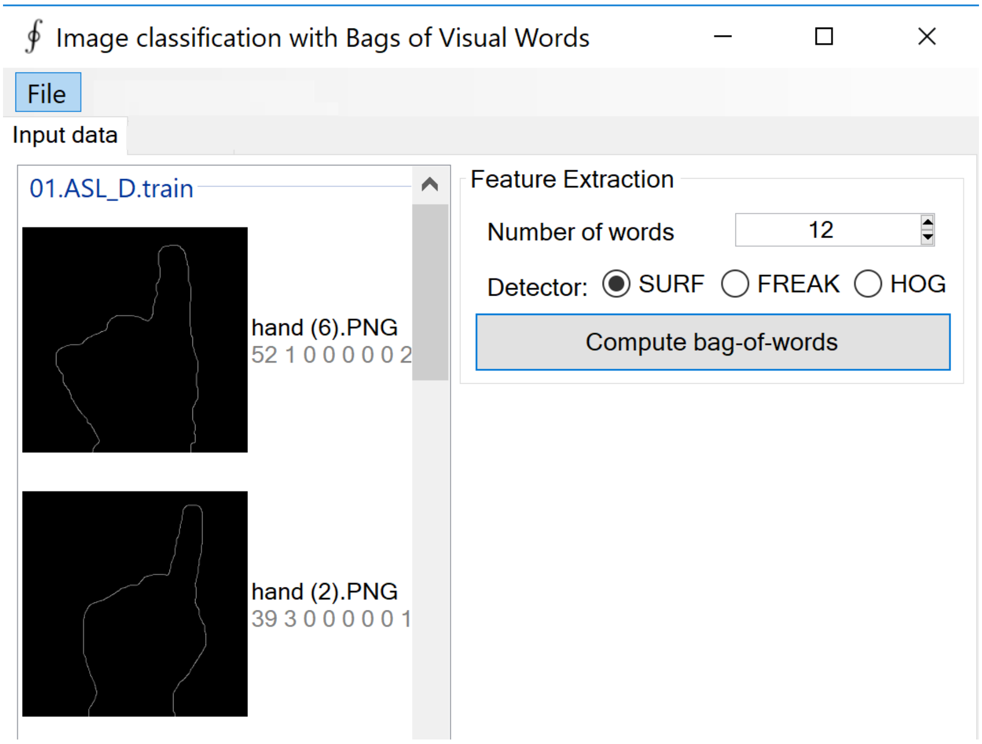

Fig. 3.12: Prototype for Training

3.6 Recognize Image

The recognition process is a combination of the classification received from the classifications done by k-NN, and the classification received from a Multi-class Support Vector Machine (SVM).

3.6.1 Classification using k-NN

Features from the target image is extracted and the k-nearest neighbors algorithm (k-NN) is used to classify the image. The Euclidian distance is used to get the distance for each feature. In mathematics, the Euclidean metric is the direct straight-line distance that exists between two points in the Euclidean space. With this distance, This Euclidean space then turns into a metric space. The classification of this process is stored to be used in later operations.

3.6.2 Classification using Multi-class Support Vector Machine

An input image is classified from the training data using SVM. In machine learning, support-vector machines (SVMs) are known as learning models which are supervised as well as are associated with algorithms that identifies data used for regression analysis including classification.

3.6.3 Logical AND operation

The logical AND operation is performed on the result of the classifications received from the two sub methods. The Boolean result is further processed along the following lines.

3.6.3.1 Result: true

If the result is 'True', both of the classifier returned the same result. Thus, the recognition is successful with high accuracy.

3.6.3.2 Result: false

If the result is 'False', then the classification of the Multi-class Support Vector Machine from the Bag of Visual Word model is assigned a degree of importance, ultimately given priority in producing the ultimate result. The study shows for a wide variety of test cases, the classification using the second sub system has a higher precision.

CHAPTER 4
RESULT AND ANALYSIS

Table 4.1 shows the sum of distances of the features for a target image from the training data. The target image was a gesture of the sign ‘D’. This is merely a subset revealed from a larger test case. The value k = 3 is used for this test case. Here, the nearest 3 neighbors match the proper class.

Table 4.2 displays a comparison of the classification using k-NN with various values of k. Gestures having outstanding difference in features were classified with 95% accuracy for the various values of k. Some of the neighbors were determined to be from the wrong class, but the ratio of classifying the target image data incorrectly is extremely low.

