iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VI **Month of publication:** June 2021

DOI: https://doi.org/10.22214/ijraset.2021.35643

Number Evaluation for Auditorily and Verbally Challenged Children

Krishna Tyagi[1], Raj Lunawat[2], Nishita Baronia[3]
[1, 2, 3]*Electronics and Telecommunication Engineering, Pune Institute of Computer Technology*

Abstract: The objective of this paper is to develop a system designed on MATLAB 2021a to serve as a holistic assessment tool for Auditorily and Verbally challenged children studying in grades 1 and 2. This system captures and recognizes specific hand gestures indicated by these children as answers to mathematical questions displayed on the system's screen. It then proceeds to evaluate the answers as right or wrong and accordingly displays the scores obtained. This project addresses the steps that must be undertaken to develop a viable, functioning number recognition assessment tool. Algorithms have been chosen taking into consideration obstacles that may arise out of poor illumination, complicated backgrounds or incorrect rotation of the images captured. We have utilized: Image pre-processing, Feature extraction, Feature Classification and the MATLAB App Designer to achieve our objective.

I. INTRODUCTION

The motivation behind undertaking this project was to try and bridge the learning gap encountered by children suffering from challenging impairments. Using Image pre-processing and machine learning concepts we have attempted to formulate an assessment application where these children can solve mathematical problems using hand gestures.

The project consists of 4 modules:

A. Image pre-processing
B. Feature Extraction
C. Feature Classification
D. Designing the User Interface

The resources used by us include:
a) Camera Module-(720p,30fps)
b) MATLAB 2021A

II. LITERATURE SURVEY

1) Sign language recognition consists of the following three phases: Segmentation, Feature Extraction and Recognition. The segmentation is performed through Otsu's algorithm while in the feature extraction phase, the components: shape descriptors, Histogram of Oriented Gradient descriptors (HOG Descriptors) and Discrete Wavelet Transform (DWT) are fused in order to compute a feature vector.
2) In the recognition phase, a multi-class Support Vector Machine (MSVM) is used to train and classify signs of ISL. Image pattern recognition has widely seen the use of "moment invariants" due to its resolute features on image translation, scaling and rotation. These moments are strictly invariant for the continuous function, however during practical application, the images are discrete. Consequently, the moment invariants may observe change over image geometric transformation. To address this problem, an analysis with respect to the variation of movement invariants was done, in order to analyze the effect of the image's scaling and rotation along with narrowing on the best possible methods to minimize the moment invariants' fluctuation.
3) For hand-gesture detection two approaches were thought of. The first centered around color and motion information and the second centered around color segmentation. In recent times, a lot of time and effort has been dedicated to the process of hand-gesture recognition. The technology used can be divided into two different categories:
a) *Vision Based:* In this method, the computer's web-cam serves as the input device for observing and capturing the gesture related data. This method requires no human-intervention and is self-sufficient as all it needs is a camera and no other external device. This system tends to complement biological vision by emulating artificial-vision system implemented in its software and/or hardware.
b) *Glove Based:* In this system data gloves are used to accurately depict the positions of the gestures by directly measuring those positions. Sensors are then used to detect these positions and to decipher the gesture.

III.PROBLEM FORMULATION AND PROPOSED WORK

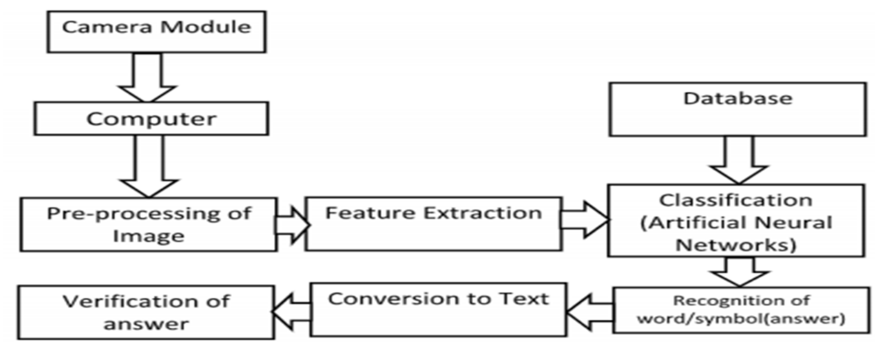


Fig. 1 Block Diagram of the Proposed System.

