An American Sign Language Recognition System using Bounding Box and Palm FEATURES Extraction Techniques

S. Shivashankara and S. Srinath

Abstract--- The sign language is absolutely an ocular interaction linguistic over and done with its built-in grammar, be nothing like basically from that of spoken languages. This research paper presents, an inventive context, whose key aim is to achieve the transmutation of 24 static gestures of American Sign Language alphabets into human or machine identifiable manuscript of English language. The gestures sets considered for cognition and recognition process are purely invariant to location, Background, Background color, illumination, angle, distance, time, and also camera resolution in nature. The gesture recognition process is carried out after clear segmentation and preprocessing stages. As an outcome, this paper yields an average recognition rate of 98.21%, which is an outstanding accuracy comparing to state of art techniques.

Keywords--- American Sign Language, Bounding Box Technique, Canny Edge Detector, CIE Color Model, Gesture Recognition.

I. INTRODUCTION

The Sign Language (SL) is fashioned using the explicit movements of right or left hand, signs made by actions of finger, and facial expressions of a person to convey their views and thoughts of deaf and dumb society to the non-deaf and dumb society. The SL may not be obviously understand by most of the normal people. So that, there exists an immense communication opening between the deaf and non-deaf communities. With the backing of advanced technology, as personal interpreters for deaf and dumb people are exceptionally challenging to patronage deaf people in their day-to-day endeavors. By the improvement in science and technology, one can consider the developing an optimal framework which interprets gestures into humanoid or machine understandable text, smoothing typical audible range people to understand and converse with hearing impairment society [1]. About more than 120 distinct sign languages are used by deaf community in the earth. American Sign Language (ASL) is an accepted and natural language instigated in early 19th Century at American School for the Deaf, USA with the help of Old French Sign Language, several village sign languages, and home sign systems.

ASL serves as the predominant SL of Deaf Communities. It is the highest and widely using SL and 4th highest usable linguistic in North America, ASL is fast perception and attractiveness as an L2 in scholastic courses [2]. ASL is also using in several countries where English is the primary language for their communication. Over and above 20 nations' SL is hereditary from ASL. Almost 20 Lakh hearing impaired people of USA and Canada are using ASL as their primary basis of communication [3]. Its foundation, existing conditions, prospect hopes, and global impact are quite amazing and eye-opening [4]. Beside through ASL Alphabets and Numbers, more than thousands of gesture signs of hand and face are exist to sign the various English words[5].

Figure 1 depicts the set of 36 gesture of english alphabets (Letter A – Z are used to spell out the English words) and numbers (0 to 9).

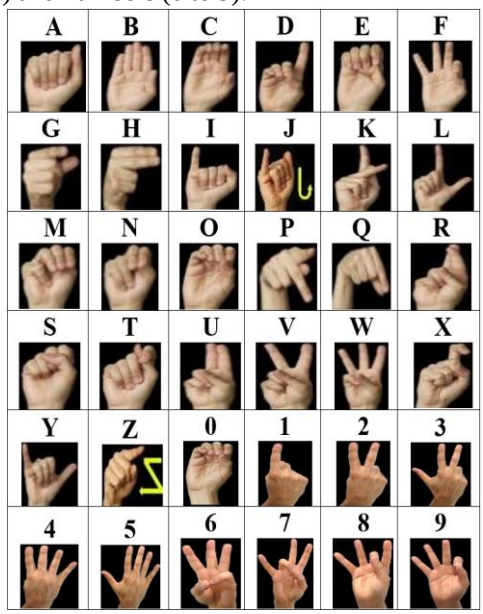

Fig. 1: Set of Gestures of ASL Alphabets and Numbers

II. DATA COLLECTION

The collection of data is a crucial role of research work in any kinds of research fields encompassing science, social sciences, technological arena, human science, and professional as well. In the process of collection of data, procedures might differ by domains, the eminence on assuring accurate and genuine assemblage remains the undistinguishable. Regardless of domains of the study, unambiguous collection of data is a necessary platform for improving the reliability of the investiga

Manuscript received February 01, 2019

S. Shivashankara, Research Scholar, Department of Computer Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, India. (e-mail:Shivashankar.research@gmail.com)

S. Srinath, Department of Computer Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, India.

tion. An in effect collections of data are constrained both in mass as well as class. A traditional development of collecting data is an important task as it endorses that the data collected are together positive as well as open. The succeeding decisions stand depending on arguments indicated in the products are valid [6].

In proposed technique, the static hand gestures and video gestures of ASL were referred and explored from the websites of ASL University (ASLU) [7][8][9][10] for ASL Alphabets gestures, Numbers gestures, and as well as Video gestures. Forming enormous marked ASL gestures databases to train and test purpose is time consuming and tedious task. Though, in this ASL recognition research work, an effort has been placed in formation of huge quantity of ASL alphabet Gestures' sets of our own which is invariant to Location (indoor and outdoor), Background (plain and complex, uniform and non-uniform), Background color (plain single color and mixed multicolor), Illumination (light: natural and artificial), Angle (angle of gestures captured), Distance (distance from gestures captured), Time (day and night), and also Mobile camera (resolution of the camera) in nature. Gestures captured for cognition and recognition process are from different angles, which are 60^0, 75^0, 80^0, and 90^0 to both left and right side of the signer and from various distances of 2 meters, 3 meters, and 4 meters.

III. SYSTEM ARCHITECTURE

The system architecture of the ASL recognition system is illustrated in Figure 2.

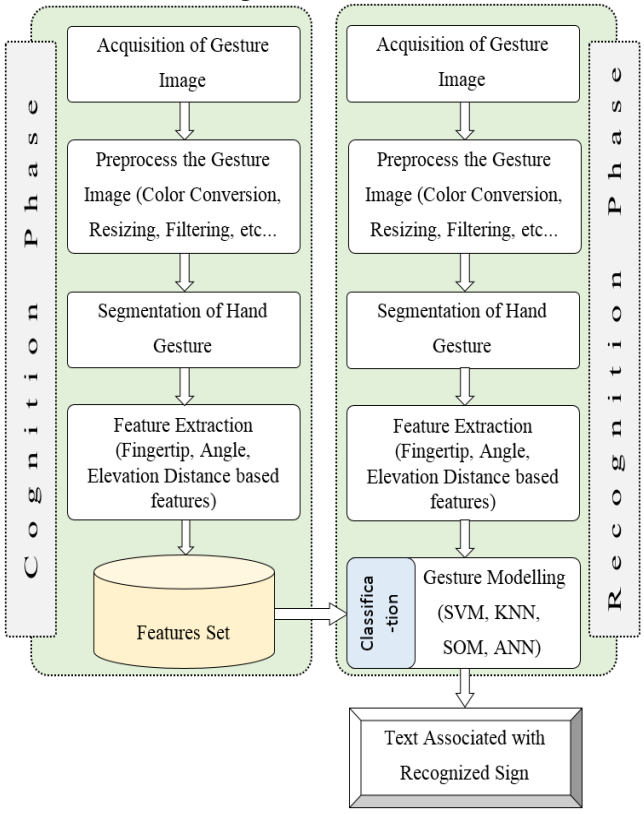

Fig. 2: Cognition and Recognition Process of ASL

In the cognition phase of Figure 2, primarily, signed gestures are captured by the digital camera of dissimilar resolution from various distances and different angles in plain uniform and complex non-uniform backgrounds with distinct illumination of natural (day) and artificial (night) lighting conditions. After the signed gestures captured, the preprocessing operation like color conversion, resizing the gesture, filtration for noise removal, and etc... are carried out. Later, an imperative task, the segmentation of hand gesture is taken place for further operations. Once the successful hand gesture segmentation is carried out, then the various features of hand gesture are extracted with respect to several factors like gesture size, roundness, centroid, mean, entropy, Local Phase Quantization (LPQ), HOG (Histogram of Oriented Gradients) feature vectors, Zernike moments for amplitude value and angle of moment value, distance transform features, and Gray Level Co-occurrence Matrix (GLCM). All the extracted consolidated features of the gestures are warehoused in feature set (knowledge base) for promote utilization in the recognition phase.

In the recognition phase, the initial four steps of processing are alike as in the cognition phase. In the fifth step of the recognition phase, the Gesture Modelling step, which does the matching of sign and gesture recognition progression is carrying out by comparing the extracted consolidated features of feature set with the extracted consolidated features of the gestures in the recognition phase using machine learning algorithms like Support Vector Machine (SVM), K-Nearest Neighbor (k-NN), Self-Organizing Map (SOM), Artificial Neural Networks (ANN), and etc. Once the successful gesture modelling is completed, then the text associated with the recognized gesture will be put on show.

IV. RELATED WORK

A very less amount of contributions have been endeavored in hand gestures recognition using various segmentation and non-segmentation algorithms with limited recognition rate and time.

