

# International Journal of Engineering Sciences & Research Technology

(A Peer Reviewed Online Journal)  
Impact Factor: 5.164



**Chief Editor**  
**Dr. J.B. Helonde**

**Executive Editor**  
**Mr. Somil Mayur Shah**

## ABSTRACT

Visual speech information plays an important role in lip-reading under noisy conditions or for listeners with a hearing impairment. Correct utterances to read Quran for beginners, there are rules of utterances to learn Quran and we need a software system to tell us if we utter correctly. For that, we built lip-reading model, the model localizes the lips efficiently.

We present in this study a classification model for some al-tajweed rules as we depended on Machine Learning - Cascade Object Detector (Viola-Jones Algorithm), HOG features, a multiclass SVM classifier and Aggregate Channel Features (ACF) object detector for features extraction. We uses Matlab to train a classifiers using a pre-trained convolutional neural network (CNN) for classifying images from the video stream of four different Rules of Holy Quran Allah Elevating (mufakhum), Allah Lowering (mouregeq), sunny ﺍﻻ and moonyﻻ. CNN acquires multiple convolutional filters, used to extract visual features essential for recognizing phoneme. CNNs produce highly accurate recognition results

**KEYWORDS:** Mouth detection, Viola-Jones, HOG + SVM classifier, (ACF), CNN

## 1. INTRODUCTION

In this age of dramatic technology shifts, one of the most significant has been the emergence of digital video as an important aspect of daily life a realization by visionary engineers that, the future of wireless relies heavily on visual communication.

Lip-reading is the task of decoding text from the movement of a speaker's mouth. Traditional approaches separated the problem into two stages: designing or learning visual features, and prediction. Lip readers have enormous practical potential, with applications in improved hearing aids, silent dictation in public spaces, security, and speech recognition in noisy environments, biometric identification, and silent-movie processing. Use the images sequence segmented from the video of the speaker's lips, which is the technique of decoding speech content from visual clues such as the movement of the lip, tongue and facial muscles.

Lip-reading actuations, besides the lips and sometimes tongue and teeth, are latent and difficult to disambiguate without context (Fisher, 1968; Woodward & Barber, 1960). For example, Fisher (1968) gives five categories of visual phonemes (called visemes), out of a list of 23 initial consonant phonemes, that are commonly confused by people when viewing a speaker's mouth. Many of these were asymmetrically confused, and observations were similar for final consonant phonemes.[33].

Also a lot of work focusing on audio-visual speech recognition (AVSR) trying to find effective ways of combining visual information with existing audio-only speech recognition systems (ASR). The McGurk effect [23] demonstrates that inconsistency between audio and visual information can result in perceptual confusion. Visual information plays an important role especially in noisy environments or for the listeners with hearing impairment. [34].

Aiming to help those who are hard of hearing and can revolutionize speech recognition.

An automatic speech-reading, or lip-reading, system as part of an audio-video speech processing (AVSP) system is expected to improve ASR in noisy conditions and can thus lead one step closer to more natural human-computer interactions.

## 2. FRAMING VIDEOS AND PREPROCESSING

Loaded the videos, reading the videos frame by frame, resize each frame in 360 columns, 640 rows, make the videos state by rotate each frame 90 degree and save each frame in specific folders to use it in the face detection stage. Prior to training and testing a classifier, a pre-processing step applied to remove noise artifacts introduced while collecting the image samples. This provides better feature vectors for training the classifier [37]

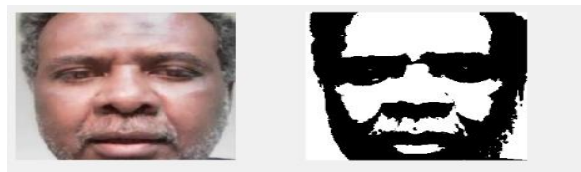


Figure (1.2) Show pre-processing results

## 3. FACE RECOGNITION

Face recognition is the process of identifying people in images or videos by comparing the appearance of faces in captured imagery to a database. Face recognition has many applications ranging from security and surveillance to biometric identification to access secure devices. There is many applications for face recognition:

### 3.1 Face Recognition Workflow

Prepare the database of facial recognition videos of readers we want to recognize also knows the face gallery then perform the processing step of feature extraction to store the discriminative information of each face reader in a compact vector. Following this, we have a learning or modeling algorithm to fit, a model appearance faces in the gallery that can discriminate between different rules in the database. The output of this stage is a classifier a module that will used to recognize the input video. When we have an input query video a face detection algorithm, we used to find where the faces are located in that video. Then crop, resize and normalize the face to match the size and pose images used in the training face gallery. From the same feature extraction step we did in the face gallery, and we run through that classifier /module, the output is the label or an indicator to signify which rule from the database, the gallery, and the query image rule belongs.