A. Image Pre-processing

Pre-processing is a common term used when improving performance of an image at the lowest level of abstraction i.e., when both input and output images are intensity images. The two kinds of images i.e., iconic and intensity differ in representation. Iconic images are of the same kind as the original data captured by the sensor while an intensity image is usually represented with the help of a matrix containing image function values (brightness). Pre-processing aims to improve the quality of an image that has been captured. It does this by suppressing distortions and also by enhancing some important features relevant for carrying out further processing more smoothly.

Image pre-processing techniques used:

1) *Capturing the Gesture:* A bounding box is used in order to capture the gesture. The student will be instructed to portray his hand gesture only in the area bounded by the box. The image captured is then cropped and sent on for further processing i.e., conversion to a binary image along with feature extraction.
2) *Dilate:* Dilation of a binary image expands the bordering regions of the "on" pixels in an image i.e., these adjacent pixels are each examined within a "Window" around each pixel. If the threshold allowing a certain number of neighboring pixels to be turned "on" is breached then the dilation automatically turns the pixel currently under examination "on". Pixels that were originally "on" are not touched.
3) *Erode:* Erosion of a binary image serves to contract the regions bordering pixels that are "on" i.e., these adjacent pixels are each examined within a "Window" around each pixel. If the threshold allowing a certain number of neighboring pixels to be turned "off" is breached then the erosion automatically turns the pixel currently under examination "off". Pixels that were originally "off" are not touched. Dilation followed by erosion has the effect of smoothing out the shapes (clusters of "on" pixels) present in a binary image along with the filling of holes within these existent shapes. Erosion followed by dilation has the effect of smoothing out the shapes (clusters of "on" pixels) present in a binary image along with the obliteration of any small, irrelevant shapes or any isolated pixels.
4) *Max, Min, Median and Mean:* These methods are used to obtain the max, min, median or mean value of the neighboring mages enclosed within a window or a box. The result procured after performing operations of binarization, dilation and erosion has been depicted in Fig 2.

Fig. 2 Image background subtraction and erosion

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

B. Feature Extraction

Feature extraction involves extracting specific features from the image. These features are vital for the analysis, interpretation and re-construction of the image.

The features extracted for the implementation of this system are:

1) *Eigen Vectors:* To determine the Eigen Vectors, we use the Principal Component Analysis. This is done to perform dimensionality reduction on the large data-sets.
2) *Corners:* The binaries along with the resized inputs are given as inputs to the system which then returns the corner points. The exact number of these corner points is then stored in the form of a feature. These corner points are plotted as shown in Fig 3 for the number 4 and Fig 4 for the number 5.

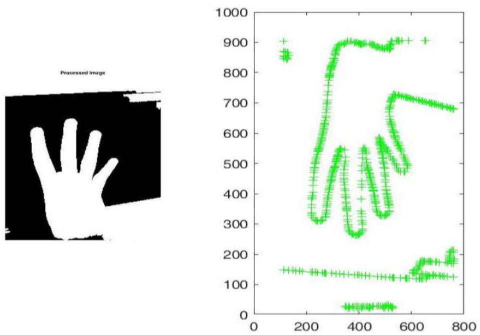


Fig.3 Corner points for gesture '4'

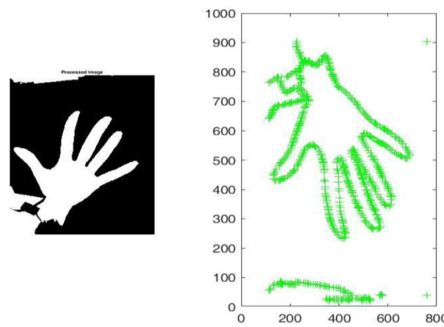


Fig.4 Corner points for gesture '5'

3) *Central and Simple Moments:* A regular moment projects onto a monomial x^p.y^q in the form of the function f(x,y) as :

$$M_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x^p)(y^q)\, f(x,y)_{dxdy}$$

It includes the simple transitional and the scale invariant properties. However, these moments are not orthogonal and the computational presentation increases with an increase in the order. This makes reconstruction of the image an arduous task.

