limited Arabic Alphabet Symbols Based on Lips Movement

Authors Names

Hassan M. N. Mohammed Zaki [a],
Khalil I. Alsaif [b]

Article History
Accepted on: 13/ 8 / 2023
Keywords: *Computer vision, digital image processing, machine learning, deep learning, artificial intelligence, communications security*

ABSTRACT

In this research, the development of deep learning algorithms was relied upon and employed in the field of computer vision for the purpose of building symbols for the Arabic alphabet with a limited capacity to cover specific requirements.

The network (CNN) is adopted for the purpose of tracking the movement of the lips when pronouncing selected words in the Arabic language by showing the borders of the lips (Lips land marks) and calculating the distance between the lips to determine the state of the mouth (open / closed) when forming a code for the selected words from the Arabic language and according to work requirements Where a special binary code is generated by tracking the movement of the lips, since the closed mouth is zero and the open mouth is the one as a basis for building codes for those movements.

I. Introduction

Relying on understanding speech through lip reading has been of great importance to researchers for a long time. Where speech is understood based on the visual information acquired from a scene that includes the movement of the lips in addition to the skill of the lip reader. Lip reading is a complex problem because of the many influences involved in determining the process of reading and understanding what is intended by the speaker's lip movements. Lip reading plays an important role in human communication, in addition to helping people with hearing impairment or speech impairment. Digital technologies have played a prominent role in this field, helped by the development of deep learning algorithms and their use in many areas of public life.

The positioning of the lip region is the main input in the process of lip reading, so accurate positioning is very important to extract features related to visual speech. The biometric features of the lip region and the physiological features in addition to the behavioral features are adopted in order to translate the lip movement into visual speech [1]. From a physiological point of view, there is a clear variation in the lips of people, in addition to the difference in behavioral speech style from one person to another.

II. Previous Literature

Detecting and identifying the face region, Képešiová & Kozák In 2018 carried out a study of the facial area where the size and condition of all parts of the face were determined, for example, the area of the eyes, nose, mouth and eyebrows. (Viola Jones) was used to detect the face area and (Cascade) for classification and Adaboost to train the algorithm that was adopted[2].

El-Melegy et al. also presented a study in 2019 they divided the work into several steps starting from the stage of entering the image into three types of face detection systems, where the researcher used the detector Dlib HOG face, Then faces are detected in a specific way by inserting it into another

Hassan M. N. Mohammed Zaki, *Col. of Fine Art. /Mosul University. Mosul, Iraq E-mail: hmn1973@uomosul.edu.iq*
Khalil I. Alsaif, Tech. Comp. Eng./Hadbaa Col. Univ. Mosul, Iraq E-mail: khalil_alsaif@hcu.edu.iq

detector and finally the Open-CV Ha detector is used as a basic factor to extract the human face area [3].

A study was conducted based on the method of computer vision in order to read the lips of people from Qail Soundarya et al. in 2020[4] by looking at the movement of the lips, and they gave a description of the difficulties they encountered when designing a lip reading system. The proposed system was based on a convolutional neural network with a hidden Markov model (CNN with HMM). The Hidden Markov Model (HMM) was adopted as a training trait for speech recognition.

In 2021, the researcher [5] presented a survey on the current understanding of deep learning, in addition to the challenges and methods used to solve those problems in the field of artificial intelligence, especially in deep learning. They also presented an improvement on computer models for deep learning, in addition to applications in the fields of visual object recognition, faces and speech, especially in the field of personal identification.

The researcher [6] presented a summary of developments in the field of deep learning by reviewing eight recent technologies: AlexNet, VGGNet, GoogLeNet ResNet, DenseNet, MobileNets, EfficientNet, and RegNet with its applications in visual speech recognition, in addition to checking the performance of these technologies.

In a study presented by Dey et al. In 2022, the researcher relied on the Dlib library[7], which uses deep learning algorithms to detect faces and uses Dlib CNN with Dlib-HOG. Through this study, it was found that Dlib CNN detects faces from different angles with long time, but better results were obtained even in the case of incomplete image input.

Lip reading and visual information extraction need to accurately identify the lip movements due to the different shape of the lips and because of the many expected movements of the lips when the word is pronounced. So a literature on Localization of the mouth based on facial features is essential:

The researchers [8] presented a method based on the combination of a color space of YCbCr and CIELa * b * with the Moore Nhbor algorithm to locate the lip region by using the chromatic transformation technique to extract the lip region and then extract and track the contours of the lips. The researcher relied on the Moore Neighbor Algorithm (MNT), in which the coordinates of twelve points were adopted on each of the upper and lower lip borders.