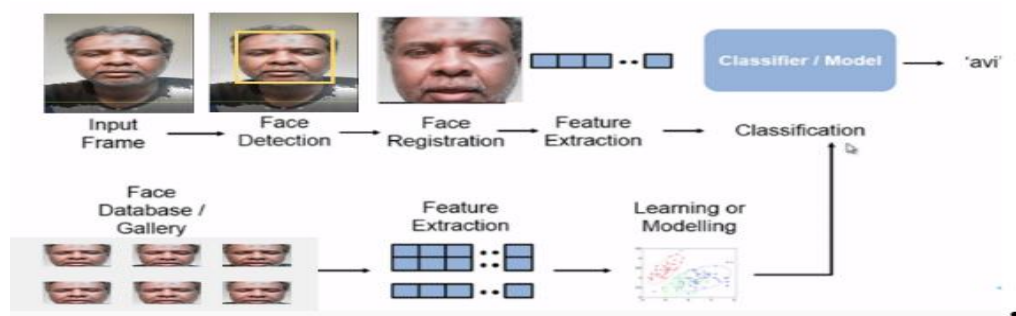


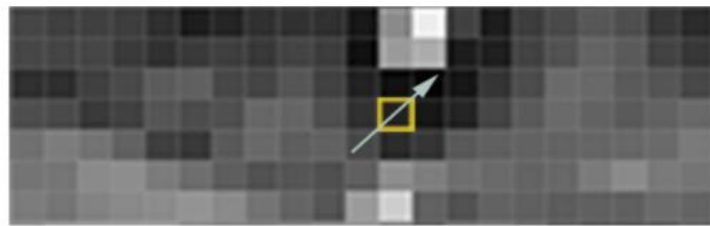
Figure (1.2) Face Recognition Workflow

## 4. FEATURE EXTRACTION

A method of dimensionality reduction that represents the discriminative or interesting parts of an image in a compact feature vector.

#### 4.1 HOG Feature

Face detection went mainstream in the early 2000's when Paul Viola and Michael Jones invented a way to detect faces that was fast enough to run on cheap cameras. We are going to use a method invented in 2005 called Histogram of Oriented Gradients (HOG). To find faces in an image, we will start by making our image black and white because we do not need color data to find faces: Then we will look at every single pixel in our image one at a time. For every single pixel, we want to look at the pixels that directly surrounding it our goal is to figure out how dark the current pixel is compared to the pixels directly surrounding it. Then we want to draw an arrow showing in which direction the image is getting, we will end up with every pixel replaced by an arrow. These arrows are called gradients and they show the flow from light to dark across the entire image but saving the gradient for every single pixel gives us excessively much detail. It would be better if we could just see the basic flow of lightness/darkness at a higher level so we could see the basic pattern of the image.



*Figure (1.2): Image is getting darker towards the upper right.*

To do this, we will break up the image into small squares of 16x16 pixels each. In each square, we will count how many gradients point in each major direction (how many point up, point up-right, point right, etc.). Then we will replace that square in the image with the arrow directions that were the strongest. The result is we turn the original image into a very simple representation that captures the basic structure of a face in a simple way



*Figure (1.2): the original image turned into a HOG representation*

The data used to train the classifier are HOG feature vectors extracted from the training images. Therefore, it is important to make sure the HOG feature vector encodes the right amount of information about the object. The `extractHOGFeatures` function returns a visualization output that can help form some intuition about just what the "right amount of information" means. By varying the HOG cell size parameter and visualizing the result, you can see the effect the cell size parameter has on the amount of shape information encoded in the feature vector.