Table 4.1: Result Comparison

Sign	**Features**					
	Contour Area	**Convex Hull Area**	**Max depth of defects**	**Sum of depths of defects**	**Defect Count**	**Distance from target-D**
B	69.51	73.77	10.14	36.92	7	14.18
B	67.22	73.10	12.00	39.58	8	12.51
B	69.00	76.44	12.19	40.34	7	13.77
B	70.25	72.82	10.50	37.90	7	14.41
D	63.24	75.97	15.05	48.3	6	**12.32**
D	64.22	78.11	16.51	50.88	5	15.54
D	63.44	76.13	15.66	48.37	5	**12.38**
D	61.49	75.96	13.04	46.33	6	**10.46**
Y	54.56	72.63	24.79	54.81	3	18.46
Y	60.56	76.63	24.99	53.92	3	18.66
Y	56.36	72.11	22.72	54.78	3	17.29
Y	54.28	72.37	23.35	54.00	3	17.13
5	90.96	139.09	23.34	93.8	6	94.03
5	92.11	143.02	24.04	93.18	6	97.03
5	94.01	139.77	24.99	96.26	8	97.28
5	90.91	139.18	23.84	92.43	6	93.33

Table 4.2: Result Variation for kNN

Training Image	Results		
	K = 3	K = 5	K = 7
B	B, B, B	B, B,B,D,D	B,B, B, D, D, B, B
D	D, D, D	D, D, D, D, B	D, D, D,D, B, D, D
Y	Y, Y, Y	Y, Y,Y, Y, Y	Y, Y, Y, Y, Y, Y, B
5	5, 5, 5	5, 5, 5,5,5	5, 5, 5, 5, 5, 5, 5

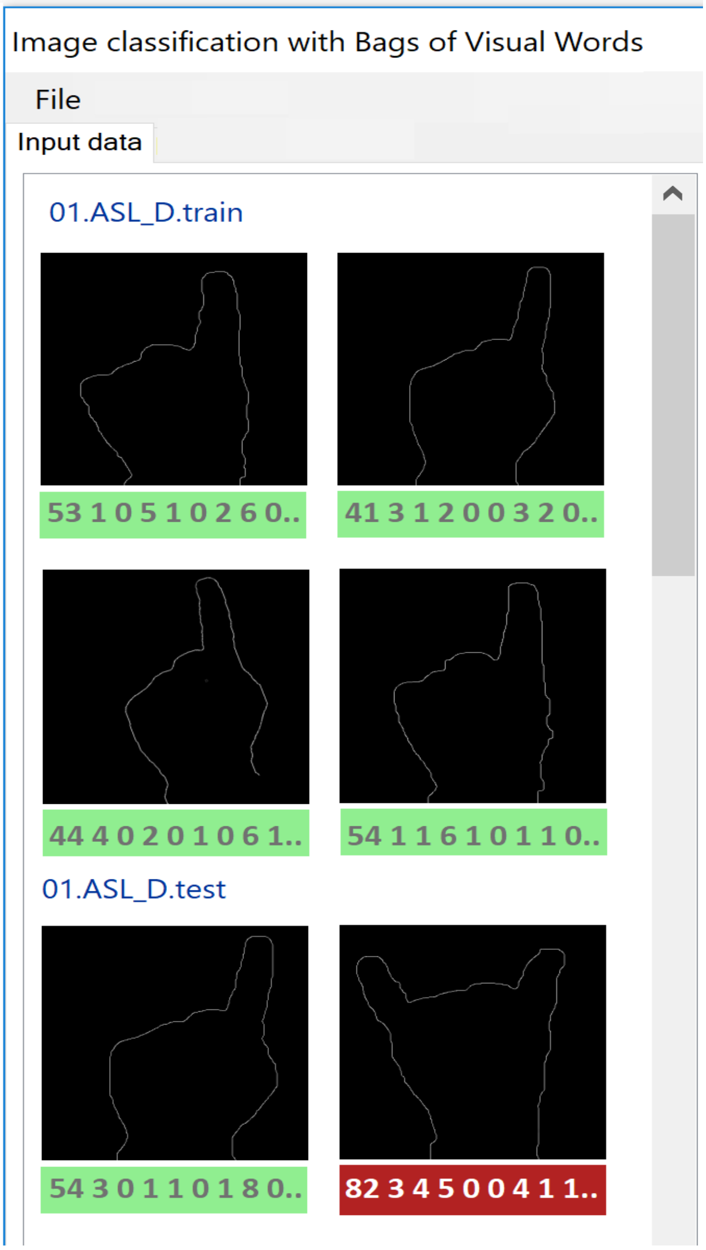

Fig. 4.1: Result of SVM on BoW

Fig. 4.1 shows the result that uses SVM for classification. Several test images of the ASL alphabet ‘D’ were correctly classified where the image of the ASL alphabet ‘Y’ was not classified as the ASL alphabet ‘D’.