ASL Recognition System using Back-Propagation Neural Network (BPNN) Algorithm for 6 hand postures (a, b, c, point, five, and v) of 10 persons in the local system or the frame captured from webcam camera is proposed [11] and obtained 70% and 85% accuracy from Raw Features Classifiers and Histogram Features Classifiers respectively. Recognition of static hand gestures corresponding to digits (0 – 9) is reported in [12] using Motion Vision Based Skin Color Segmentation (MVBSCS) Technique with Radial Basis Function Neural Network (RBFNN) as a classifier. Recognition system has conducted in uniform lighting condition and obtained 93.5% of recognition rate. In 2017 [13], the 24 ASL alphabet recognition system implemented in hardware using a Field-Programmable Gate Array (FPGA) with Neuromorphic Camera and Artificial Neural Networks (ANN). The recognition accuracy of 79.58% is obtained for images with an angular variation of approximately $\pm 8^0$. The ASL Alphabet recognition system is implemented using Principle Component Analysis and Euclidian Distance Classifier (PCA & EDC) [14] and achieved

98% recognition rate under good illumination and plain light colored background for limited ASL Alphabets. The finger spelled ASL gesture recognition system is presented in [15] using shape based algorithm and Euclidian Distance Classifier. Images were taken in black or dark background. With the 180 point descriptor, an accuracy rate of 79.9% is obtained in real time, for rotation, luminance, and translation invariance gestures in less than half second.

The ASL Hand Gestures Recognition (HGR) System of Alphabets is proposed in [16] using Adaptive Network-based Fuzzy Inference System (ANFIS) and K-Nearest Neighbor (K-NN) Algorithm. An accuracy of 80.77% is achieved when the number of epochs was 10 in shorter recognition time. An occluded and non-occluded static hand gesture recognition system has been designed using Non-negative Matrix Factorization (NMF) and Compressive Sensing (CS) Theory. Here, 2 databases are used for gestures. Database 1 contains 10 static hand gestures of ASL Alphabets (A, B, C, D, G, H, I, L, V, Y) of 3 different backgrounds of uniform light, uniform dark, and complex [17]. Database 2 consists of 10 ASL Numbers (0 – 9). All the hand gestures of database 1 and database 2 are cropped to 80 X 80 Size and 75 X 120 Size respectively. After cropping those gestures are sent as input to experiment. An average recognition rate (Protocol P1 and P2 achieves 96.67% and 98.13% of accuracy respectively) achieved is 97.4%. The Sign Language Recognition (SLR) System to recognize the sign languages images which are obtained from ASL Lexicon Video dataset [18] using Gabor Transformation and ANFIS Classifier and achieved 98% recognition accuracy. An optimal approach [19] for recognizing ASL has been developed to recognize 24 static gestures of ASL Alphabets (J and Z not included) and 10 static gestures of ASL Numbers (0-10) and achieved 93.05% and 95% recognition rate respectively.

V. THE PROPOSED SYSTEM DESIGN

5.1. Face detection

It is a process of detecting the face by identifying different parts of human faces like eyes, nose, mouth and etc. The earliest system developed for face detection applications was viola and jones algorithm, broadly using in digital cameras, smart phones and also many other such places. Though, researchers bring into being that the system workings fine merely to nearby front faces in usual lighting condition with no occlusion. It has a major constraint for detection of faces which are invariant to pose, angle, and distance with different illumination conditions considered.

5.1.1. Bounding box technique

One among the most significant tasks in computer vision and image processing area is object identification / detection. It is well-defined as identifying the precise localities of objects in an image. In Digital Image Processing (DIP), the Bounding Box (BB) is purely the coordinates of the rectangular / circle / line / polygon boundary, which completely surrounds a digital image or some particular probable Region Of Interest (ROI) or multiple probable regions of interest of an image.

In common, every single feature recognition / detection algos returns the ROI in the manner of row and column pixel coordinates and also the height and width. By using the initial coordinates alongside width and height (in pixels), usually code the algorithm to draw bounding boxes. Drawing a BB from place to place, a proposed segmentation focus has turn out to be both a widely held user interface and a communal production for learning-driven detection algorithms.

A user stipulated BB of an object is an uncomplicated and admired communication standard, well-organized by means of numerous current interactive image segmentation frameworks. Such frameworks incline to adventure the stipulated BB simply to eliminate its surface from the importance and on occasion for initializing the energy minimization [20].

In this proposed system design, BB technique detects the various parts (regions) of the human face such as left eye, right eye, mouth, and nose. Based on the various parts of the face detected, the face region will be detected. While detecting face parts and face using BB technique, the ROI to be detected and stored in the various BB technique parameters such as *bbox*, *bbX*, *faces*, and *bbfaces*.

The 1st parameter, *bbox*, is used to detect the face parts such as left eye, right eye, mouth, and nose. In detecting face parts, minimum size required for left eye, right eye, mouth, and nose are (12 18), (12 18), (15 25), and (15 18) respectively. Figure 3 shows the BB shapes to detect the various above said face parts. Here, the 1st and 2nd rectangular boxes are for left and right eye and also the 3rd and 4th boxes are for mouth and nose respectively.

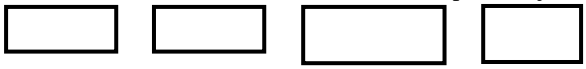

Fig. 3: Bounding boxes for face parts detection

Here, the face parts are detected by 4 different regions using the following equations.

For left eye region detection:

$$\text{region}=\left[1,\left(\frac{\text{stdsize}*2}{3}\right);1,\left(\frac{\text{stdsize}*2}{3}\right)\right] \quad (1)$$

For right eye region detection:

$$\text{region}=\left[\left(\frac{\text{stdsize}}{3}\right),\text{stdsize};1,\left(\frac{\text{stdsize}*2}{3}\right)\right] \quad (2)$$

For Mouth region detection:

$$\text{region}=\left[1,\text{stdsize};\left(\frac{\text{stdsize}}{3}\right),\text{stdsize}\right] \quad (3)$$

For Nose region detection:

$$\text{region}=\left[\left(\frac{\text{stdsize}}{5}\right),\left(\frac{\text{stdsize}*4}{5}\right);\left(\frac{\text{stdsize}}{3}\right),\text{stdsize}\right] \quad (4)$$

Where, *stdsize* = standard size of the image = 176.

In all the 4 equations, the regions are detected by con

sidering from which column location (starting column) to which location (ending column), and also from which row location (starting row) to which location (ending row) of the image. By finding the regions using equation (1) to (4), the detected face parts of the image shown as given in Figure 4. From the various regions of detected face parts, the face is detected using BB technique.

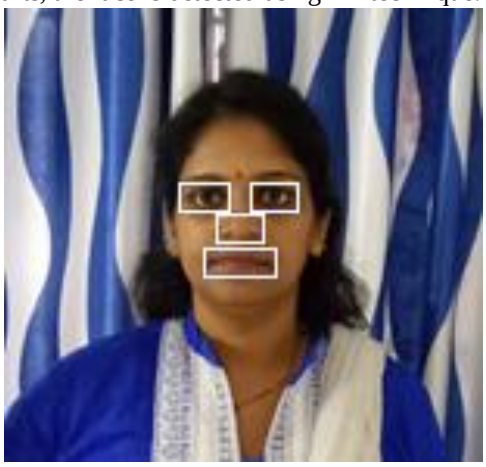

Fig. 4: Detected face parts from the image

The 2nd parameter, *bbX*, an image with detected face of an image is surrounded with the box. The box might be drawn using rectangle, line, circle, and polygon as mentioned in the code. In proposed technique, the BB shape circle is used to surround the face by using the shape property and the location. The circle property contains 3-element row vector [*x*, *y*, and *r*] where, *x* and *y* are coordinates for the center of the circle and the r is the radius of the circle, which must be greater than 0. The detected face surrounded with the circle as shown in Figure 5.

The 3rd parameter, *faces*, the detected face of an image is stored as cell array for further process. At last, the parameter, *bbfaces*, the detected face of an image with box is stored as cell array for masking the detected face.

Fig. 5: Detected face surrounded with circle

*5.1.2. The CIE 1979 L*a*b* color model*

The CIE (Commission Internationale Eclairage) 1976 *L*a*b** Color Model (or CIE LAB), which specifies 2 color spaces; for use with self-luminous colors and surface colors. CIE LAB permits the description of color insights in the form of 3-D space, which is shown in Figure 6.

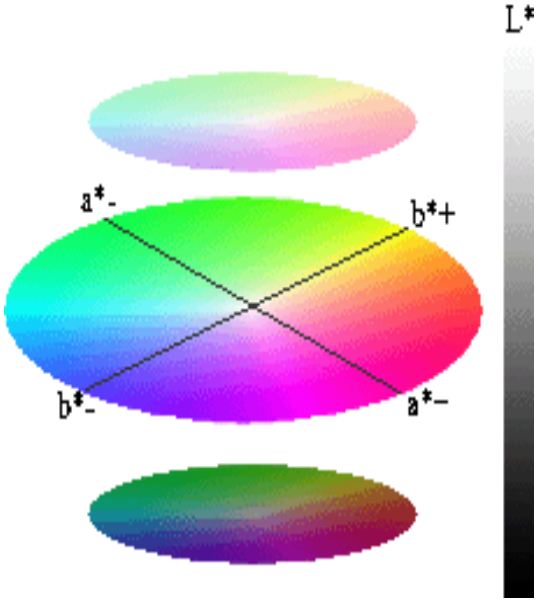

Fig. 6: CIE LAB description in 3-D space.

The *L** axis is identified as lightness and ranges from 0 to 100 (say black to white). The other 2 co-ordinates *a** and *b** signify redness-greenness and yellowness-blueness respectively. Samples for which *a* = b* = 0* are achromatic and thus the *L** axis signifies the achromatic scale of grays from black to white.