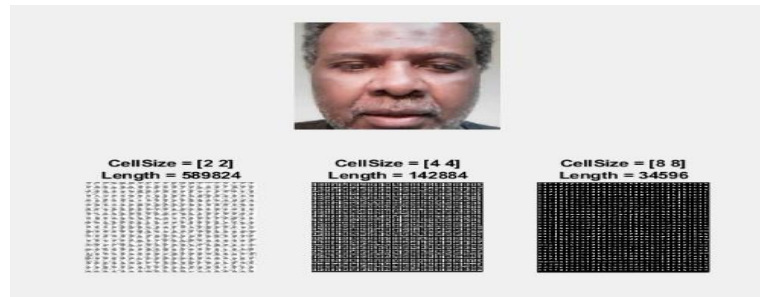


Figure (1.2) Visualize the HOG features

The visualization shows that a cell size of [8 8] does not encode much shape information, while a cell size of [2 2] encodes a lot of shape information but increases the dimensionality of the HOG feature vector significantly. A good compromise is a 4-by-4 cell size. To identify a face shape visually this size setting encodes enough spatial information while limiting the number of dimensions in the HOG feature vector, which helps speed up training. In practice, should be varied the HOG parameters with repeated classifier training and testing to identify the optimal parameter settings.[37]

To find faces in this HOG image, all we have to do is find the part of our image that looks the most similar to a known HOG pattern that was extracted from a bunch of other training faces:

#### 4.2 Support Vector Machine (SVM)

Face recognition systems are definitely still an active area of research. The support vector machine algorithm invented by Vapnik (1998) has been effectively applied to many pattern recognition problems. In this, we are using support vector machine (SVM) as a classifier. What is SVM is a Supervised Machine Learning Algorithm, which solves both the Regression problems and Classification problems. SVM finds a hyperplane that segregates the labeled dataset (Supervised Machine Learning) into two classes.

[39]

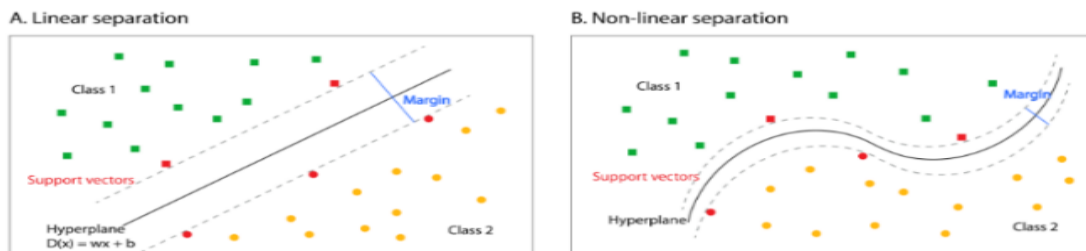


Figure (1.3): SVM ability to Classify linear and Nonlinear Data

#### 4.3 Local Binary Pattern (LBP)

Local Binary Pattern, is a feature descriptor we used it for feature extraction, load the segmentation face regions images you want to calculate LBP features of. Split the data randomly 70% for training, 30% for testing, two new feature vectors was saved (feature\_train, feature\_test)

The local binary pattern (LBP) operator is a gray-scale invariant texture primitive statistic, which has shown excellent performance in the classification of various kinds of textures. For each pixel in an image, a binary code is produced by thresholding its neighborhood with the value of the center pixel (Fig. 1 (a) and Eq. 1).

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p, s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \quad (1)$$

Where,  $gc$  corresponds to the gray value of the center pixel ( $x_c, y_c$ ) of the local neighborhood and  $gp$  to the gray values of  $P$  equally spaced pixels on a circle of radius  $R$ . By considering simply the signs of the differences between the values of Neighborhood and the center pixel instead of their exact values, LBP achieves invariance with respect to the scaling of the gray scale.

Local Binary Pattern (LBP) is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel with the value of the center pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Perhaps the most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes caused, for example, by illumination variations. Another important property is its computational simplicity, which makes it possible to analyze images in challenging real-time settings. [30]

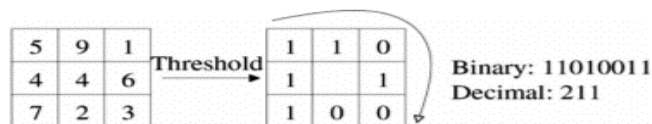


Figure (1.3): Example of the basic LBP operator

Face analysis using local binary patterns • Face recognition is one of the major challenges in computer vision a face descriptor based on LBPs •has, excellent results in face, recognition

It was a powerful feature for texture classification. It has further been determined that when LBP is combined with histograms of oriented gradients (HOG) descriptor, it improves the detection performance considerably on some datasets.

#### 4.4 SVM Train

The LBP feature are extracted from dataset in training and validation set and saved in the features\_test and features\_train respectively ,dct applied the features\_test and features\_train The DCT block computes the unitary discrete cosine transform (DCT) of each channel in the M-by-N input matrix, u ,selecting the labels name(name of the folders) SVM will be train using (error correction out put code ) fit cecoc to solve multi selection proplem .evaluate the model to see performance of the model , plot the confussion matrix , calculate the accuracy then calcaluate the precision .