As for [9], in their study, deep learning techniques were adopted in order to localize the mouth to facilitate the process of extracting the different features of the movement of the lips. One of the proposed models is to use the Dlip and OpenCV libraries to detect the face by predicting the shape based on the points of interest in it, which consist of facial features such as eyes, mouth and nose area, and other facial features. The main step in this research is to detect the main facial points in the initial stage, and then the mouth region is determined based on the previous points, followed by the introduction of a deep learning approach to produce the corresponding sentence.

Parra-Dominguez et al. A set of measurements used to design the shape of the face without relying on facial movement and relying on facial features to calculate those measurements [10]. The proposed method determines the perimeter of a part of the face using the Euclidean distances between two places in the face properties and then measures the degree of asymmetry between the two sides. The proposed system consists of three models: the first is the extraction of facial features, the second is the calculation of facial measurements, and finally the classification of paralysis in the facial area and the determination of levels of asymmetry within the facial elements.

III. Features Geometry in Lips Reading

The interest of researchers in the topic of lip reading has led to many applications, where automatic lip reading has gained great importance, especially in speech recognition system in noisy environments, where the visual signal provides complementary information to the audio signal. The geometric features of the lip region provide valuable information for speech recognition extracted from the movement of the lips during speech. The following is a brief survey of the literature that adopted the geometric features of the lips in reading visual speech:

A system that relied on a set of frames extracted from a video clip of a person speaking silently. During that research[11], an innovative method was proposed by adopting techniques to isolate the lip area in order to reach the geometric visual features of the movement of the lips, which represent a characteristic of the speech spoken by the speaker. The geometric feature of the lip region was represented by a set of key points surrounding the lip region, from which the height and width of the mouth region were calculated. These features were processed using a digital virtual generalization classifier (KNN), which was used to determine the speech spoken by the speaker.

Ivanko et al at 2019[12] they presented research in helping the deaf and hard of hearing in order to activate their skills to speak with each other. To automatically increase accuracy and reliability, the researcher used the lip reading method in addition to hand gestures (RusLan database), which was collected in the year 2018-2019 in SPIIRAS and consists of 13 video recordings in Full HD. The images available in the database were used to train the visual speech recognition system, and the main step in the system is to search for the region of interest represented by the region of the moving lips. Two methods were proposed to search for this region. The algorithm was implemented using AAM to detect the face and mouth region, where the main idea is to extract the shape and appearance of the object, depending on the predictor in the Dlip library where the five point features are determined by adopting 68 main points for the face, including 20 points for the mouth region. After extracting the geometric features of the mouth region and after interpolating the coordinates of 20 points for the features of the mouth region, the distance between the points of the mouth was measured according to the Euclidean distances.

At 2022, the researcher Zhu et al.[13] Adopt a study to design a system that detects driver fatigue based on facial features through videotapes. Where a model was built to assess drivers fatigue based on geometric measurements, the eye aspect ratio (EAR) and the mouth aspect ratio (MAR) were calculated. The system is built on the basis of a deep convolutional network that detects facial regions to solve the face convergence problem. The geometric feature of the mouth was extracted from the ratio of the mouth's height to its width. The convolutional network used to detect the face area is based on the 68 initial points in the facial features, and the K-Nearest Neighbor (KNN) algorithm is used to classify the driver's condition by calculating (EAR) in addition to (MAR) and adopting the standard measurement of the Euclidean distance between the points of the eye features in addition to the distance between landmark points in the mouth. The system gave, based on engineering measurements, a definition of the driver's fatigue state, such as yawning and resting.

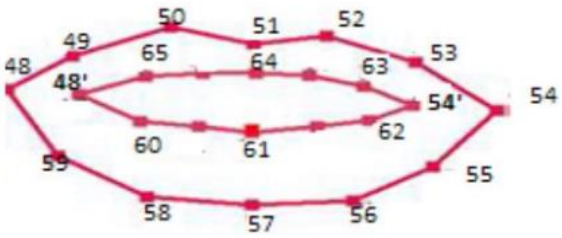
Measuring the distance between the upper and lower lip corners leads to the extraction of lip geometry which are important features in noisy environments and are desirable features as a visual feature [14]. Also the talking mouth image analysis focuses on:

• The height and width of the borders of the outer and inner lips during the proposed opening of the mouth and its extension.