After analyzing the results for both of the sub systems with varying parameters and training images, it can be concluded that the tendency to classify an image incorrectly is a bit low in the second sub system, that uses SVM. Hence, this system is provided with higher weight in ranking. If the logical AND operation turns out to be False, the result of the SVM classifier is preferred.

CHAPTER 5
CONCLUSION & FUTURE SCOPE

This thesis paper investigates on the static gestures of ASL alphabets and numerals. The training images and the test images both have sets of images with different rotation angles to replicate real world scenario. Both of the classifications produce satisfactory results. The parameters (the value of k for k-NN and the number of words for computing Bag of Visual Words) can be tuned for larger training data set. The lighting condition may cause the test images to be poorly captured, resulting in an incorrect shape identification of a gesture. The prospects are: improving performance under low light condition and producing a cost-effective mobile application to detect sign language gestures. Since, the base is finely defined in this proposal, a mobile application will be a perfect representative of the system proving it to be quite handy as well as useful in many days to day life events.

REFERENCES

1 R Syed Atif Mehdi, Yasir Niaz Khan, "Sign language recognition using sensor gloves", Neural Information Processing, 2002. ICONIP '02.

2 Kunal Purohit, Prof. Kailash Patidar, Mr. Rishi Singh Kushwah , "A Wearable Hand Gloves Gesture Detection basedon Flex Sensors for disabled People", IJRASET.

3 Lei Gu, Xiaoyang Yuan and Takeshi Ikenaga, "Hand Gesture Interface Based on Improved Adaptive Hand Area Detection and Contour Signature", IEEE International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS 2012) November 4-7, 2012.

4 Joyeeta Singha and Karen Das, "Indian Sign Language Recognition Using Eigen Value Weighted Euclidean Distance Based Classification Technique", International Journal of Advanced Computer Science and Applications, March 2013

5 Lae-Kyoung Lee, Su-Yong An, and Se-Young Oh, "Robust Fingertip Extraction with Improved SkinColor Segmentation for Finger Gesture Recognitionin Human-Robot Interaction", WCCI 2012 IEEE World Congress on Computational IntelligenceJune, 10-15, 2012.

6 Francesca Gasparini, Raimondo Schettini, "Skin segmentation using multiple thresholding", Proceedings of SPIE - The International Society for Optical Engineering, January 2006.

7 Son Lam Phung, Abdesselam Bouzerdoum, and Douglas Chai, "Skin Segmentation using Color and Edge Information", Signal Processing and Its Applications, 2003.

8 Julka, Abhinandan & Bhargava, Sandeep. (2013). A Static Hand Gesture Recognition Based on Local Contour Sequence. International Journal of Computer Science and Software Engineering.

9 Jun Li and Sheng Ding, "A Research on Improved Canny Edge Detection Algorithm", International Conference on Applied Informatics and Communication, 2011.

10 Satoshi Suzuki, "Topological Structural Analysis of Digitized BInary Images by Border Following. A Research on Improved Canny Edge Detection Algorithm", Computer Vision, Graphics, and Image Processing 30 , 1985.

11 Green's theorem,[online] https://en.wikipedia.org/wiki/Green's_theorem, accessed on 18 June, 2019.

12 Jack Sklansky, Finding the convex hull of a simple polygon, Pattern Recognition Letters, 1982.

13 OpenCV, [online] https://opencv.org/about/, accessed on 18 June, 2019.

14 Fereshteh Falah Chamasemani and Yashwant Prasad Singh, "Multi-class Support Vector Machine (SVM) Classifiers - An Application in Hypothyroid Detection and Classification", 2011 Sixth International Conference on Bio-Inspired Computing: Theories and Applications.

15 Daniela Ramirez-Giraldo, Santiago Molina-Giraldo, Andres M. Alvarez-Meza, Genaro Daza-Santacoloma, and German Castellanos-Dominguez, "Kernel Based Hand Gesture Recognition using Kinect Sencor", STSIVA 2012.