The measures *L**, *a**, and *b** are acquired by the tri-stimulus values rendering to the subsequent transformations:

$$L^*=116.\sqrt[3]{\left(\frac{Y}{Y_n}\right)}-16 \qquad (5)$$

for $Y/Y_n > 0.008856$

$$L^*=116.(7.787.\frac{Y}{Y_n}+0.138)+16 \qquad (6)$$

for $Y/Y_n <= 0.008856$

$$a^*=500.\left[\sqrt[3]{\left(\frac{X}{X_n}\right)}-\sqrt[3]{\left(\frac{Y}{Y_n}\right)}\right] \qquad (7)$$

for $X/X_n > 0.008856$)

$$a^*=500.\left[\left(7.787.\frac{X}{X_n}+0.138\right)-\left(7.787.\frac{Y}{Y_n}+0.138\right)\right] \qquad (8)$$

for $X/X_n <= 0.008856$

$$b^*=200.\left[\sqrt[3]{\left(\frac{Y}{Y_n}\right)}-\sqrt[3]{\left(\frac{Z}{Z_n}\right)}\right] \qquad (9)$$

for $Z/Z_n > 0.008856$

$$b^*=200.\left[\left(7.787.\frac{Y}{Y_n}+0.138\right)-\left(7.787.\frac{Z}{Z_n}+0.138\right)\right] \qquad (10)$$

for $Z/Z_n <= 0.008856$

Where, X_n, Y_n, and Z_n are the tri-stimulus values of the illuminant (perfect reflecting diffuser), which is used for the computation of *X*, *Y*, and *Z* of the sample.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

5.1.3. The Otsu thresholding

Converting a greyscale image to monochrome is a common image processing task. The Otsu thresholding [21] works out a universal threshold, level, which is used for translate an intensity image to a binary image. The Otsu's method selects the threshold to minimalize the intraclass variance of the binary pixels, defined as a weighted sum of variances of the 2 classes.

$$\sigma^2_{\omega}(t)=\omega_0(t)\sigma^2_0(t)+\omega_1(t)\sigma^2_1(t) \quad (11)$$

Here, the weights ω_0 and ω_1 are the probabilities of the 2 classes detached using threshold *t*, and also σ^2_0 and σ^2_1 are the variances of these 2 classes.

The class probability $\omega_{0,1}(t)$ is calculated by the bins of the histogram:

$$\omega_0(t)=\sum_{i=0}^{t-1} p(i) \quad (12)$$

$$\omega_1(t)=\sum_{i=t}^{L-1} p(i) \quad (13)$$

Otsu displays that minimalizing the intra-class variance is as similar as maximizing inter-class variance:

$$\sigma^2_b(t)=\sigma^2-\sigma^2_{\omega}(t)=\omega_0(\mu_0-\mu_T)^2+\omega_1(\mu_1-\mu_T)^2 \quad (14)$$

$$=\omega_0(t)+\omega_1(t)[\mu_0(t)-\mu_1(t)]^2 \quad (15)$$

Which is expressed in terms of class probabilities ω and class means μ.

While the class mean $\mu_{0,1,T}(t)$ is:

$$\mu_0(t)=\frac{\sum_{i=0}^{t-1} ip(i)}{\omega_0(t)} \quad (16)$$

$$\mu_1(t)=\frac{\sum_{i=t}^{L-1} ip(i)}{\omega_1(t)} \quad (17)$$

$$\mu_T=\sum_{i=0}^{L-1} ip(i) \quad (18)$$

The relations shown in equation (19) and (20) can be easily verified:

$$\omega_0\mu_0+\omega_1\mu_1=\mu_T \quad (19)$$

$$\omega_0+\omega_1=1 \quad (20)$$

Both the class probabilities as well as class means are calculated reiteratively. This yields an effective thresholding.

5.1.4. Edge detection

It is a DIP technique to find out the boundaries of the objects inside images. Edge detection is carried out by identifying cutoffs based on the brightness. Edge detection is used for segmentation, data extraction, obtain shape information and many such applications in DIP, computer vision, and machine vision. The usual edge detection algorithms in DIP are Sobel, Canny, Prewitt, Roberts, and Fuzzy Logic methods.

In proposed technique, used Canny Edge Detector (CED), also called optimal detector, developed by John F Canny [22] in 1986. The CED is a most powerful, widely used, and uses a multi-stage process to detect an extensive range of edges of the images. It extracts suitable structural info from various visualization objects and affectedly decrease the quantity of data to be treated. Canny also provides Low error rate, Good localization, and Minimal response in detection. The CED process fall in 5 different steps as follows.

1. Apply Gaussian filter to smooth out the image in order to take out the noise. The Gaussian filter kernel of size *(2k+1) X (2k+1)* is given in equation (21).

$$g(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right) \quad (21)$$

Where, *g(x, y)* is the Gaussian operator, σ is the standard deviation, and finally *x* and *y* are the coordinates.

2. Find out the intensity gradients of an image. Smoothened image is filtered with a kernel in both horizontal and vertical direction to get 1st derivative in horizontal and vertical directions, (G_x) and (G_y) respectively. By these 2 images, edge gradient and direction for each pixel is calculated as in equation (22) and (23)

$$G=\sqrt{G^2_x+G^2_y} \quad (22)$$

$$\theta=\tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (23)$$

3. Apply non-maximum suppression to acquire rid of spurious reply for the detection of edge. Once the gradient magnitude and direction is obtained, an image scan is performed to eliminate unnecessary pixels that may not establish the edge. Aimed at this, at each pixel, the pixel is tested if it is a local maximum in its neighborhood in the direction of gradient.
4. Apply double threshold to define possible edges. After the non-maximum suppression, left over pixels offer additional exact depiction of real edges. Though, few edge pixels continue which are affected by noise and color variation. It is necessary for filtering pixels of edge by weak gradient value and reserve edge pixels by high gradient value, for these reactions. This is proficient through choosing high and low threshold values. If the gradient value is greater than the high threshold value, then it is noticeable as strong edge pixel. If the gradient value is lesser than the high threshold value and greater than the low threshold value, then marked as weak edge pixel. Suppose the value of the edge pixel is lesser than the low threshold value, then it will be suppressed. The 2-threshold values are empirically defined and definition is based on the contents of the input given.
5. Trace the edges through hysteresis. Firm up the finding of edges by defeating all other edges which are weak and disconnected to solid edges. The minor pixel noises are also eliminated on the hypothesis, edges are extensive lines. Thus, at last the strong edges are obtained in the image.

5.2. Hand / Palm features extraction

Following are the hand / palm features extraction techniques used is proposed technique.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

5.2.1. Centroid

It obtains the center position of an object based on how centroid needs to be computed. There are many possible methods: Calculate the features' central XY coordinates, use the features to point tool.

5.2.2. Boundary

It traces the boundaries of an object / image. Boundary features are the structures that separate one property from another.

5.2.3. LPQ (Local Phase Quantization)

This method is built on blur invariance property of the Fourier phase spectrum used to extract the texture features of an object. It usages the local phase info pull out via the 2-D Discreate Fourier Transform (DFT) or, specifically, Short Term Fourier transform (STFT) calculated above a rectangular M-by-M neighborhood N_x at all the pixel location x of an image $f(x)$ is mark out by equation (24)

$$F(u,x)=\sum_{y\epsilon N_x} f(x-y)e^{-j2\pi u^T y}=w_u{}^T f_x \qquad (24)$$

Where, w_u is a base vector of the 2-D DFT at the frequency u, and f_x is a one more vector comprising all the M^2 samples of an image from N_x.

In LPQ, only 4 complex co-efficients are reflected, conforming to 2-D frequencies u_1=[a,0]T, u_2=[0,a]T, u_3=[a, a]T, and u_4=[a,−a]T, where 'a' is a scalar frequency under 1st zero traversing of H(u).

5.2.4. HOG (Histogram of Oriented Gradients) feature vector

It is a feature descriptor, used to detect the object, recognize the object, and also to extract the texture feature of an object. This method aggregates the incidences of gradient orientation in localized portion of an image. HOG descriptors can be used to recognize the object via offering them as features to machine learning. Thus, it is not tied to a particular machine learning algorithm. HOG can be calculated by using the following steps.

1. Normalization of global image.
2. Compute the gradient image in x and y.
3. Compute the gradient histograms.
4. Normalize across blocks.
5. Flattening into a feature vector.

5.2.5. Zernike moment

A set of rotation invariant features are introduced here. They are the magnitudes of a set of orthogonal complex moments of an image known as Zernike moments. It consists 2 values: Amplitude Value and Angle of Moment Value. The Zernike moment offers higher accuracy, less information redundancy, and much better at image reconstruction. The Zernike moments, Z_{nl}, for an image can be calculated using equation (25).

$$Z_{nl}=\left(\left(\frac{n+1}{\pi}\right)\sum_x.\sum_y V^*_{nl}(x,y)f(x,y)\right) \qquad (25)$$

Where, $x^2 + y^2 \leq 1$, $0 \leq l \leq n$, and n-l is even, $f(x, y)$ designates the intensity values of the normalized image and V^*_{nl} is a complex conjugate of a Zernike polynomial of degree n and angular dependence l.

5.2.6. GLCM (Gray Level Co-occurrence Matrix)

It is a statistical technique, which examines the texture that reflects the spatial relationship. The GLCM represents the distance and angular spatial relationship over an image sub relationship over an image sub region of specific region of specific size. GLCM functions characterize an image texture through computing how frequently pixel pairs by precise values and in an indicated spatial relationship arise in an image, forming a GLCM, and later mining statistical measures from this matrix.

5.2.7. Mean2

It computes the mean value or average value of an entire 2-dimensional array / matrix.