4) *Zernike Moments:* Zernike moments are used in order to extract the features of binary digits present in binary images. They describe the function on the unit desk in a unique manner and can also be extended on to images. They have invariance properties which make them appealing as descriptors for the purpose of gesture recognition. Different feature extraction methods have been designed to carry out the different representations of characters such as the solid binary character and the grey sub level image of each individual character. All the pixels present in an image have random orientations, this factor makes it a must for all the features to be invariant when it comes to the translation and rotation of the pixels present in a field of view. Since Zernike polynomials are orthogonal, we can use them to reconstruct an image easily.
5) *Peaks:* Hand gestures can be recognized by the means of finger detection. After segmentation of the image that is dividing the image into multiple segments to locate objects and boundaries, the edges are detected using Canny edge detector. Canny edge detector is an edge detection operator that helps to detect the edges in an image. Noise can be removed by considering the segmented hand as the biggest region and removing the small objects. K-curvature algorithm has been used for peak detection. Hand contour, the larger segment, has been extracted and represented with boundary point, A. Two other points, B and C have been considered at distance k. Angle α is computed between the two vectors by simply taking their dot product. In order to find whether it is a peak or a valley the vectors are converted into 3-D lying in the XY plane and then their cross product is computed. If the sign of the Z component is positive, then it is considered as weak whereas for a negative sign it is considered as a valley. By knowing the number of peaks and valleys the hand gesture can be recognized.

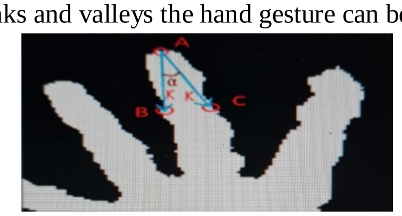


Fig.5 Peaks

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

6) *Fourier Descriptors:* Fourier descriptors is used to represent the boundary shape of a segment in an image. The basis of the descriptor is given by the first few terms in a Fourier series. This type of object descriptor is useful for recognition tasks because it can be designed to be independent of scaling, translation, or rotation. The Fourier transformed coefficients from the Fourier Descriptors of the shape represent the shape of the object in the frequency domain. The general features of the shape can be found in the lower frequency descriptors, while the higher frequency descriptors contain information about the shape details. After normalization we apply Fourier transform to the shape signature. A shape signature is any 1-D function representing 2-D areas or boundaries. The shape signature we used is complex coordinates. A complex coordinates function is the complex form of the boundary coordinates.
7) *Hu Moments:* Hu Moments are used to signalize the outline of an object in an image. Moment invariants are firstly introduced by Hu. Hu derived six absolute orthogonal invariants and one skew orthogonal invariant based upon algebraic invariants, which are independent of position, size, orientation and parallel projection. Image moments can basically be defined as weighted average of image pixel intensities of an image. The pixel intensity at location (x,y) is given by I(x,y).

The 7 moments are calculated using the following formulae:

$$h_0 = \eta_{20} + \eta_{02}$$

$$h_1 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$h_2 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$

$$h_3 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$h_4 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

$$h_5 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})]$$

$$h_6 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

8) *Histogram of Oriented Gradients (HOG):* Histogram of oriented gradients describe gestures by distributing local intensity gradients. It is essentially a feature descriptor used to detect objects when it comes to image processing and computer vision. It does this by counting occurrences of gradient orientation in localized portions of an image. The effect of varying the cell size in order to acquire Histogram of oriented gradient descriptors can be observed through Fig 6 representing a gesture depicting the number one. This helps us zero onto the exact cell size required.