The model for two data set are well-trained, best accuracy, very good data.



Figure (1.3): LBP operator Feature Extraction

#### 4.5 LIP Detection and Localization

This section tries to determine the best feature and color component to use for a lip detector. The lip detectors were made by training cascaded classifiers. This approach depends on building lip model(s), with or without using training face images and subsequently using, the defined model to search for the lips in any freshly input image, some images cant cropping mouth with cascade object detector (viola jones ) .

##### 4.5.1 Viola Jones algorithm

Viola-Jones is quite powerful, and its application has proven to be exceptionally notable in real-time face detection. Has four main steps:

1. Selecting Haar-like features
2. Creating an integral image
3. Running AdaBoost training
4. Creating classifier cascades

The job of the cascade is to quickly discard non-faces or non-mouth, and avoid wasting precious time and computations. Thus, achieving the speed necessary for real-time face / mouth detection. We set up a cascaded system in which we divide the process of identifying a face/mouth into multiple stages. In the first stage, the sub region passes through the best features. In the next stages, we have all the remaining features. When an image sub region enters the cascade, it evaluated by the first stage. If that stage evaluates the sub region as positive, meaning that it thinks it is a face/ mouth, the output of the stage is maybe. When a sub region gets a maybe, it sent to the next stage of the cascade and the process continues as such until we reach the last stage. If all classifiers approve the image, finally classified as a human face/ mouth and presented to the user as a detection. [43].

Show that the method localizes the lips efficiently, with high level of accuracy (91.15%).

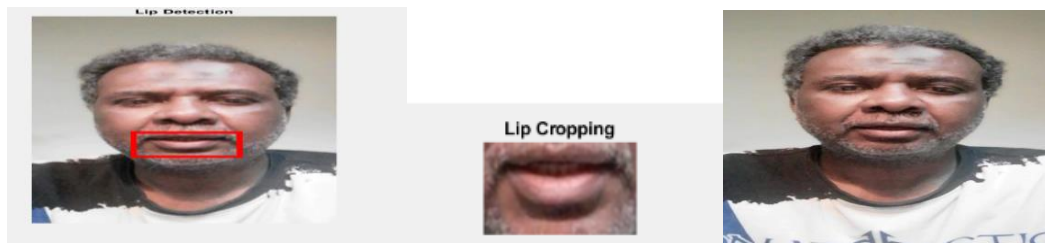


Figure (1.3): Lip Detection

Figure of Mouth not cropping

#### 4.5.2 Detect objects using ACF object detector

This MATLAB function detects objects within image we using the input aggregate channel features (ACF) object detector, its two parts Training part and testing Detect and label people using aggregate channel features (ACF). This algorithm is based on the peopleDetectorACF function. To use this algorithm, you must define at least one rectangle ROI label. You do not need to draw any ROI labels.[44] .



Figure (1.3): Classification accuracy

##### 4.5.2.1 Image labeler

The Image Labeler app provides built-in algorithms that you can use to automate labeling. From the app tool strip, click Select Algorithm and then select an automation algorithm.



Figure (1.3): Image Labeler

##### 4.5.2.2 ROI and Scene Label Definitions

- An ROI label corresponds to either a rectangular or pixel region of interest. These labels contain two components: the label name, such as "cars," and the region you create.

- A Scene label describes the nature of a scene, such as "sunny." You can associate this label with a frame.[44]

**5. CONVOLUTIONAL NEURAL NETWORK (CNN OR CONVNET)**

Is a network architecture for deep learning which learns directly from data, eliminating the need for manual feature extraction.

CNNs are particularly useful for finding patterns in images to recognize objects, faces, and scenes. They can also be quite effective for classifying non-image data such as audio, time series, and signal data.

Applications that call for object recognition and computer vision such as self-driving vehicles and face-recognition applications rely heavily on CNNs.

**5.1 Important factors of Using CNNs for deep learning:**

- CNNs eliminate the need for manual feature extraction—the features are learned directly by the CNN.
- CNNs produce highly accurate recognition results.
- CNNs can be retrained for new recognition tasks, enabling you to build on pre-existing networks.[41]

**5.2 Feature Learning Layers and Classification**

Like other neural networks, a CNN is composed of an input layer, an output layer, and many hidden layers in between.