5.2.8. Entropy

It is a statistical measures of randomness that can be used to characterize the texture of the input image. Entropy is defined as in equation (26)

$$H(X)=-\sum_{i=1}^{N} p_i \log_2 p_i \qquad (26)$$

Where, N is the number of probable values of X, p_i is the probability that X imagines the i^{th} value. The entropy $H(X)$ can be interpreted as an average information acquired by noticing X.

5.2.9. Generalized Distance Transform (GDT)

The GDT is used to extract the distance related features. Precisely, GDT is used for Feature matching / Tracking, Dynamic programming / Stereo matching, Belief propagation, Markov Random Fields (MRFs), and many more such applications. The GDT is defined as in the equation (27)

$$D_f(P)=\min_{q\epsilon G}(d(p,q)+f(q)) \qquad (27)$$

Where, $f(q)$ is sampled on the grid G, $f(q)$ not necessarily a 2-D image, it can represent any dimensional, discrete space that encodes spatial relationships through $d(p, q)$.

VI. PROPOSED WORK

In this proposed work, an optimal effort has been placed to recognize the gestures of ASL Alphabets. The process of recognizing ASL Alphabets is distributed as preprocessing the input image by an efficient segmentation using BB Technique.

5.2.1. Preprocessing and segmentation

Step 1: Start
Step 2: Read an input RGB image.
Step 3: Detect the face by finding the face parts using BB technique.
Step 4: Mark the face by placing circle on the face.
Step 5: Convert the BB technique applied RGB image values into CIE LAB values by creating color transformation structure.

Step 6: Apply device-independent color space transformation using CIE LAB values.
Step 7: Compute global threshold (level) that can be used to convert an intensity image to a binary image using Otsu's method.
Step 8: Convert the grayscale threshold based image into a binary image.

Step 9: Detect the binary converted Hand.
Step 10: Crop and extract the detected binary hand from the previous step.
Step 11: Display the extracted hand by eliminating rest of the binary image.
Step 12: Crop and extract the palm by extracting only upper half part (by vertically dividing) of the extracted binary hand.
Step 13: Display the extracted binary palm by eliminating rest of the extracted binary hand.
Step 14: Identify the edge boundary of the extracted binary Palm using canny approximation to the derivative.
Step 15: Display the boundary of the palm.
Step 16: Stop.
Figure 7 illustrates the block diagram of proposed method of input gesture preprocessing.

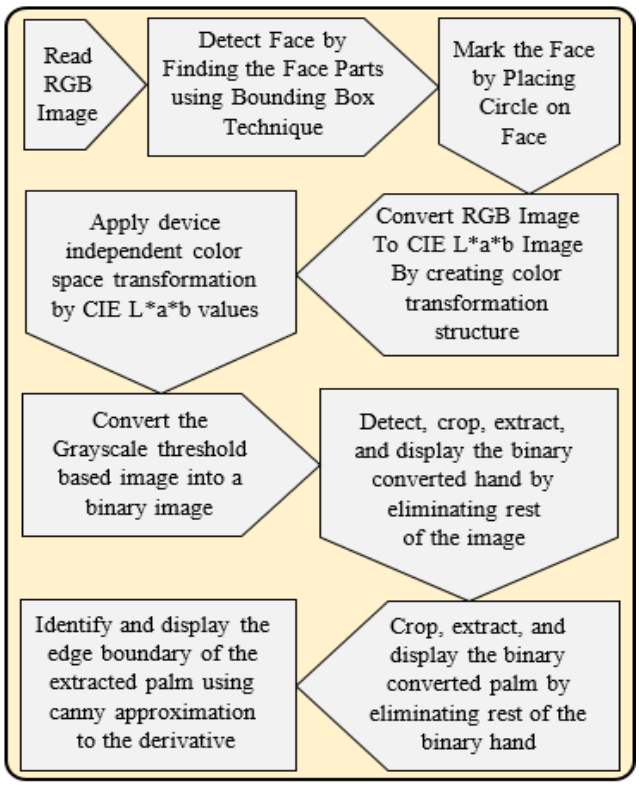

Fig. 7: Block diagram of proposed method (Preprocessing and Segmentation)

5.2.2. Feature extraction and transliteration

Step 1: Start.
Step 2: Extract the hand features by considering following feature extraction factors.
a) Area features: Boundary and centroid.
b) Texture features: LPQ and HOG feature vector.
c) Rotation invariant features: Distance transform, Zernike moment.
d) Statistical features: Mean and entropy.
e) GLCM features: Mean, Entropy, and Threshold.
Step 3: Match the gestures by choosing any of the machine learning algorithm techniques like SVM, SOM, K-NN, ANN and etc...
Step 4: Recognize the gestures and display the text associated to it.
Step 5: Stop.
Figure 8 spectacles the block diagram of proposed method of Feature extraction, and Transliteration process.

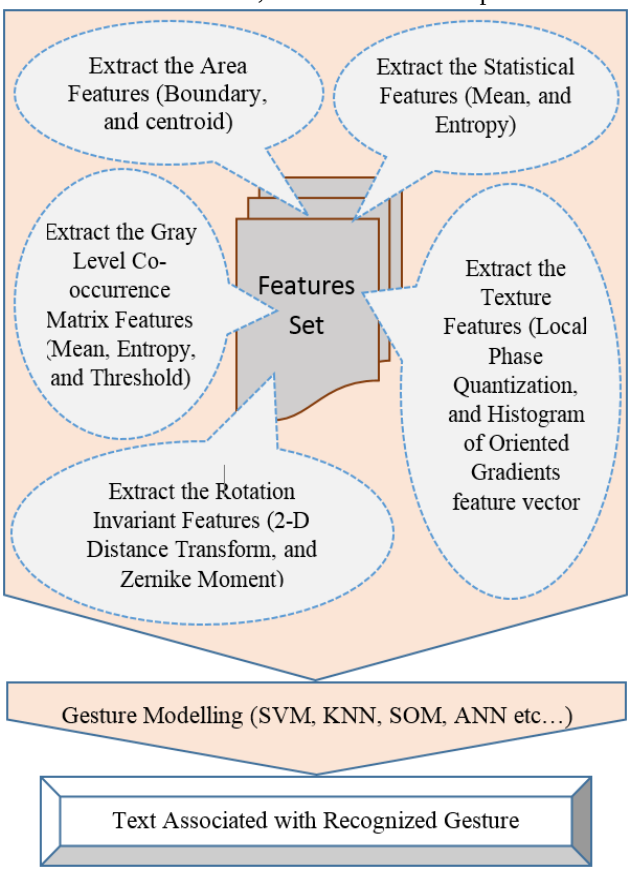

Fig. 8: Block diagram of proposed method (Feature Extraction, and Transliteration).

5.2.3. Snapshots

Following are the various screenshots obtained from the proposed algorithm.

Figure 9 depicts the initial screen of the proposed ASL gesture recognition technique. It has 6 command buttons such as 'Input Image', 'Binary Image', 'Extracted Hand', 'Extracted Palm', 'Identified Boundary', and finally the 'K-NN Output' to perform their respective operations when user clicks on that particular command button.

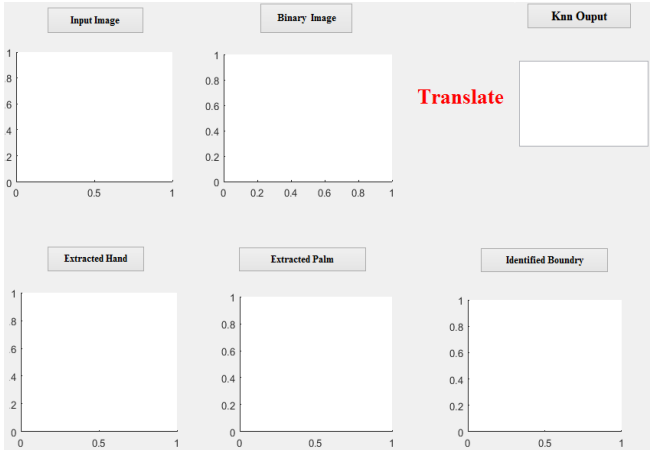

Fig. 9: ASL Alphabet Recognition Input

Figure 10 read out the input gesture of ASL alphabet when user clicks on the 'Input Image' command button for further preprocessing operations of gesture recognition task.

Figures 11 shows the BB technique applied input ges ture. Once the BB technique is applied to the input gesture, the face in the input gesture is detected and mark the face by placing the circle on the face for further segmentation process.

Figure 12, shows the input gesture and grayscale threshold based gesture, converted to binary gesture by creating color transformation structure using CIE LAB values.

In Figure 13, the process of binary hand detection and hand extraction from the binary gesture is carried out via eliminating the rest of binary gesture. Hand extraction is carried out to extract the binary palm as a next step in recognition process.