• Variation in the thickness and shape of the lip while pronouncing the word.

The distance is measured in Euclidean space between any two points on the inside and outside of the lip circumference. This process describes the width and height between the outer and inner upper and lower lip borders, from which the thickness of the lips is calculated. Figure 1 illustrate the lips Landmark based on the geometric features of 68 pixels as key feature points.

Fig. (1): lips Landmark Relying on the geometric features of 68 pixels



The contour points of the lips are determined as follows:

• External: Point 48 is the corner of the left mouth, Point 54 is the corner of the right mouth, Point 51 is the middle of the upper lip and represents the meeting point of the vertical groove between the border of the upper lip with the base of the nose, Points 50 and 52 at the bottom of the columns. Points 56, 57 and 58 lie on the underside of the lip, as opposed to Points 50, 51 and 52 on the upper lip. In contrast, points 49, 53, 55 and 59 are placed in the position covering the entire lip area. The mechanism for placing these points at equal distances from the points next to them.

• The inner lip circumference points are somewhat similar to the outer lip circumference points, which should accurately describe the inner mouth opening. Since point 48 is the left inner lip angle and corresponds to point 48 of the outer lip, point 54 is the right inner lip angle. It corresponds to point 54 of the outer lip. Points 61 and 64 are located in conjunction with points 51 and 57, as these four points lie on the same vertical plane below the philtrum.

$$D(l_1, l_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad \textbf{(1)}$$

While the remaining pairs of points are represented by points 62 and 63, 60 and 65 on the inner lip region at equal distances. As seen in Figure (1), the feature-dependent distance is calculated using the Euclidean distance (ED) from equation (1).

where D: The distance between the two primary points, l1(x1, y1), l2(x2, y2):two essential facial landmark points.

IV. Deep Leaning

A specialized branch of the machine learning field, inspired in the 1950s by models of computation and human cognition [15], the field of deep learning represents a small branch in the artificial intelligence, this field has brought fundamental changes in several areas, DL field created by the neural network (NN) sub-field on it and developed in whole area of science as seen in Fig(2) which show a comprehensive overview of deep learning techniques for classification and practical applications. Since its inception, DL has caused even greater success in every field of application.

The DL can be either a hierarchical learning models or a deep construction method; learning means a method used to estimate the parameters of the model so that an algorithm is designed to perform a specific task as best as possible; on the other hand, DL has numerous layers between input-output layers which allow many non-linear processing units for information in several stages hierarchical structure which is used for feature learning purpose. Learning methods based on data representation are known as representation the learning, and some recent literature described DL as a learning model that is comprehensive and capable of providing solutions to all kinds of problems in various scientific fields [16].

In the design of this research, the deep network (CNN) of the LeNet type was adopted, with minor direct changes made to the network structure. In order to capture large numbers of complex patterns, the number of filters used in the deep layers has been increased, in addition to the use of Adam's optimizer instead of SGD to speed up network training. The first Convolutional layers in the LeNet are adopted for as Edge detectors. For mouth direction classification issues for each category, the location of the mouth is relative to the face components. A CNN model with three convolution stages was employed in this research methodology. Images of the mouth region with a fixed dimension make up the input stage.

There are 32,48,64 filters of dimension 5×5 employed in the three convolutional layers. These steps were followed by a rectifier linear unit (ReLU). A 2×2 max pooling layer is placed after each of the three convolutional layers. All of the activations' outputs are combined into a layer that is fully connected. There are five kinds of output in the system with respect to the eye pupil directions as shown in Fig. (3). The softmax loss is used over classes as the error measure.