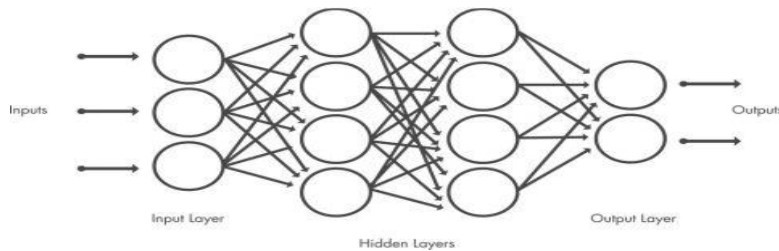


Figure (1.3): Convolutional Neural Network Layers

These layers perform operations that alter the data with the intent of learning features specific to the data. Three of the most common layers are: convolution, activation or ReLU, and pooling.

**5.3 Transfer learning workflow**

Load network, replace layers, train network, and assess accuracy.

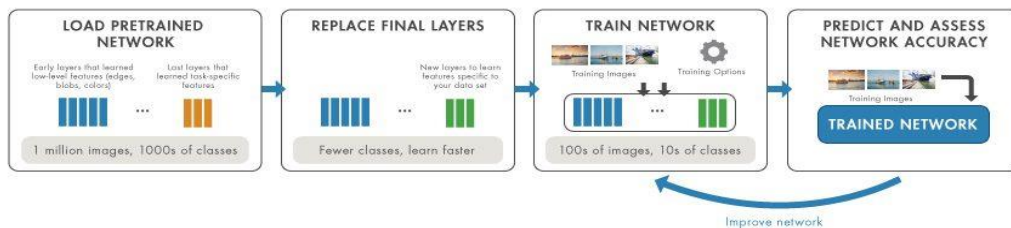


Figure (1.3): Transfer Learning Workflow

**5.4 Database and Experiments**

We have collected 4 rules of holy Quran with 29 subjects (rules), each subjects recorded 4 times the data base collected 116 words large-vocabulary, single words audio-visual speech database, using Hawaii camera. The database contains full frontal face color video of the subjects with minor face-camera distance and lighting variations (see Fig. 2). The video is captured at a resolution of 280 × 720 pixels (interlaced), a frame rate of 30 Hz (i.e., 60 fields per second) The audio is captured in a relatively “clean” office environment, at a sampling rate



of 16 KHz, and it is time-synchronous to the video stream. database videos are randomly split them into 70% training, 30% test.

**Implementation and Result**

The 7 rules data set \_1 are trained and identification achieved

- 1- for Allah Elevating (mufakhum) there is 3 cases in Quran each case recorded 4 times accuracy = 100% , Recall = 99.5000, Precision =100, elapsed time is 35 min and 18 Sec , Epoch = 20 of 20, Iteration 260 of 260

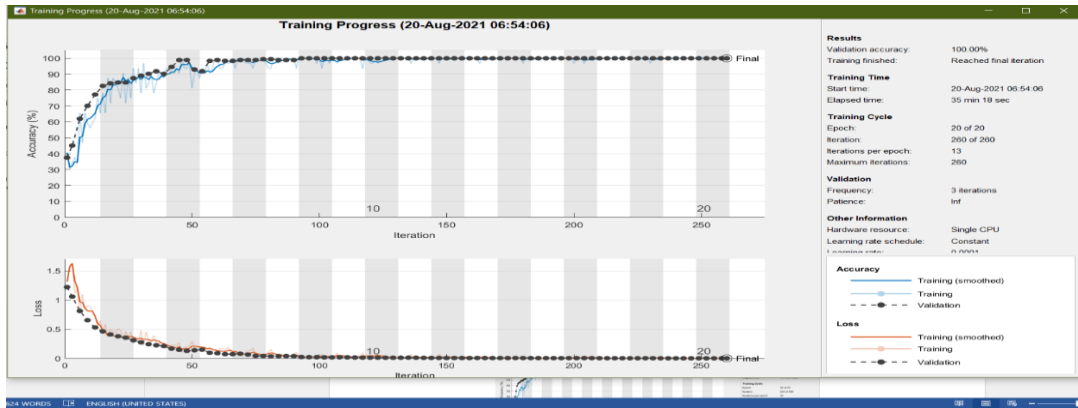


Figure of CNN Training Progress Allah Elevating (mufakhum)