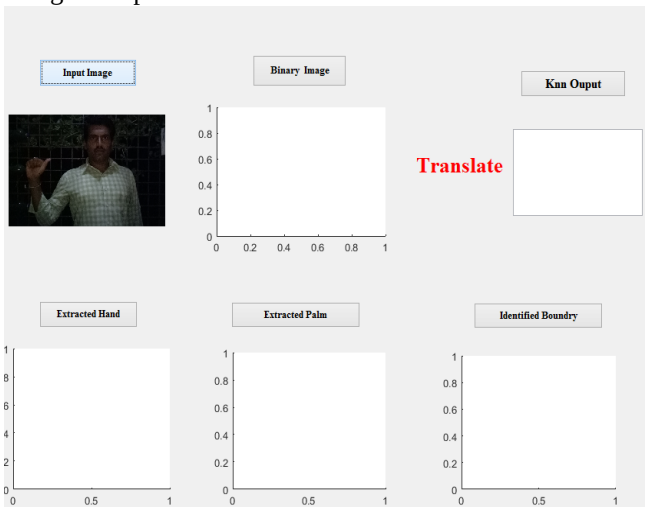

Fig. 10: Reading an input gesture

Fig. 11: BB technique applied input gesture

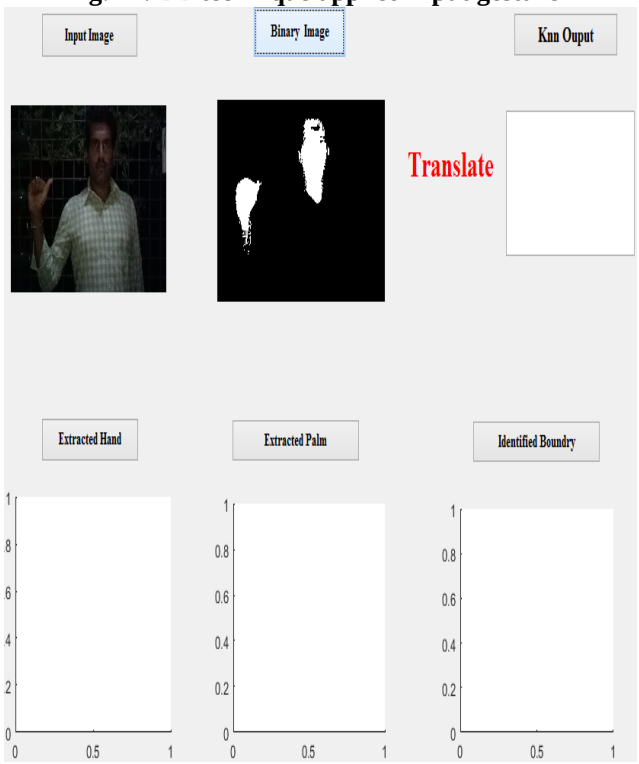

Fig. 12: Input gesture and binary converted gesture.

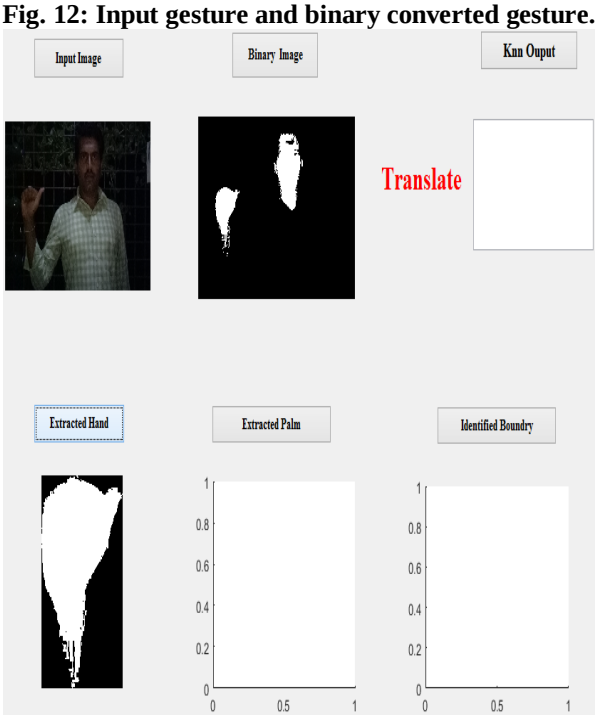

Fig. 13: Binary hand extraction from binary gesture

Figure 14 illustrates the binary palm extraction process by cropping only upper half part (by vertically dividing) of the extracted binary hand and eliminating the rest of the extracted binary hand.

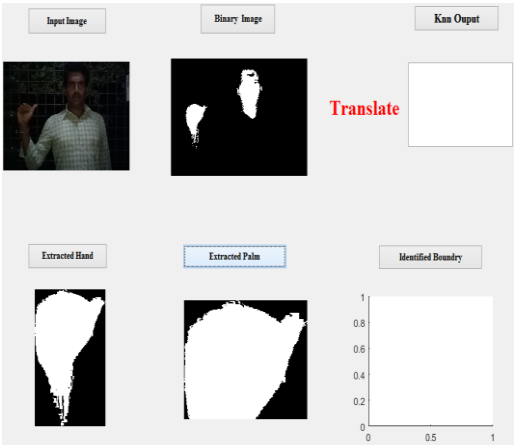

Fig. 14: Binary palm extraction from extracted hand

In Figure 15, the identified palm boundary from extracted binary palm is displayed. The palm boundary is obtained by applying canny approximation to the derivative. This palm boundary is used to extract the hand / palm features of various factors, which are utilized for gesture matching and recognizing the gestures of ASL alphabets.

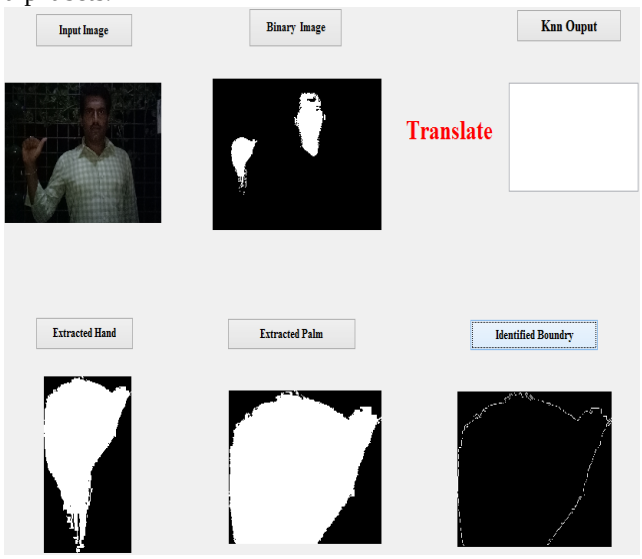

Fig. 15: Identified boundary from the extracted palm.

The final stage of the proposed algorithm i.e., gesture matching and translation of recognized gestures of ASL alphabets into human or machine readable text of English language using various hand / palm features extracted from the previous step is shown in Figure 16 along with input gesture, binary gesture, extracted binary hand, extracted binary palm, and identified palm boundary.

VII. RESULTS EVALUATION

Table 1 clearly shows the recognition rate of 28 tested data sets of ASL Alphabet gestures of the proposed technique. A total of 10 separate and 28 separate gestures data sets, which are invariant to location (indoor and outdoor), background (plain & uniform, Complex & non-uniform), illuminations (day natural light and night artificial light), distance (gestures captured from 2, 3, and 4 meters distance), and angle (gestures captured from 60^0, 75^0, 80^0, and 90^0 angle from each side of the signer) of ASL Alphabets with various resolution mobile cameras (5, 8, and 13 Mega Pixel) are used for cognition and recognition process respectively. Each data set contains the 24 gestures of ASL alphabets. Here, the alphabets 'J' and 'Z' have not included as they involve hand movement.

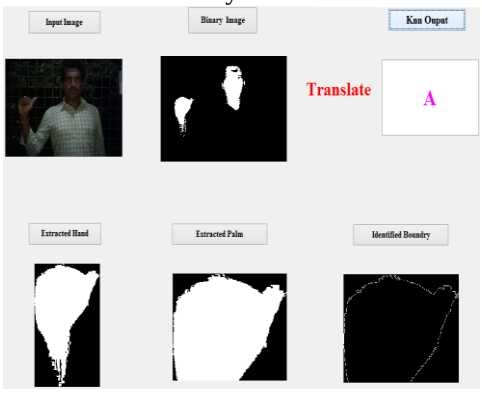

Fig. 16: Translation of gestures of ASL alphabets

The setwise gestures recognition rate of ASL Alphabets of indoor and outdoor locations of 28 data sets is presented in Figure 17. Here, 20 data sets (Set 1-5, 7-9, 11, 13-17, 19-22, 26, and 27) provides an excellent cent percent recognition rate and data sets 6, 10, 18, and 25 obtains 91.67% of accuracy as the gestures 'P' and 'Q' are not recognized. Also data sets 12, 23, 24, and 28 offers a very good recognition rate of 95.83%. In data sets 12 and 23, gesture 'P' is not recognized and in datasets 24 and 28, gesture 'Q' are not recognized.

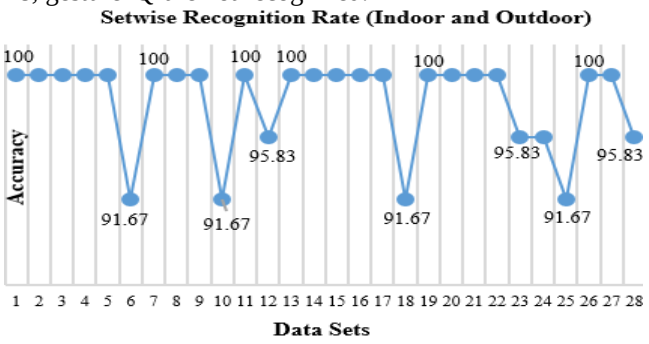

Fig. 17: Setwise Gestures Recognition Rate

Figure 18 illustrates the alphabetwise average recognition rate, where the gestures of ASL alphabets 'P' and 'Q, provides the satisfactory average recognition rate of 78.57%, and rest of the gestures of ASL alphabets yields an excellent average recognition rate of 100%.