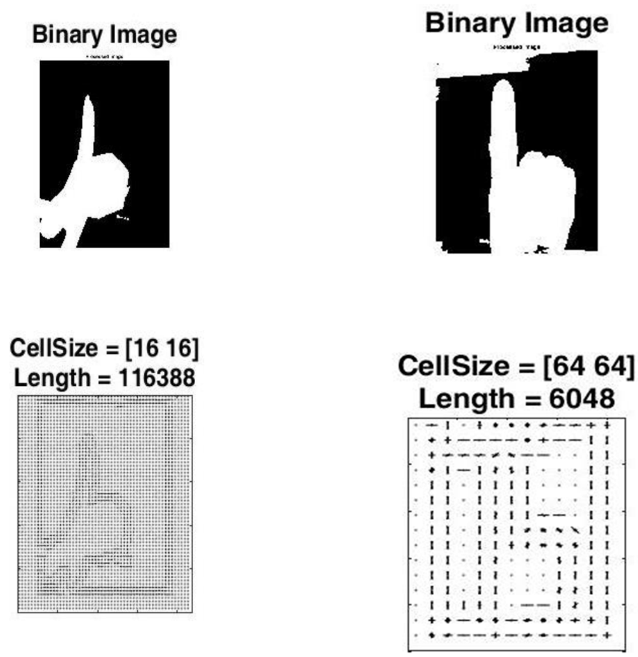


Fig. 6 HOG for gesture ‘1’ (for cell size [16,16] and [64, 64])

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

C. Feature Classification

We have computed two different training models, trainedClassifier1 for numbers 0 to 5 and trainedClassifier2 for numbers 6 to 9. We trained our database using the following algorithms and obtained the following results:

1) Support Vector Machine

Fig.7 Confusion Matrix for SVM

Fig 7 shows the confusion matrix generated for SVM. The diagonal elements show the number of correctly classified points. Ideally, the value of diagonal elements must be maximum while the others should be minimum (ideally zero). For example, in the figure above, the first row has two elements, 37 in the diagonal box means that in the input training database for the gesture 'two' which had 40 training images, 37 images were correctly classified as 'two', while 3 images were misclassified as 'four.

2) K-Nearest Neighbour

Fig.8 Confusion Matrix for KNN

Fig 8 shows the confusion matrix generated for KNN. The diagonal elements show the number of correctly classified points. Ideally, the value of diagonal elements must me maximum while the others should be minimum (ideally zero). For example, in the above figure the first row has two elements, 33 in the diagonal box means that in the input training database for the gesture 'two' which had 40 training images, 33 images were correctly classified as 'two', while 7 images were misclassified as 'four'.