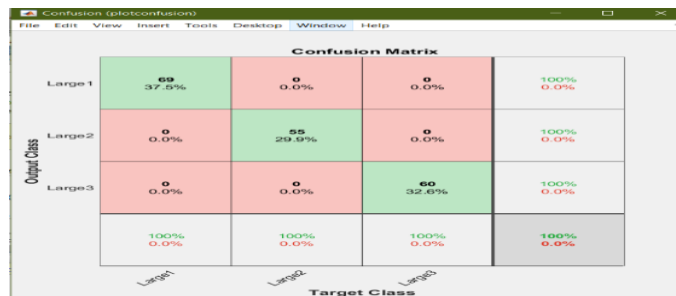


Figure Confusion Matrix of CNN Training Progress Allah Elevating (mufakhum)

- 2- for Allah Lowering (moureqeq) there is 4 states in Quran each state recorded 4 times accuracy = 98.84%, Recall = 98.5088, Precision = 99.0213, elapsed time is 50 min and 25 Sec , Epoch = 20 of 20, Iteration 360 of 360

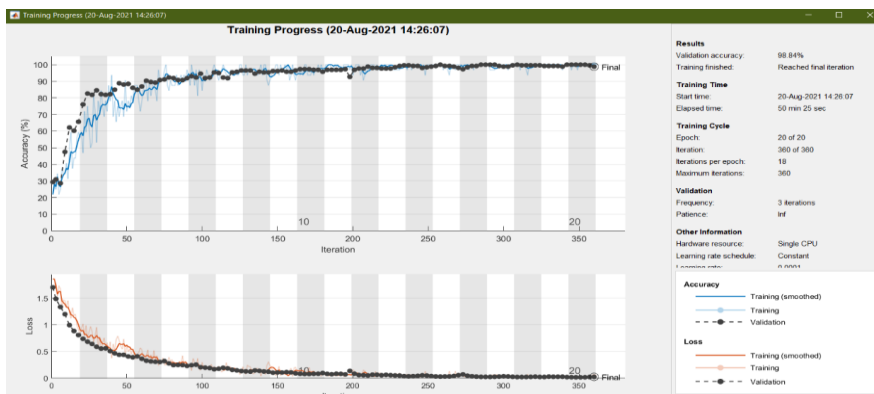
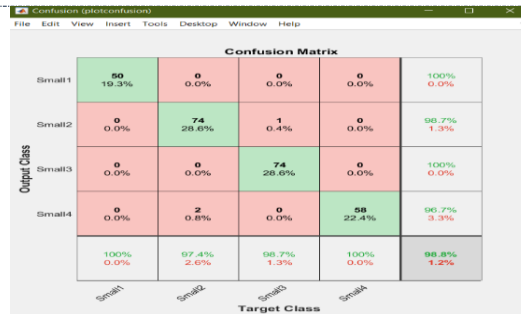


Figure of CNN Training Progress Allah Lowering (moureqeq)



Output Class	Small1	Small2	Small3	Small4	Accuracy
Small1	89 19.3%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
Small2	0 0.0%	74 28.6%	1 0.4%	0 0.0%	98.7% 1.3%
Small3	0 0.0%	0 0.0%	74 28.6%	0 0.0%	100% 0.0%
Small4	0 0.0%	2 0.8%	0 0.0%	58 22.4%	98.7% 3.3%
Target Class	100% 0.0%	97.4% 2.6%	98.7% 1.3%	100% 0.0%	98.8% 1.2%

Figure of CNN Training Progress Allah Lowering (moureqeq)

## 6. RESULTS

We choose MATLAB version 9.5.0.944444 (R2018b) as our programming environment as it offers many advantages. It contains a variety of image processing

All experiments emphasize the superiority of the visual words approach as a solution for the visual speech recognition problem. However, many challenges still remain in this area of research, particularly the large unseen part of speech, as the word recognition rate is greatly affected by its phoneme components, because producing these sounds involves the movement (dynamic) of some parts of the speech production system, and some of these parts can be seen (lips and mouth) and others cannot

In this research has been done to the performance of a face detection to detect mouth to extract visual words system by making use of feature extraction with Histogram of Oriented Gradients (HOG). Features extractions +support vector machine classifier (SVM) and Local Binary Patterns. (LBP).