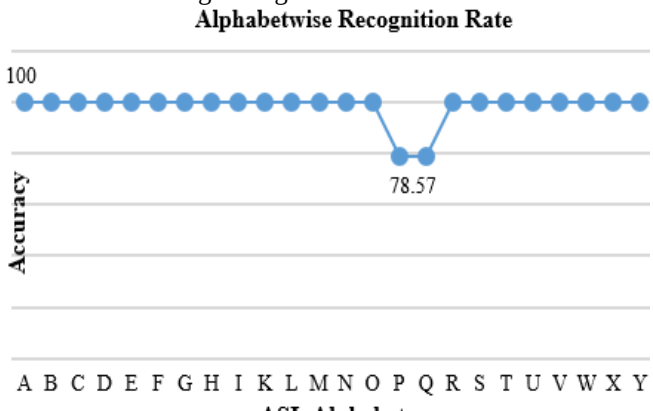

Fig. 18: Alphabetwise Average Recognition Rate

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Table 1: Recognition Rate of ASL Alphabets gestures by considering various factors.

Set No.	Location	Background	Background Color	Light	Angle from Gesture Captured	Camera Resolution (Mega Pixels)	Distance from gestures captured	Recognition Rate (%)
1	Indoor	Plain & Uniform	Gray	Day Natural	90^0	5	3 Meters	100
2	Indoor	Plain & Uniform	Gray	Day Natural	90^0	5	3 Meters	100
3	Indoor	Plain & Uniform	Gray	Day Natural	90^0	5	3 Meters	100
4	Indoor	Plain & Uniform	Gray	Day Natural	90^0	5	3 Meters	100
5	Indoor	Plain & Uniform	Gray	Day Natural	90^0	5	3 Meters	100
6	Outdoor	Complex & Non-uniform	Mixed	Night Artificial	90^0	13	4 Meters	91.67
7	Indoor	Plain & Uniform	White	Day Natural	90^0	8	3 Meters	100
8	Indoor	Complex & Non-uniform	Mixed	Day Natural	60^0	13	4 Meters	100
9	Outdoor	Complex & Non-uniform	Mixed	Day Natural	60^0	13	3 Meters	100
10	Outdoor	Complex & Non-uniform	Mixed	Night Artificial	60^0	13	4 Meters	91.67
11	Indoor	Complex & Non-uniform	Mixed	Day Natural	75^0	8	4 Meters	100
12	Indoor	Complex & Non-uniform	Mixed	Day Natural	60^0	8	3 Meters	95.83
13	Indoor	Complex & Non-uniform	Mixed	Day Natural	60^0	8	3 Meters	100
14	Indoor	Plain & Uniform	White	Day Natural	75^0	8	4 Meters	100
15	Indoor	Complex & Non-uniform	Mixed	Day Natural	60^0	8	2 Meters	100
16	Indoor	Complex & Non-uniform	Mixed	Day Natural	60^0	8	2 Meters	100
17	Indoor	Complex & Non-uniform	Mixed	Day Natural	80^0	8	4 Meters	100
18	Indoor	Complex & Non-uniform	Mixed	Day Natural	80^0	8	2 Meters	91.67
19	Indoor	Plain & Uniform	White	Day Natural	90^0	13	4 Meters	100
20	Indoor	Complex & Non-uniform	Mixed	Day Natural	90^0	13	4 Meters	100
21	Outdoor	Complex & Non-uniform	Mixed	Day Natural	90^0	13	4 Meters	100
22	Indoor	Complex & Non-uniform	Mixed	Day Natural	90^0	5	2 Meters	100
23	Indoor	Complex & Non-uniform	Mixed	Day Natural	75^0	5	2 Meters	95.83
24	Indoor	Complex & Non-uniform	Mixed	Day Natural	75^0	5	2 Meters	95.83
25	Indoor	Complex & Non-uniform	Mixed	Day Natural	75^0	5	2 Meters	91.67
26	Indoor	Plain & Uniform	White	Day Natural	75^0	5	2 Meters	100
27	Indoor	Plain & Uniform	White	Day Natural	75^0	5	2 Meters	100
28	Indoor	Complex & Non-uniform	Mixed	Day Natural	75^0	5	3 Meters	95.83

In Figure 19, the 24 setwise gestures recognition rate of ASL Alphabets of indoor location is depicted. Here, 18 data sets (Set 1-8, 10-14, 16-18, 22, and 23) provides an outstanding cent percent recognition rate and data sets 9, 19, 20, and 24 obtains satisfactory recognition rate of 95.83%. In data sets 9 and 19, gesture 'P' not recognized and in datasets 20 and 24, gesture 'Q' not recognized. Also data sets 15, and 21 offers a recognition rate of 91.67% as the 2 gestures 'P' and 'Q' were not recognized.

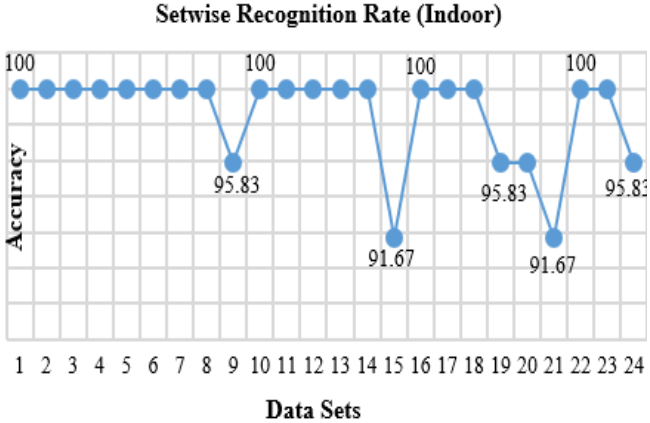

Fig. 19: Setwise Gestures Recognition Rate (Indoor location considered)

In Figure 20, the 4 setwise gestures recognition rate of ASL Alphabets of outdoor location is depicted. Here, 2 data sets (Set 2, and 4) provides an excellent cent percent recognition rate and data sets 1, and 3 obtains satisfactory recognition rate of 91.67% as the 2 gestures 'P' and 'Q' were not recognized.

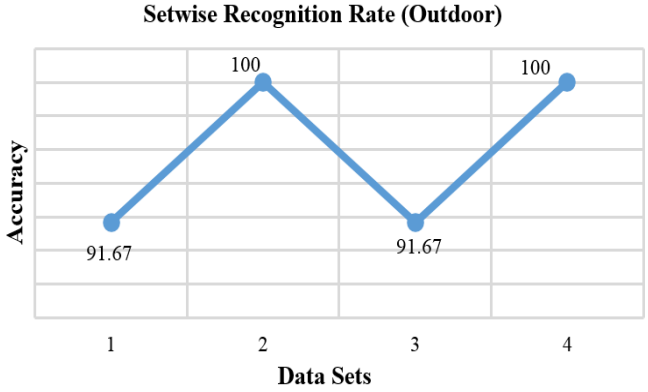

Fig. 20: Setwise Gestures Recognition Rate (Outdoor location considered)

Figure 21 shows the locations-wise average gesture recognition rate of ASL Alphabets. Here, the gestures captured in both indoor and outdoor locations yields an over and above 95% average recognition rate with invariant to background, single or mixed background colors, distance, angle, mobile camera resolution, and illumination. In particular, the indoor gestures offers an outstanding average recognition rate of 98.63%, which contains both plain and uniform as well as complex and non-uniform background gestures, whereas the outdoor gestures provides the better average recognition rate of 95.83% as the gestures captured in outdoor location are all in complex and non-uniform background.

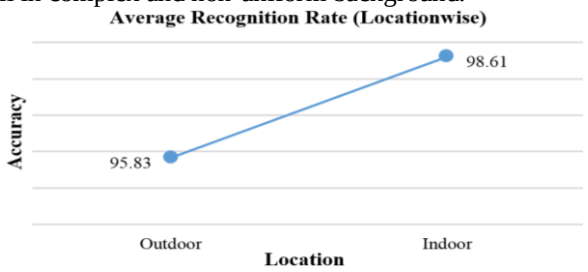

Fig. 21: Location-wise Average Gestures Recognition Rate

Figure 22 highlights the background-wise average gesture recognition rate of ASL Alphabets. Here, the gestures captured in both Complex and Non-uniform background as well as plain and uniform background yields over 97% average recognition rate invariant to location, single or mixed background colors, distance, angle, mobile camera resolution, and illumination. In detail, the complex and non-uniform background gestures gives very good average recognition rate of 97.22%, whereas the plain and uniform background gestures offers an excellent 100% average recognition rate.

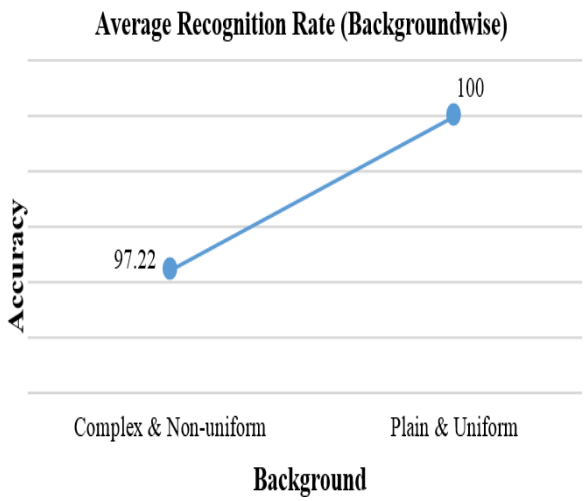

Fig. 22: Average Gestures Recognition Rate (Background considered)

The Mega Pixel-wise average gesture recognition rate of ASL Alphabets is shown in Figure 23. The gestures captured with irrespective of 3 various mobile camera resolution, it yields more than 97% average recognition rate invariant to background, location, single or mixed background colors, distance, angle, and illumination. The gestures captured from 5 mega pixel (MP) mobile camera gives 98.26% of average recognition rate, whereas 8 MP mobile camera offers highest average recognition rate of 98.61% among other 2 mobile camera resolution. Also gestures captured for the dataset from 13 MP camera yields an average recognition rate of 97.62%, which is bit lesser compare to other 2 mobile camera resolution. It is due to the maximum gestures captured in outdoor location with complex and non-uniform background.