3) Decision Trees

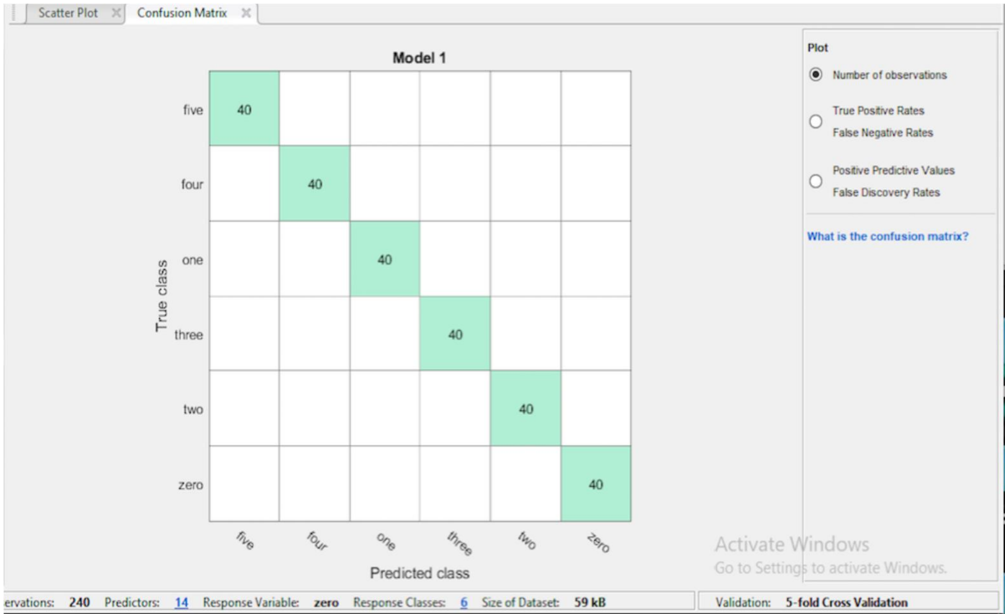


Fig.9 Confusion Matrix for Decision Trees.

Fig 9 depicts the confusion matrix generated for Decision tree. The diagonal elements represent the number of correctly classified points. Ideally, the diagonal elements must possess the highest possible or maximum value while the others ought to be close to or ideally zero. For example, in the above figure all the diagonal elements have attained the maximum value. This can be used to safely assume that the images have been accurately classified

We obtained the best accuracy using Decision trees learning method. Also, after extensive research and comparison, we used the decision trees learning algorithm to train our model.

Table I. Accuracy obtained for numbers 0 to 5

Model Used	Accuracy
SVM	86.7%
KNN	51.7%
DECISION TREES	100%

For the Classifier2 used for identifying numbers ranging from 6 to 9, features were extracted using HOG and classified using SVM algorithm.

Table II. Accuracy obtained for numbers 6 to 9

Model Used	Accuracy
QUADRATIC SVM	48.7%
SIMPLE TREE	43.6%
MEDIUM GAUSSIAN SVM	43.6%
MEDIUM KNN	33.3%

This table shows the accuracy percentages obtained when the HOG features were extracted from training gestures representing numbers 6 to 9 and given to various classifiers.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

D. GUI Design

We used the GUIDE to build our graphical user interface. We created the test in the following main parts-

1) *Student Information:* The student is required to enter his/her details such as Name and Roll Number.

Fig.10 Student Details Tab

2) *Test*: Our test consists of ten carefully crafted mathematical questions. The student must read the question and answer it using the sign language by showing it in front of the camera. A sample question is shown in Fig. 11

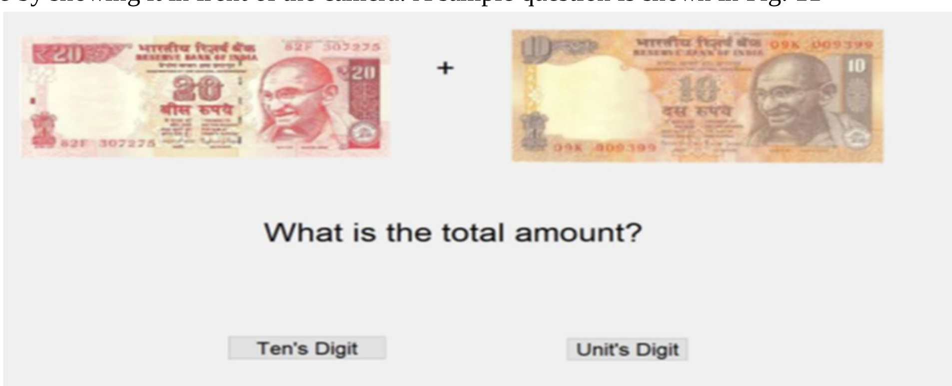


Fig.11 Sample Question

3) *Score:* The final score is displayed on the screen as shown in Fig.12.