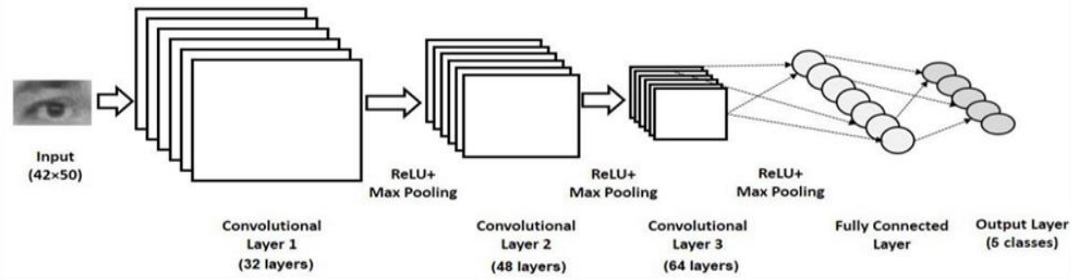


Fig. (2) CNN structure

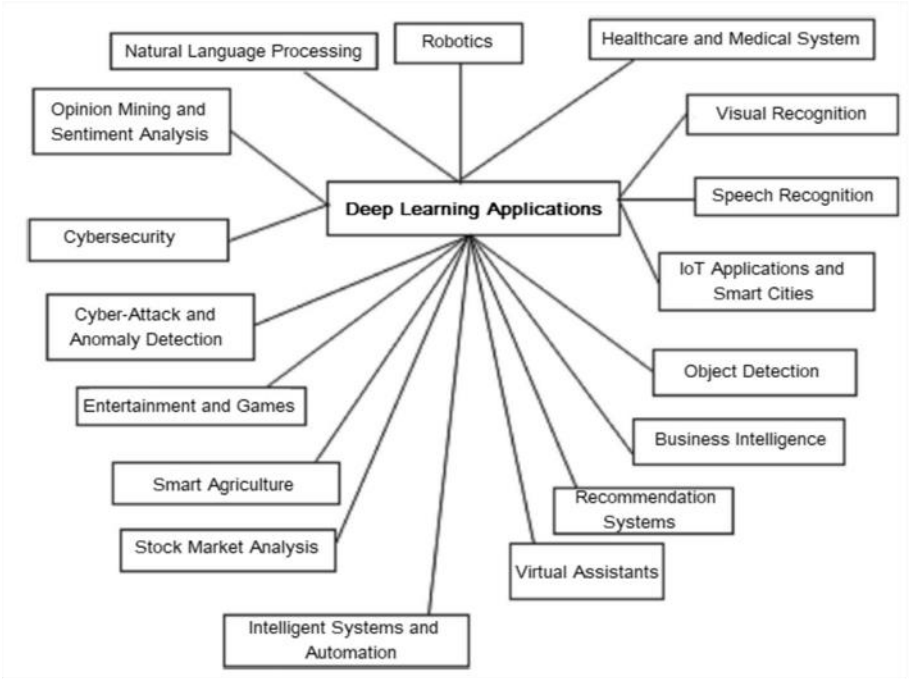


Figure (3): Deep Learning Techniques, Taxonomy, Applications and Research Directions [21]

Representational learning is a set of methods that feed pure data to a machine to discover the representations needed by the classification and discovery process. Deep learning methods are representational learning modes with several levels of representations obtained from simple non-linear units that transform representations from lower to higher levels in a hierarchical manner and, with such enough transformation, can learn highly complex functions. Deep learning relies on data-driven learning algorithms to create

feature layers, these features are represented by several levels in a hierarchy; this is the point of strength compared to the traditional methods used in machine learning[17].

Deep learning provided the most appropriate solutions to various problems in several fields through the adoption of deep learning algorithms in the field of science and government business; Here are some of these capabilities:

• The ability to simulate human behavior (understanding language, image recognition and speech recognition).

• The ability to devise solutions to changing problems over time (tracking, weather forecasting, anticipating price changes).

• Solutions presented to solve some special cases and how to adapt these solutions to be compatible with those special cases (biometrics, prediction of the effect of mutations in nucleic acids).

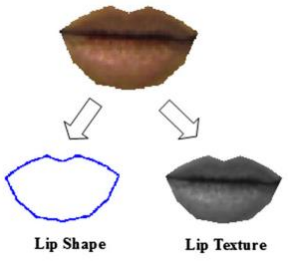
Some problems of a sensory or emotional nature cannot be imagined and are the capabilities of the human mind (for example, how to analyze feelings. Deep models are highly flexible in representing the nonlinear functions that make up a large part, in addition to innovative approaches that have been adopted for the efficient training of those networks with high capabilities; In addition, it adopts various regulatory techniques that ensure that the deep model that contains huge numbers of statistically desirable parameters that reflects positively on the unseen data[18].