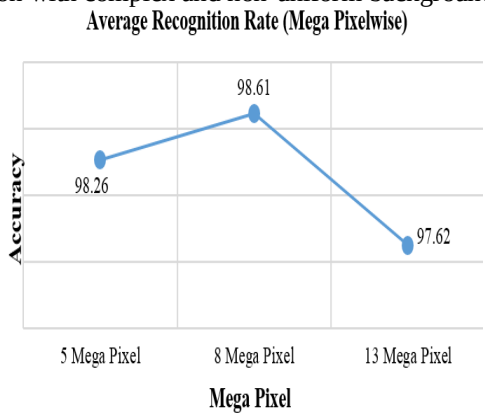

Fig. 23: Average Gestures Recognition Rate (Camera resolution considered)

The average gesture recognition rate of ASL Alphabets captured from various distances is shown in Figure 24. The gestures captured in 3 various distances, yields over and above 97% average recognition rate invariant to background, location, angle, single or mixed background colours, mobile camera resolution, and illumination. The gestures captured from 4 and 3 meters distance obtains an excellent average recognition rate of 98.15% and 99.17% respectively. The gestures captured from 2 meters distance yields 97.22% average recognition rate, which is bit lesser compare to other 2 distances as all the gestures captured here are by 5 and 8 MP cameras and the maximum gestures are captured in complex and non-uniform background.

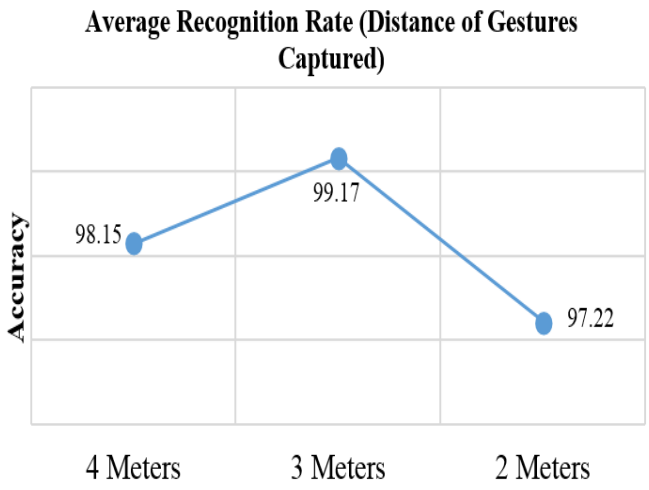

Fig. 24: Average Gestures Recognition Rate (Distance of gestures captured considered)

In Figure 25, the average gestures recognition rate of ASL Alphabets captured under various background colors is depicted. The gestures captured in 3 various background colors, yields over 97% average recognition rate invariant to background, location, mobile camera resolution, distance, angle, and illumination. Specifically, the gestures captured with mixed background colors obtains a very good average recognition rate of 97.22%, whereas in gestures captured in single plain background colors (here, gray, and white background colors considered) offers an outstanding cent percent average recognition rate.

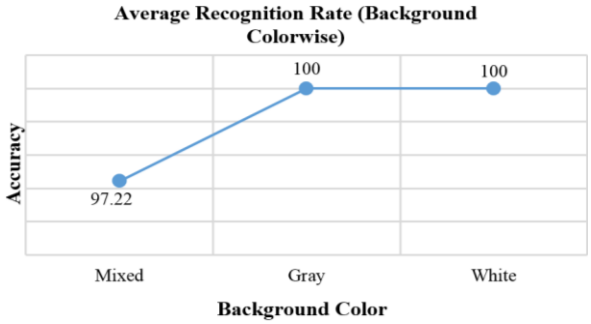

Fig. 25: Average Gestures Recognition Rate (Background color considered)

Figure 26 point up the average gesture recognition rate of ASL Alphabets captured under natural and artificial lighting conditions. The gestures captured with irrespective of natural or artificial lighting conditions, offers above 91% average recognition rate invariant to background, location, mobile camera resolution, single or mixed background color, distance, and angle. In particular, the gestures captured in natural lightings obtains a very good average recognition rate of 97.22%, whereas gestures captured in artificial lightings offers 91.67% satisfactory average recognition rate.

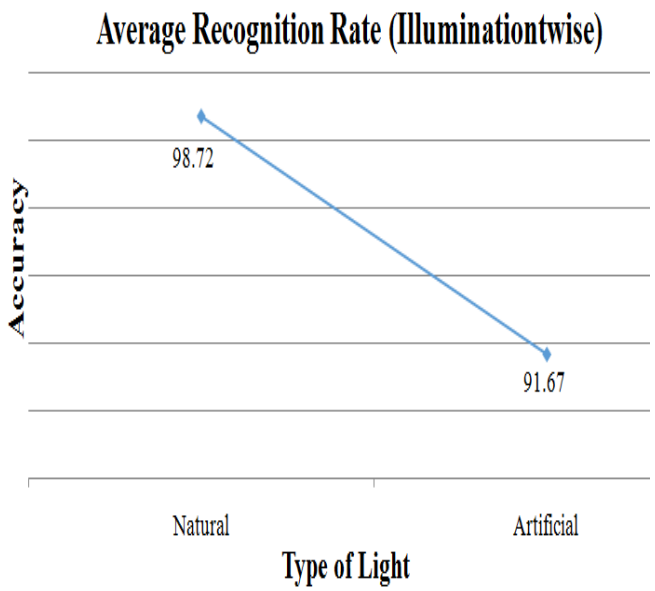

Fig. 26: Average Gestures Recognition Rate (Illumination considered)

An average gesture recognition rate of ASL Alphabets captured in various degree of angles is highlighted in Figure 27. The gestures captured in various degree of angles, offers above 95% average recognition rate invariant to background, location, mobile camera resolution, single or mixed background color, distance, and illumination. In precise, the gestures captured in 90^0 and 60^0 angle obtains an excellent average recognition rate of 98.72% and 99.17% respectively. The gestures captured in 80^0 and 75^0 offers a very good average recognition rate of 95.83% and 97.4% respectively. It is noticed that, gestures captured at 90^0, 80^0, 75^0, and 60^0 angles yields almost same average recognition rate with ± 2%.

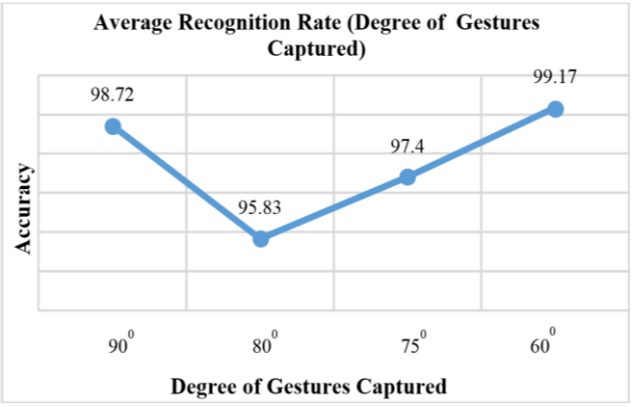

Fig. 27: Average Gestures Recognition Rate (Degree of gestures captured considered)

The distance and background considered average gesture recognition rate of ASL Alphabet are illustrated in Figure 28. The gestures captured from the distance 2, 3, and 4 meters with plain and uniform background are all offers an outstanding 100% average recognition rate. At the same time, the gestures captured from the distance 2, 3, and 4 meters with complex and non-uniform background are offers a very good average recognition rate of 96.4%, 98.6%, and 97.6% respectively.