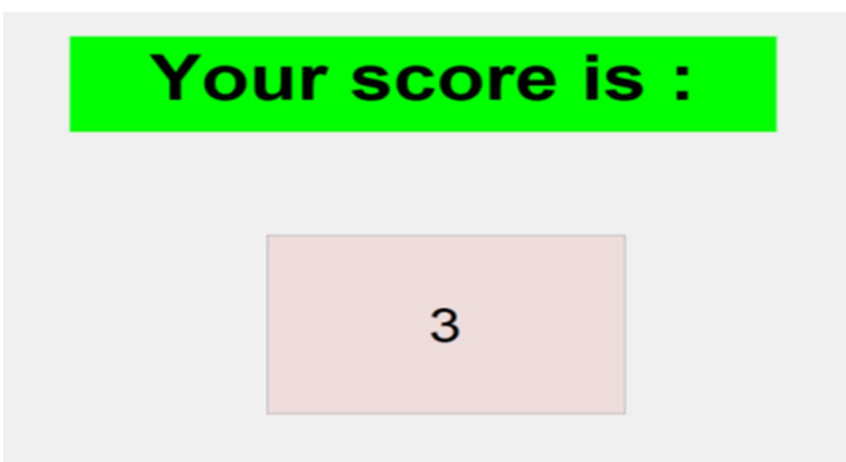


Fig.12 Final Score

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

IV. RESULTS AND OBSERVATIONS

A. Capturing the Image

Fig.13 Capturing the Image using a bounding

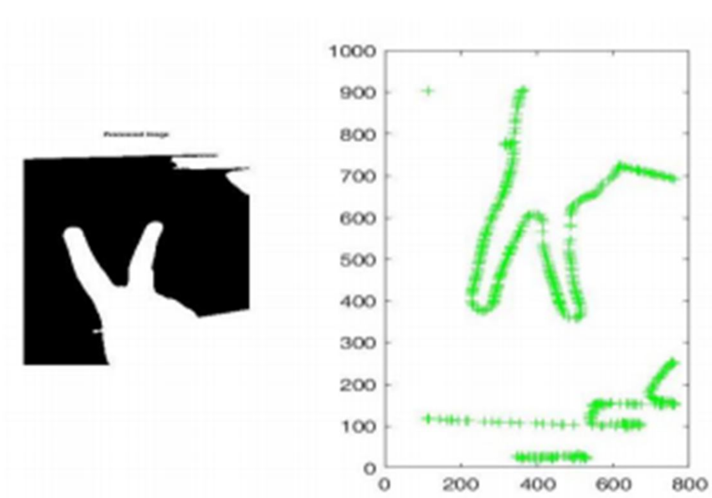
The figure above (Fig 13) shows the image captured using bounding box. The student has to show the answer gesture within the bounding box.

B. Pre-processed Image

Fig.14 Obtained Gesture

The figure above (Fig 14) shows the pre-processed image. The gesture obtained from the bounding box is pre-processed by using various operations such as binarization, dilation and erosion.

C. Feature Extraction



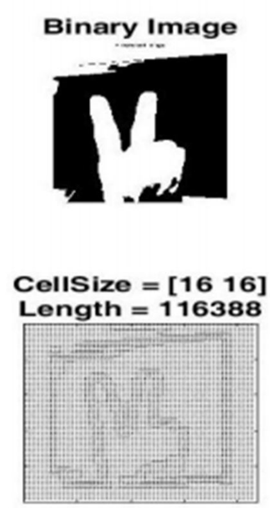


Fig.15 Corner Points for the gesture

Fig.16 HOG Feature extraction for Cell size [16 16]

The figures 14 and 15 respectively show the output image after the corners and the HOG features of the binary image are extracted.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VI Jun 2021- Available at www.ijraset.com

D. Feature Classification

A comparative table of features (Eigen Values, Corners, Central Moments, Simple Moments, Zernike Moments, Peaks, and Fourier Descriptors) and classification based on different algorithms (SVM, KNN, Decision Trees) is constructed to identify the most suitable classification technique.