V. Visual Feature Extraction

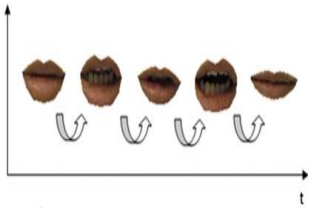
Extraction of visual feature is the basis of research in the field of visual recognition of speech, and the purpose of extracting these features is to obtain as much speech information as possible from the speaker's image and within a reasonable range of parameters[18]. After detection, extract the lip region, followed by extracting the visual features from this region which represents the region of the interest since it contains the information of visual speech and because the video frame series may contain a lot of additional information such as lighting, skin color, etc., that is not Related to the lip-reading assignment, so how to extract the features which related to the visual lip-reading task from the sequenced lip images is the key to recognition visual speech, on the other hand

A. Discriminative Analysis of The Lip Movement Features for Reading The Speech

The process of extracting facial features is a difficult problem because of the great differences between people in terms of appearance, in addition to the change in appearance during speech. What makes image analysis more difficult is the difference in face poses and environmental conditions in terms of the difference in lighting. In recent years, the features of the lips extracted during the pronunciation process have been considered a new feature of biometrics, where the features of the lips have two characteristics: First, the feature of the lips possesses physiological information represented in the shape and appearance of the lips, while the second, lips feature possesses the behavioral information of changing the lips during pronunciation. Before conducting a comprehensive biometric analysis of the lips, the physiological and behavioral characteristics of the lips must first be determined, and the description provided by the physiological features takes into account the physiological characteristics of the speaker's lips, which are usually instant of time(both the shape and appearance of the lip representing a moment in time), without any kind of temporal information, these features are extracted from the isolated images of the lips, where the behavioral features described the deformation of the lips due to the lips



(a) lip Physiological features



(b) lip behavioral features

Fig. (4) illustrates the difference in lip features between physiological and behavioral.

movement during pronunciation, knowing that it is dynamic and possesses temporal information, and these features are extracted from the sequence of the lip images (features represent the movement of the lips over time during speech.), these types of lip region features can provide discriminatory information for visual speech. Fig. (1) illustrate the difference in lip features between both physiological and behavioral [20].

1) Physiological Features of The Lip:, The physiological features of the lip region can be obtained by adopting techniques based on the geometrical features of that region, given that some measures assumed the height or width of the mouth region opening as an important feature in tracking visual speech, or adopting model-based techniques, where the model is built based on what gives the details of visual speech, the contour of the shape lip region is usually adopted and described by a set of parameters assume that visual speech information contains in lips contour.

Another physiological feature of the lip lies in the different intensity that is distributed within the speak lips region, and this information is known as the lips texture. The physiological features of the lips consist of both the shape and the texture of the lips. Lip geometry is the appropriate technique for modelling the movement of the lips, especially in the problem of speech reading, but on the other hand, it is considered one of the difficult tasks as the algorithms used They are very sensitive to lighting conditions and image quality.

2) Behavioral Features of The Lips, The behavioral features of the lips are divided into the shape of the moving lips and the texture of the moving lips. The main difference of this type from the physiological features of the lips is that these features capture as well as describe only the dynamic information of the lips during speech (deformation information of the lips). Therefore the behavioral features extraction needs a series of lip images instead of isolated images, as well as the behavioral features, are a sequential time signal.

Features of lip movement when reading speech: The large differences between people in terms of appearance, in addition to the change in appearance during speech, made the process of extracting facial features a difficult problem, as well as the process of analyzing images more difficult, in addition to the facial conditions and environmental conditions in terms of the difference in lighting. In recent years, the features of the lips extracted during the articulation process have been considered as a new feature of biometrics, in addition to adopting them for personality recognition, since the features of the lips are characterized by two characteristics: First, the lips possess physiological information represented in the shape and appearance of the lips. The second feature possesses the behavioral information of changing the lips during vocalization.

(a) lip Physiological features (b) lip behavioral features

Biometric analysis includes the determination of the physiological and behavioral characteristics of the lips, and these characteristics are taken into consideration as these features are extracted from isolated images of the lips. Lip movement features are associated with time during speech as biological information for the speaker. These types of features of the lip region can provide discriminatory information for visual speech. Figure (1) shows the difference in the features of the lips between both physiological and behavioral [21].

VI. Suggested Algorithm

In computer vision systems, tracking movements, comprehensive processing is used to extract visual features, especially in lip reading systems, in addition to preliminary processing. Figure (5) represents a general scheme of the lip-reading system.