It mainly consists of four parts, namely face detection, mouth detection and cropping, and feature extraction and classification. Face detection represents how the model detect a face and a mouth, determines the successive algorithms of detection and recognition. The most useful and unique features of the face image / mouth image are extracted in the feature extraction phase. In the classification, the mouth image is determine the correct word (word of the rule). The accuracy of the system for face detection is above 100% by Histogram of Oriented Gradients (HOG). Features extractions +support vector machine classifier (SVM) and the Local Binary Patterns algorithm, the accuracy of the system is above 100%.

Moreover, Cascade Object Detector (Viola-Jones Algorithm), Show that the method localizes the lips efficiently, with high level of accuracy (91.15%).

The accuracy of the system is low scores detection by Aggregate Channel Features (ACF).

Many difficulties has been faced for any VSR system to detect/localize such regions to capture the related visual information, such as we cannot read lips without seeing them first. Therefore, lip localization is an essential process for any VSR system . The lips and mouth region are the visual parts of the human speech production system; these parts hold the most visual speech information difficulties to capture the related visual information, i.e. pose and lighting variations, , and lips occlusions , Blurring lips when pronouncing some letters.

In future to improve the our system performance, Vision Transformers (ViTs) techniques can be combined with Convolutional neural networks (CNNs),in the system for video-based visual speech recognition on real time.

## REFERENCES

- [1] Mary Hepburn parsons, "The reading of Speech from the lips", Gallaudet college Kendall green Washington, 9, 1900.
- [2] D.G. Stork and M.E. Hennecke, "Speech reading by Humans and Machines", Berlin, Germany: Springer, 1996.
- [3] W. H. Sumby and I. Pollack, "Visual contributions to speech intelligibility in noise," Journal of the Acoustical Society of America, 26:212–215, 1954.

- [4] Jie yang, Alex waibel, "Tracking Human Faces in Real -Time ", Pittsburgh, pennsylvania 15213, 1995.
- [5] Robert august Kaucic Jr, "Lip tracking for Audio-visual Speech recognition", Merton college University of oxford, 1997 .
- [6] Sharmila Sengupta , Arpita Bhattacharya , Pranita Desai , Aarti Gupta "Automated Lip Reading Technique for Password Authentication", New York, USA, 212 .
- [7] Md. Hasan Tareque1, Ahmed Shoeb Al Hasan, "Human Lips-Contour Recognition and Tracing", (IJARAI) International Journal of Advanced Research in Artificial Intelligence, Vol. 3, No. 1, 2014.
- [8] Alin Chițu, Léon J.M. Rothkrantz, Zaidi Razak, Zulkifli Mohd Yusoff,
- [9] Kuniaki Noda , Yuki Yamaguchi , Kazuhiro Nakadai , "Lipreading using Convolutional Neural Network", Japan, 2014 .
- [10] Ikrami A. Eldirawy, "Visual Speech Recognition", Islamic University of Gaza, May 2011.
- [11] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal
- [12] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal
- [13] T. Chen and R.R. Rao, "Audio-Visual Integration in Multimodal Communication" , Proc. IEEE, vol. 86, no. 5, pp. 837-852, May 1998
- [14] David G. Stork, Marcus E. Hennecke, "Speech reading by Human and Machines", Ricoh California Research center 2882 Sand Hill Road # 115 Menlo Park, 94025-7022, USA .
- [15] Audrey R. Nath and Michael S. Beauchamp , "A Neural Basis for Interindividual Differences in the McGurk Effect, a Multisensory Speech Illusion", PMC3196040, 20 Jul 2012.
- [16] K. Neely. "Effect of visual factors on the intelligibility of speech," Journal of the Acoustical Society of America, 28(6):1275-1277, 1956.
- [17] C. Binnie, A. Montgomery, and P. Jackson, "Auditory and visual contributions to the perception of consonants," Journal of Speech Hearing and Research, 17:619-630, 1974.
- [18] D. Reisberg, J. McLean, and A. Goldfield, "Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli, " In B. Dodd and R. Campbell, "Hearing by Eye, " pages 97-113. Lawrence Erlbaum Associates, 1987.
- [19] K. P. Green and P. K. Kuhl, "The role of visual information in the processing of place and manner features in speech perception," 45(1):32-42, 1989.
- [20] D. W. Massaro, "Integrating multiple sources of information in listening and reading, " In Language perception and production. Academic Press, New York.
- [21] R. Campbell and B. Dodd, "Hearing by eye," Quarterly Journal of Experimental Psychology, 32:85-99, 1980.
- [22] Yannis M. Assael1, Brendan Shillingford1, Shimon Whiteson, "LIPNET: END-TO-END SENTENCE-LEVEL LIPREADING" , 3 Department of Computer Science, University of Oxford, Oxford, UK 1 Google DeepMind, London, UK 2 .
- [23] Alin Chițu and Léon J.M. Rothkrantz, "Automatic Visual Speech Recognition
- [24] ", Delft University of Technology, Netherlands Defence Academy, the Netherlands.
- [25] Achraf Ben-Hamadou, Walid Mahdi, ahmed rekik's "An adaptive approach for lip-reading using image and depth data", Multimedia Tools and Applications · July 2015.
- [26] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", 3rd Edition. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [27] F. G. Smith, K. R. Jepsen, and P. F. Lichtenwalner, "Comparison of neural network and Markov random field image segmentation techniques" , in Proceedings of the 18th Annual Review of progress in quantitative nondestructive evaluation, vol. 11, 1992, pp. 717-724.
- [28] A. Blake and M. Isard, "Active Contours, Springer", 1998.
- [29] J. Shi and J. Malik, "Normalized cuts and image segmentation", in CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97). Washington, DC, USA: IEEE Computer Society, 1997, p. 731.
- [30] J. Sethian, "Level set methods and fast marching methods: Evolving interfaces in computational geometry," 1998.
- [31] D. Reisberg, J. McLean, and A. Goldfield, "Easy to hear, but hard to understand: A lipreading advantage with intact auditory stimuli, " In B. Dodd and R. Campbell, "Hearing by Eye " , pages 97-113. Lawrence Erlbaum Associates, 1987.