16 Ashish S. Nikam and Aarti G. Ambekar, "Sign Language Recognition Using Image Based Hand Gesture Recognition Technique", Online International Conference on Green Engineering and Technologies (IC-GET), 2016.

Turnitin Originality Report

Processed on: 29-Aug-2019 13:23 +06

ID: 1164738514 Word Count: 5756 Submitted: 1 183-25-705 By Jayed Talukder	Similarity Index 16%	**Similarity by Source** Internet Sources: 9% Publications: 5% Student Papers: 14%

2% match (student papers from 04-Apr-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 940891790

2% match (Internet from 17-Jan-2013)
http://en.wikipedia.org/wiki/YCbCr

1% match (student papers from 08-Jun-2018)
Submitted to Neshaminy School District on 2018-06-08

1% match (Internet from 02-Mar-2019)
http://dspace.uiu.ac.bd/bitstream/handle/52243/118/Nashtarin%20Nur%28ID_012153004%29.pdf?isAllowed=y&sequence=1

1% match (publications)
"Computer Vision -- ACCV 2014", Springer Science and Business Media LLC, 2015

1% match (student papers from 07-Apr-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 942523876

1% match (Internet from 14-Jul-2012)
http://www.tak-chung.com/projects/fyp/report.htm

< 1% match (publications)
Mehdi, S.A., and Y.N. Khan. "Sign language recognition using sensor gloves", Proceedings of the 9th International Conference on Neural Information Processing 2002 ICONIP 02, 2002.

< 1% match (student papers from 10-Aug-2015)
Submitted to Universiti Sains Malaysia on 2015-08-10

< 1% match (publications)
Ajin T Pullan, Arun Sasi, Divya Krishnan, Irina Merin Baby, Mithun Krishnan, Dhanaraj K. J. "High Resolution Touch Screen Module", 2018 IEEE Recent Advances in Intelligent Computational Systems (RAICS), 2018

< 1% match (student papers from 20-May-2017)
Submitted to Rajarambapu Institute of Technology on 2017-05-20

< 1% match (publications)
Nikita P. Nagori, Vandana Malode. "Communication Interface for Deaf-Mute People using Microsoft Kinect", 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), 2016

http://cs229.stanford.edu/proj2016/report/ZhuWang-Data-Driven%20Earthquake%20Location%20Method-report.pdf

< 1% match (publications)
Lei Gu, Xiaoyang Yuan, Takeshi Ikenaga. "Hand gesture interface based on improved adaptive hand area detection and contour signature", 2012 International Symposium on Intelligent Signal Processing and Communications Systems, 2012

< 1% match (student papers from 14-Apr-2010)
Submitted to The Hong Kong Polytechnic University on 2010-04-14

< 1% match (publications)
"Artificial Intelligence and Computer Vision", Springer Science and Business Media LLC, 2017

< 1% match (student papers from 07-Apr-2018)
Class: Article 2018
Assignment: Journal Article
Paper ID: 942529588

< 1% match (student papers from 13-Dec-2018)
Submitted to Segi University College on 2018-12-13

< 1% match (Internet from 02-Apr-2019)
http://aircconline.com/ijcsit/V8N5/8516ijcsit01.pdf

< 1% match (student papers from 30-Apr-2010)
Submitted to Brunel University on 2010-04-30

< 1% match (Internet from 01-Dec-2016)
http://www.studymode.com/essays/Export-Procedure-46196223.html

< 1% match (Internet from 07-Sep-2017)
http://ro.ecu.edu.au/cgi/viewcontent.cgi?article=1609&context=theses_hons

< 1% match (student papers from 04-Nov-2008)
Submitted to University of Bristol on 2008-11-04

< 1% match (student papers from 12-Dec-2014)
Submitted to Institute of Technology, Nirma University on 2014-12-12

< 1% match (student papers from 28-Apr-2013)
Submitted to University of Westminster on 2013-04-28

< 1% match (Internet from 08-Jul-2019)
https://www.sv-jme.eu/article/classification-of-malignancy-suspicious-lesions-in-autofluorescence-bronchoscopy/

< 1% match (publications)
Łukasz Sztyber. "Chapter 41 Lane Finding for Autonomous Driving", Springer Nature, 2020

< 1% match (Internet from 06-Sep-2018)
http://dspace.daffodilvarsity.edu.bd:8080/bitstream/handle/20.500.11948/2861/Farming_information_Support_system.docx.p isAllowed=y&sequence=1