Fig. (5): of the lip-reading system

The following steps are followed in order to form the symbol for the proposed alphabet:
1- An image of the face is acquired (visual input image frames are acquired from a web camera or video file).
2- Extraction region of interest (ROI) i.e. the mouth region in each frame.
3- Conducting preliminary processing in order to obtain borders for the lips in each frame, as shown in Figure (1).
4- The movement of the mouth is traced from a closed mouth state to an open mouth state and then to a closed mouth state, according to the requirements of the symbol to be configured, as shown in Figure (6).

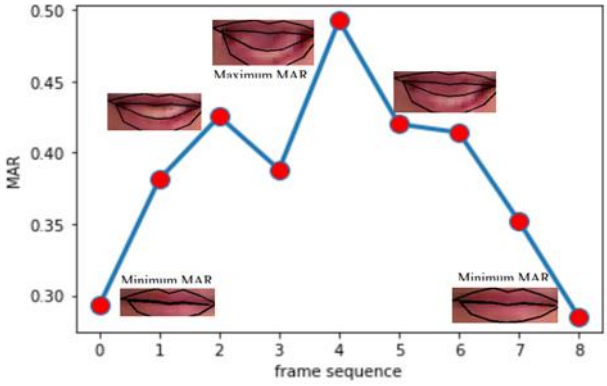


Fig. (6) Traced mouth from a closed mouth state to an open mouth state and then to a closed mouth state

5- Based on the above principle, the symbols shown in Table (1) were adopted in order to build a vocabulary for symbols that could be a model for control orders or dialog signals between people with special needs through a system based on computer vision.

Mouth state				order	Code
Closed	Open	Closed	closed	Need help	0100
Closed	Open	Open	closed	It is cold	0110
Closed	Open	Closed	open	It is hot	0101
Closed	closed	Closed	open	Hungary	0001
.	.	.	.	.	
.	.	.	.	.	
.	.	.	.	.	
Last code (4 state = 16 order)					1111

Table (1) mouth state sequence (as a codes)

VII. Result Discussion

implanting the proposed algorithm to represent mouth contour for any word, which can be adopted to translate mouth movement to speech word, fig(7) show an example to represent lips countor landmarks for pronouncing word ALSAIF.

Fig. (7) lips contour land marks of ALSAIF word

Measuring the distance between the upper lip and the lower lips extract a graphical feature for a pronounced word as illustrated in fig (8) which can be adopted character features in identification systems as well as person detection systems

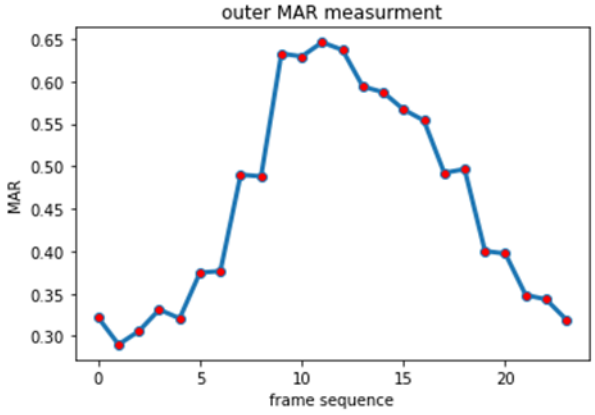


Fig. (8) Distance between upper/lower lips during pronouncing ALSAIF word

VIII. Conclusion

The idea of this research is implemented to generate 16 code (4 bit binary) to produce audial signal which can be translated as a control orders.

A lot of projects can modify the proposed approach be added to any control system used with smart house or serve a persons of special requirements.

REFERENCES

[1] Niu,X. , Zhao, X., Han, H., Das, A., Dantcheva, A., Shan, S. & Chen, X. (2019). Robust Remote Heart Rate Estimation from Face Utilizing Spatial-temporal Attention. 14th IEEE International Conference on Automatic Face & Gesture Recognition. DOI: 10.1109/FG.2019.8756554.