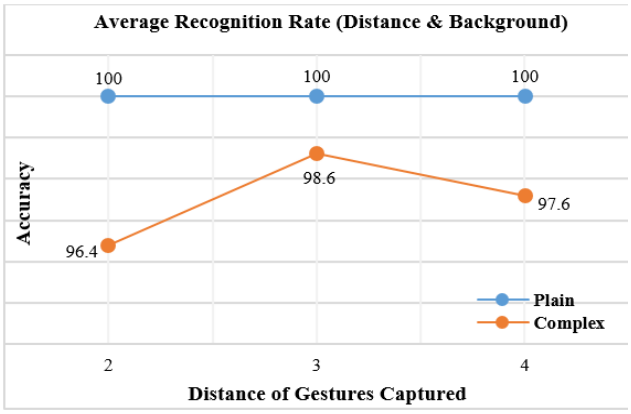

Fig. 28: Comparative average recognition rate (Plain and complex background with capturing distance of 2, 3, and 4 meters is considered)

Table 2: Recognition Rates of various Traditional Techniques with Proposed Technique

Reference and Year	Technique Used	Success Rate (%)	Remarks
[11] 2017	BPNN with Raw Features Classifiers	70	Used only 6 gestures (A, B, C, Point, 5, v) from local system or web camera.
[11] 2017	BPNN with Histogram Features Classifiers	85	
[13] 2017	Field-Programmable Gate Array with Neuromorphic Camera and ANN	79.58	24 ASL Alphabets with an angular variation of approximately $\pm 8^0$.
[14] 2017	Principle Component Analysis and Euclidian Distance Classifier	98	18 ASL Alphabets in good illumination and plain light colored background.
[15] 2017	Shape based Algorithm and Euclidian Distance Classifier	79.9	Rotation, Luminance, and Translation invariance Finger Spelled ASL gesture taken in dark background with 180 point descriptor.
[16] 2017	ANFIS and K-Nearest Neighbor Algorithm	80.77	ASL Alphabets when the number of epochs was 10 in less recognition time.
[17] 2017	Non-negative Matrix Factorization (NMF) and Compressive Sensing (CS) Theory	97.4	Occluded and Non-occluded static 10 ASL cropped hand gesture (A, B, C, D, G, H, I, L, V, Y) in uniform light, dark, and complex background.
[18] 2018	Gabor Transformation and ANFIS Classifier	98	Used readily available ASL Lexicon Video dataset which are plain background.
[19] 2018	HSV and YCbCr Color Models, Otsu Method, and various region properties	93.05	144 static occluded and non-occluded ASL Alphabets of 6 data sets (24 letters in each set). Alphabets J and Z were not included.
2018	**Proposed Technique**	**98.21**	

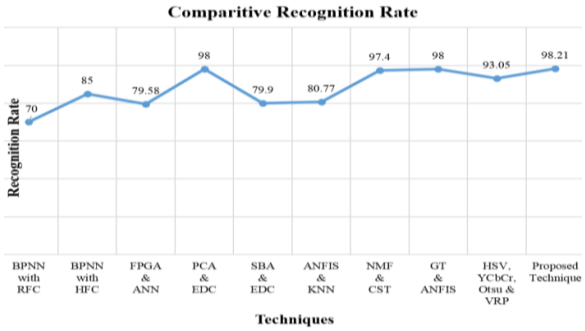

Fig. 29: A Comparative Recognition Rate of some Traditional Existing Techniques and Proposed Technique

Table 2 and Figure 29 highlights the comparative average gesture recognition rate of ASL Alphabets of existing (papers published in year 2017, and 2018) techniques with the proposed technique. It is noticed that, the proposed technique offers very good recognition rate of 98.21% which is marginally better compared to the 2 existing recognition techniques as both Principle Component Analysis and Euclidian Distance Classifier (PCA & EDC) and also Gabor Transformation (GT) and ANFIS Classifier techniques offers 98% recognition rate. But PCA and EDC technique has taken only 18 ASL Alphabets in good illumination and plain light coloured background where as GT and ANFIS classifier technique used readily available ASL Lexicon Video dataset which are plain in background.

(Note: Figure 23 Abbreviations, BPNN with RFC: Back-Propagation Neural Networks with Raw Features Classifiers, BPNN with HFC: Back-Propagation with Histogram Features Classifiers, FPGA and ANN: Field-Programmable Gate Array with Neuromorphic Camera and Artificial Neural Network, PCA and EDC: Principle Component Analysis and Euclidian Distance Classifier, SBA and EDC: Shape Based Algorithm and Euclidian Distance Classifier, ANFIS and KNN: Adaptive Network-based Fuzzy Inference System (ANFIS) and K-Nearest Neighbour Algorithm, NMF and CST: Non-negative Matrix Factorization (NMF) and Compressive Sensing (CS) Theory, GT and ANFIS: Gabor Transformation and ANFIS Classifier, HSV, YCbCr, Otsu and VRP: HSV and YCbCr Colour Models, Otsu method, and Various Region Properties).

VIII. CONCLUSION

This research paper exhibits an inventive framework, to achieve the transliteration of 24 static alphabets (Letter J and Z not included as they involve hand movement) of American Sign Language into English text and achieved an average recognition rate of 98.21% which is the best in recent (papers published in year 2017, and 2018) existing traditional work carried out. This paper also summarizes the system architecture, state of art, data collection for the proposed work, proposed system design, and the detailed results evaluation by showing comparative graphical depiction of the proposed technique with the existing techniques average recognition rate and also depicts the average gesture recognition rate chart by considering various factors like background complexity, background color, location, time, distance, angle, mobile camera resolution, and illumination. This paper also highlights on face detection and edge detection technique, and also the various hand / palm features extraction techniques.

FUTURE PERSPECTIVE

As part of the future perspective, this research work can be extended to recognize the ASL Alphabets and Numeric gestures as well as some complex gestures in both static and real time environment considering in plain and complex background with different lighting conditions.

REFERENCES

[1] B.M.Chethana Kumara, H.S.Nagendraswamy, R.Lekha Chinmayi, "Spatial Relationship Based Features for Indian Sign Language Recognition", *International Journal of Computing, Communications & Instrumentation Engineering*, Vol. 3, Issue 2, (2016), pp.206-212, available online: http://dx.doi.org / 10.15242 / IJCCIE.IAE0516005.
[2] Rachael Locker McKee, David McKee, "What's so hard about learning ASL?: Students' AND Teachers' perceptions", *Sign Language Studies, Linstok Press Inc.*, (1992), pp.129-157,
[3] Srinath.S, Ganesh Krishna Sharma, "Classification approach for sign language recognition", *International Conference on Signal, Image Processing, Communication & Automation*, (2017), pp.141-148.
[4] Shivashankara S, Srinath S, "A comparative Study of Various Techniques and Outcomes of Recognizing American Sign Language: A Review", *International Journal of Scientific Research Engineering & Technology*, Vol 6, Issue 9, (2017), pp.1013-1023, available online: http://www.ijsret.org
[5] Shivashankara S, Srinath S, "A Review on Vision Based American Sign Language Recognition, its Techniques, and Outcomes", *7th IEEE International Conference on Communication Systems and Network Technologies*, (2017), pp293-299, DOI 10.1109 / CSNT.2017.58
[6] Dr.Roger Sapsford, Victor Jupp, *Data Collection and Analysis*, Sage Publishing Ltd, (2006)
[7] URL: http://www.asluniversity.com
[8] URL: http://www.asl.tc
[9] URL: http://www.lifeprint.com
[10] URL: http://www.youtube.com
[11] Tülay Karay1lan, Özkan K1l1ç, "Sign Language Recognition", *2nd International Conference on Computer Science and Engineering*, (2017), pp.1122-1126.
[12] N.S Sreekanth, N.K Narayanan, "Static Hand Gesture Recognition using Mon-vision Based Techniques", *International Journal of Innovative Computer Science & Engineering*, Vol. 4, Issue 2, (2017), pp.33-41, 2017, available online: http://www.ijicse.in.
[13] Miguel Rivera-Acosta, Susana Ortega-Cisneros, Jorge Rivera, Federico Sandoval-Ibarra, "American Sign Language Alphabet Recognition Using a Neuromorphic Sensor and an Artificial Neural Network", *Sensors*, (2017), pp.1-17, available online: http://www.mdpi.com / journal / sensors.
[14] Nitesh S. Soni, Prof. Dr. M. S. Nagmode, Mr. R. D. Komati, Online Hand Gesture Recognition & Classification for Deaf & Dumb, International Conference on Inventive Computation Technologies, 2017.
[15] Amit Kumar Gautam, Ajay Kaushik, "American Sign Language Recognition System Using Image Processing Method", *International Journal on Computer Science and Engineering*, Vol. 9, No.07, (2017), pp.466-471, available online:

http://www.enggjournals.com/ijcse/doc/IJCSE17-09-07-028.pdf
[16] Fifin Ayu Mufarroha, Fitri Utaminingrum, "Hand Gesture Recognition using Adaptive Network Based Fuzzy Inference System and K-Nearest Neighbor", *International Journal of Technology*, (2017), pp.559-567, available online: https://www.researchgate.net / publication / 316581665
[17] Huiwei Zhuang, Mingqiang Yang, Zhenxing Cui, Qinghe Zheng, "A Method for Static Hand Gesture Recognition Based on Non-Negative Matrix Factorization and Compressive Sensing", *IAENG International Journal of Computer Science*, (2017), available online: http://www.iaeng.org / IJCS / issues_v44 / issue_1 / IJCS_44_1_07.pdf
[18] Neethu P S, Dr. R Suguna, Dr.Divya Sathish, "Real Time Hand Gesture Recognition System", *Taga Journal*, Vol. 14, (2018), pp.7982-792, available online: http://www.tagajournal.com
[19] Shivashankara S, Srinath S, "American Sign Language Recognition System: An Optimal Approach", *International Journal of Image, Graphics and Signal Processing*, Vol.10, No.8, (2018), pp. 18-30, available online: http://www.mecs-press.org/
[20] Victor Lempitsky, Pushmeet Kohli, Carsten Rother, Toby Sharp, "Image Segmentation with A Bounding Box Prior", *Microsoft Research Cambridge*, (2009), available online: https:// www.microsoft.com / en-us / research / publication / image-segmentation-with-a-bounding-box-prior/
[21] Nobuyuki Otsu, "A threshold selection method from gray-level histograms", *IEEE Trans. Sys., Man., Cyber*, Vol 9, Issue 1, (1979), pp.62–66, available online: https:// ieeexplore.ieee.org / document / 4310076/
[22] John Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol PAMI-8, No.6, (1986), pp.679-698, available online: https:// ieeexplore.ieee.org / abstract/document/4767851/