Table III. Efficiency Using Decision Trees (0 to 5)

EIGEN VALUES	CORNERS	CENTRAL MOMENT	SIMPLE MOMENT	ZERNIKE MOMENT	PEAKS	FOURIER DESCRIPTOR	I1	I2	I3	I4	I5	I6	I7	EFFICIENCY
Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	100
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	100
Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	100
Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	100
Y	Y	N	N	N	Y	N	Y	Y	Y	N	N	N	N	100
Y	Y	Y	Y	Y	Y	Y	N	N	N	N	N	N	N	100

Table IV. Efficiency Using SVM (0 to 5)

EIGEN VALUES	CORNERS	CENTRAL MOMENT	SIMPLE MOMENT	ZERNIKE MOMENT	PEAKS	FOURIER DESCRIPTOR	I1	I2	I3	I4	I5	I6	I7	EFFICIENCY
Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	N	N	92.1
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	92.1
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	85.8
Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	92.9
Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	95.4
Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	92.9
Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	N	N	N	92.5

Table V. Efficiency Using KNN (0 to 5)

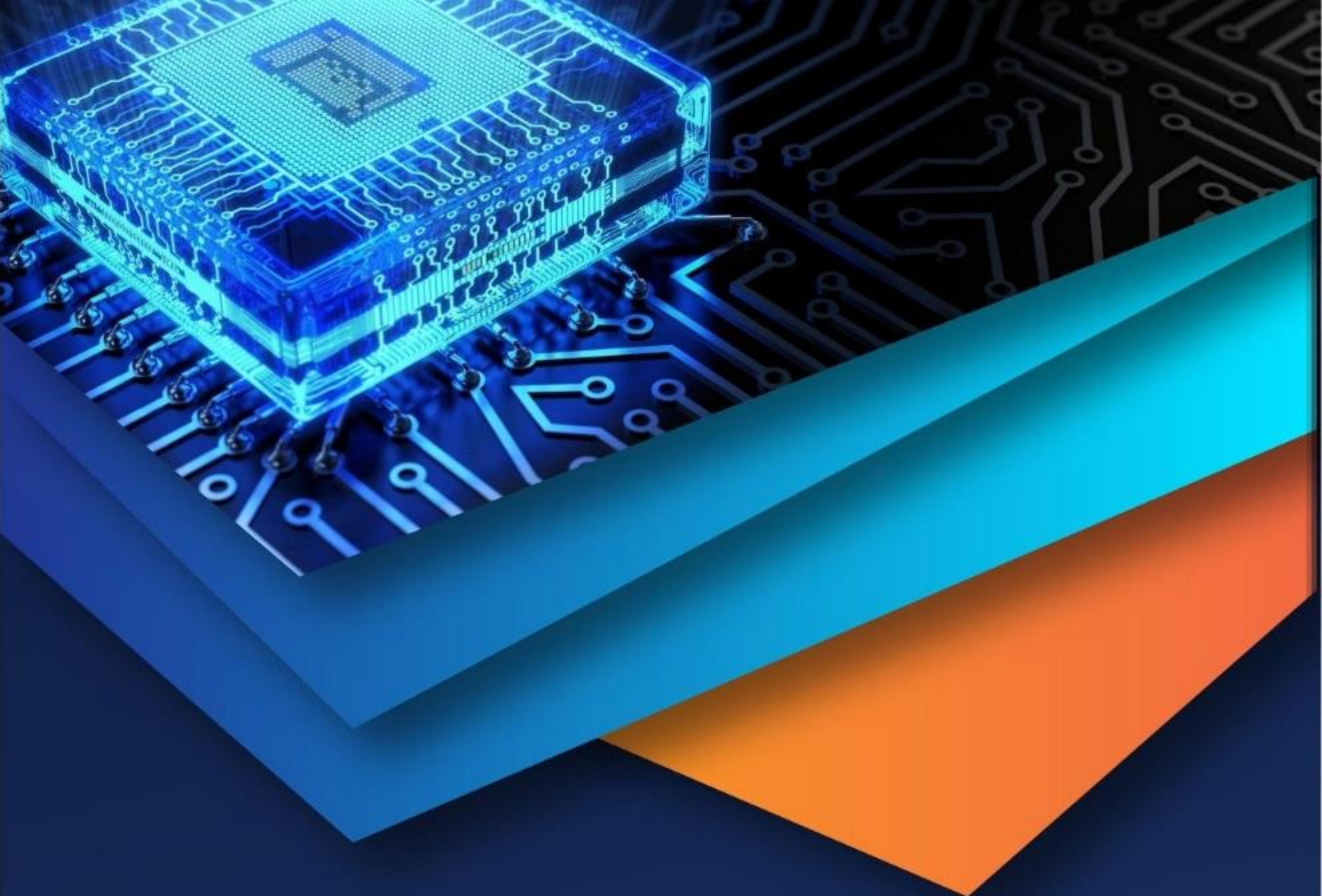
EIGEN VALUES	CORNERS	CENTRAL MOMENT	SIMPLE MOMENT	ZERNIKE MOMENT	PEAKS	FOURIER DESCRIPTOR	I1	I2	I3	I4	I5	I6	I7	EFFICIENCY
Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	N	N	N	57.9
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	40.8
Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	57.9
Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	64.6
Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	N	N	N	60.8
Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	N	N	63.3
Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	N	N	N	67.9