[2] Képešiová, Z., & Kozák, Š. (n.d.). An Effective Face Detection Algorithm. Conference: 2018 Cybernetics & Informatics (K&I). doi: 10.1109/CYBERI.2018.8337566

[3] El-Melegy, M. T., Haridi, H. A. M., Ali, S. A., & Abdelrahman, M. A. (2019). Optimal Score Fusion via a Shallow Neural Network to Improve the Performance of Classical Open Source Face Detectors. n Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019), pages 660-667. ISBN: 978-989-758-354-4 Copyright. doi: 10.5220/0007691206600667

[4] Soundarya, B., Krishnaraj, R., & Mythili, S. (2020). Visual Speech Recognition using Convolutional Neural Network. IOP Conference Series: Materials Science and Engineering PAPER, ICCSSS. doi:10.1088/1757-899X/1084/1/012020

[5] Mishra, R. K., Reddy, G. Y. S., & Pathak, H. (2021). The Understanding of Deep Learning: A Comprehensive Review Ranjan. Mathematical Problems in Engineering. Article ID 5548884, 15 pages

[6] Chai, J., A, H. Z., Li, A., & Ngai, E. W. T. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Machine Learning with Applications.Science Direct Machine. doi: http://dx.doi.org/10.30684/etj.36.2A.4 Mahmuod.H.

[7] Dey, S., Dudhatra, J., Shah, S., & Mane, P. M. P. (2022). Attendance System using Facial Recognition. International Research Journal of Engineering and Technology (IRJET). ISO 9001:2008 Certified Journal.

[8] Thein, T., & San, K. M. (2018). Active Lip Localization Based on Lip Movements Recognition Using YCbCr and CIEL a*b* Color Space. International Journal of Scientific & Engineering Research Volume 9, Issue 6. ISSN 2229-5518. https://doi.org/10.1049/bme2.12073

[9] Sarhan, A. M., Elshennawy, N. M., & Ibrahim, D. M. (2021). HLR-Net: A Hybrid Lip-Reading Model Based on Deep Convolutional Neural Networks. Computers, Materials & Continua 68(2), 1531- 1549. https://doi.org/10.32604/cmc.2021.016509

[10] Parra-Dominguez, G. S., Sanchez-Yanez, R. E., & Garcia-Capulin, C. H. (2021). Facial Paralysis Detection on Images Using Key Point Analysis. Appl. Sci 11, 2435. https://doi.org/10.3390/app11052435

[11] Mahmmed, M. H., Saeed, T. R., & Ali, W. H. (2018). Robust Visual Lips Feature Extraction Method for Improved Visual Speech Recognition System. Engineering and Technology Journal. Vol. 36, Part A, No. 2. doi: http://dx.doi.org/10.30684/etj.36.2A.4 Mahmuod.H.

[12] Zhu, T., Zhang, C., Wu, T., Ouyang, Z., Li, H., Na, X., Liang, J., & Li, W. (2022). Research on a Real-Time Driver Fatigue Detection Algorithm Based on Facial Video Sequences Tianjun. Appl. Sci.
[13] Ivanko, D., Ryumin, D., & Karpov, A. (2019). AUTOMATIC LIPREADING OF HEARING-IMPAIRED PEOPLE. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-2/W12. Int. Worksh. on "Photogrammetric & Computer Vision Techniques for Video Surveillance, Biometrics and Biomedicine", 13–15 May. https://doi.org/10.5194/isprs-archives-XLII-2-W12-97-2019
[14] Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P., & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. npj Computational Materials.
[15] Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., Hasan, M., Essen, B. C. Van, Awwal, A. A. S., & Asari, V. K. (2019). A State-of-the-Art Survey on Deep Learning Theory and Architectures. electronics. doi:10.3390/electronics8030292
[16] Sheng, C., Kuang, G., Bai, L., Hou, C., Guo, Y., Xu, X., Pietikäinen, M., & Liu, L. (2022). Deep Learning for Visual Speech Analysis: A Survey. http://arxiv.org/abs/2205.10839
[17] Shi-Lin Wang, Alan Wee-Chung Liew (2012) Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power, Pattern Recognition, Volume 45, Issue 9, September 2012, Pages 3328-3335
[18] Prof. Khalil I. Alsaif, Nagham Salim Allella, "Lips Reading Spoken Arabic Word Based on The Geometric Shape Features of The Lip", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN: 2395-602X, Print ISSN: 2395-6011, Volume 10 Issue 1, pp. 624-634, January-February 2023. Available at doi: https://doi.org/10.32628/IJSRST2310164
[19] Jadhav, A., Lone, S., Sagarika Matey, Madamwar, T., & Jakhete, S. (2021). Survey on Face Detection Algorithms. International Journal of Innovative Science and Research Technology. Volume 6, Issue 2.
[20] Mishra, R. K., Reddy, G. Y. S., & Pathak, H. (2021). The Understanding of Deep Learning: A Comprehensive Review Ranjan. Mathematical Problems in Engineering. Article ID 5548884, 15 pages.
[21] Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, 2(6), 4.