- [32] G. W. Greenwood, "Training partially recurrent neural networks using evolutionary strategies", *IEEE Trans. Speech and Audio Processing*, 5(2):192–194, 1997.
- [33] E. Owens and B. Blazek, "Visemes observed by hearing impaired and normal hearing adult viewers", *Journal of Speech Hearing and Research*, 28:381–393, 1985.
- [34] Ikrami A. Eldirawy "Visual Speech Recognition", Islamic University of Gaza, May 2011
- [35] Ahmad Basheer Hassanat, "Visual Words for Automatic Lip-Reading", University of Buckingham United Kingdom, December 2009
- [36] Yuanhang Zhang, Shuang Yang, Jingyun Xiao, Shiguang Shan, Xilin Chen, "Can We Read Speech Beyond the Lips? Rethinking RoI Selection for Deep Visual Speech Recognition", arXiv: 2003.03206v2 [cs.CV] 9 Mar 2020.
- [37] Elena Alionte, Corneliu Lazar, "A Practical Implementation of Face Detection by Using Matlab Cas,cade Object Detector", *International Conference on System Theory, 2015 19th Control and Computing (ICSTCC)*, Cheile Gradistei, October 14-16, Romania
- [38] Guoying Zhao, Mark Barnard, "Lip-reading with Local Spatiotemporal Descriptors", *IEEE Transactions on Multimedia* · December 2009
- [39] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection", *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 886-893, 2005.
- [40] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, "Gradient-based learning applied to document recognition.", *Proceedings of the IEEE*, P. (1998). 86, 2278-2324.
- [41] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, "Reading Digits in Natural Images with Unsupervised Feature Learning NIPS Workshop on Deep Learning and Unsupervised Feature Learning", 2011.
- [42] Lindsay I Smith, "A tutorial on Principal Components Analysis", February 26, 2002
- [43] Sushma Ronanki, Sonia Gundu, Rupavathi Baratam, P Mounika, J Rajesh Kumar, "FACE DETECTION AND IDENTIFICATION USING SVM", B.TECH, Electronics and Communication, SSCE, Srikakulam, Andhra Pradesh (India), April 2017 .
- [44] [https://www.mathworks.com/videos/introduction-to-deep-learning-what-is-deep-learning--1489502328819.html?s\\_tid=vid\\_pers\\_recs](https://www.mathworks.com/videos/introduction-to-deep-learning-what-is-deep-learning--1489502328819.html?s_tid=vid_pers_recs)
- [45] <https://www.mathworks.com/discovery/convolutional-neural-network-matlab.html>
- [46] [https://www.mathworks.com/discovery/machine-learning.html?s\\_tid=srchtitle](https://www.mathworks.com/discovery/machine-learning.html?s_tid=srchtitle)
- [47] <https://www.mygreatlearning.com/blog/viola-jones-algorithm/>
- [48] <https://www.mathworks.com/help/vision/ref/imagelabeler-app.html>