V. CONCLUSIONS

In this project we have implemented a sign-language recognition model, which assesses the mathematical proficiency of verbally and auditorily challenged children, using template machinery in order to obtain the required results. In our proposed work, we implemented a system of gesture recognition based on the extraction of different features such as: Eigen values, corners, HOG, Moments and classification using SVM and Decision Trees. We created a hybrid model I.e., a separately trained model for numbers 0-5 in order to achieve a rate of accuracy equaling 100% and another model for numbers 6-9 which may achieve an accuracy rate of about 55%, on MATLAB.

REFERENCES

[1] Solomon C, Fundamentals of Digital Image Processing, 1st Edition; John Wiley Sons Inc. New York, 2011.
[2] Li, Linghua & Zhang, Lili. (2012). Corner Detection of Hand Gesture. TELKOMNIKA Indonesian Journal of Electrical Engineering. 10. 10.11591/telkomnika. v10i8.1646.
[3] G. G. Nath and C. S. Arun, "Real time sign language interpreter," 2017 IEEE International Conference on Electrical, Instrumentation and Communication Engineering (ICEICE), 2017, pp. 1-5, doi: 10.1109/ICEICE.2017.8191869
[4] H. Muthu Mariappan and V. Gomathi, "Real-Time Recognition of Indian Sign Language," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), 2019, pp. 1-6, doi: 10.1109/ICCIDS.2019.8862125.
[5] K. Nimisha and A. Jacob, "A Brief Review of the Recent Trends in Sign Language Recognition," 2020 International Conference on Communication and Signal Processing (ICCSP), 2020, pp. 186-190, doi: 10.1109/ICCSP48568.2020.9182351.
[6] S. Shivashankara and S. Srinath, "A review on vision based American sign language recognition, its techniques, and outcomes," 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), 2017, pp. 293-299, doi: 10.1109/CSNT.2017.8418554.
[7] Adithya, V., Vinod P.R. and Gopalakrishnan U., "Artificial neural network based method for Indian sign language recognition, " Information and Communication Technologies (ICl], 2013IEEE Conference on, vol., no., pp.1080,1085, 11-12 April 2013.
[8] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, “Image Features Detection, Description and Matching” Springer, 2016, IX 438p, 238 llus.
[9] Zacniewski, Artur. (2017). Using Zernike moments in the process of automatic identification. Autobusy - Technika, Eksploatacja, Systemy Transportowe. 6. 1629-1632.
[10] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, “Image Features Detection, Description and Matching” Springer, 2016, IX 438p, 238 llus.
[11] M. Hassaballah, Aly Amin Abdelmgeid and Hammam A. Alshazly, “Image Features Detection, Description and Matching” Springer, 2016, IX 438p, 238 llus.
[12] Sign Language videos by Noida Deaf Society.



10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY