REPUBLIC OF TURKEY

FIRAT UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

Lip Recognition Using OpenCV

Bnar Azad H.Ameen

Master Thesis

Department: Software Engineering

Supervisor: Prof. Dr. Asaf VAROL

April – 2017

REPUBLIC OF TURKEY

FIRAT UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE

Lip Recognition Using OpenCV

MASTER THESIS

BNAR AZAD H. AMEEN

142137107

Submission Date to the Institute: 2017

Thesis Presentation Date: 14 /4 / 2017

Thesis Supervisor: Prof. Dr. Asaf VAROL (F.Ü.)

Other Juries: Assoc. Prof. Dr. Muhammet FATİH TALU (İ.Ü.)

: Yrd. Doc. Dr. Fatih OZKAYNAK (F.U.)

April – 2017

DECLARATION

I am presenting the thesis entitled "Lip Recognition Using OpenCV" for fulfilling the requirements of Master's Degree in Software Engineering. I declare that the presented content of this thesis is my own work with all simulations and programming.

Bnar Azad H.Ameen

ELAZIĞ – 2017

ACKNOWLEDGMENT

TABLE OF CONTENTS

ÖZET

OpenCV Kullanılarak Dudak Tanıma

Genellikle dudak lokalizasyonu ve tespiti, vurgu çıkarma, okuma tekniği, onay tekniği, kombinasyon stratejisi, ağız ve korpusun düzenlenmesini birleştiren dudak tanımada pek çok yenilikler olmaktadır. Bu tez dudak okuma tanımada görsel temsilin bilgisayara nasıl aktarıldığı ve makina öğrenme tekniklerinin üretilen sesler ile dudak şekillerini eşleştirmede nasıl kullanıldığı göz önüne alındığında kullanılan kavramları tanıtmaktadır. Dudak tanıma için ilgi noktası, dudak kısıtlama stratejileri, dudak sınırlaması vurgulamaları gibi çeşitli alanlardaki yenilikler ile geliştirilmiş bir prosedür tanıtılmaktadır. Ek olarak araştırma, dudak tanıma sistemlerinin zorluklarını ve kısıtlılıklarını tartışmaktadır ve pratik çalışmanın bulgularını kullanarak dudak tanımanın doğruluğunu geliştirmek için önemli öneriler sunmaktadır.

Anahtar Kelimeler: sınıflandırma, yüz tanıma, makine öğrenmesi, dudak algılama, ağız algılama

ABSTRACT

There have been many innovations in lip-recognition which usually incorporates lip localization and detection, highlight extraction, reading technique, acknowledgment technique, combination strategy, the arrangement of the mouth and the corpus. This thesis introduces the concepts used in lip-reading recognition considering both how the visual representation is transferred to the computer, and how machine learning techniques are used to associate lip shapes with the sounds produced. An improved procedure for lip-acknowledgement is introduced, with the innovations discussed in several areas such as point of interest, lip restriction strategies, lip limitation highlights and so forth. In addition, the research discusses the difficulties and limitations of lip recognition systems, and provides valuable recommendations and suggestions for improving the accuracy of lip recognition using the findings of the practical work.

Keywords: classification, face detection, machine learning, lip detection, mouth detection.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AAM   : Active Appearance Model

ACM   : Active Contour Model

ANN   : Artificial Neural Network

ASM   : Active Shape Model

DT   : Deformable Template

GMM   : Gaussian Mixture Model

HMM   : Hidden Markov Model

HSV   : Hue Saturation Value

KNN   : K-Nearest Neighbor

LACM   : Localized Active Contour Model

MIT   : Management of Information Technology

NB   : Naïve Bayes

NN   : Neural Network

PC   : Personal Computer

PCA   : Principle Component Analysis

RGB   : Red Green Blue

ROI   : Region of Interest

SVM   : Support Vector Machines

1. INTRODUCTION

Lip recognition is one of the newer research fields in computer vision. This growing interest comes from a broad range of usages, which is a part of the visual information or can be used to improve overall system robustness and performance [1]. These usages include identifying audio-visual speech, analysis of facial expressions and lip synchronization. Isolating the lips from the face is considered to be non-trivial. Due to the fact that there is a large variety of lip structures amongst people, different colored lips, different lighting conditions, the appearance of the teeth, tongue, and even the facial hair which makes it to be more difficult. In recent years, many techniques have been suggested for lips detection. One of the first methods used to achieve segmentation of lips is edge information [2].

When there are no restrictions on the shape or smoothness of the definition of the lips against the facial skin, the segmentation produces many errors with very rough boundaries. A major group of techniques is known as model-based, where models built for lips are using different parameters to define the lips. These lip recognition techniques include deformable template [3], snakes [4], models based on active contour [5], etc.

These techniques are highly desirable because they do not have many variables. They also do not have problems with images that are rotated, scaled or have different lighting. Getting precise parameters can be troublesome due to the need for large training sets required for a large variety of lip structures. In addition, the adjustment parameters are generally very difficult to achieve, and many need to be selected and reset manually. Color has been found to be beneficial for lip reading recognition and has been widely used [3, 4, 6, 7].

A proposal for lip-reading was presented in 1954 by Sumby and Pollack. Here, the visually observed lip movement was used as an aid for understanding human speech in a noisy environment. Lip recognition is used for explaining the movement of the lip, and which allows the recognition of people facial expressions and emotional states as well [8].

At the University of Illinois, a Ph.D. student called Petajan created the first lip-reading system in 1984. His system was designed to use lip movement to enhance automatic speech

recognition. The recognition was limited to isolated utterances and speaker dependent recognition. The accuracy of the system is said to be much better than relying on speech alone [9]. Apart from this, in 1989 and 1991 artificial neural network was introduced by Yuhas and Mase from MIT respectively [8]. Artificial Neural Networks (ANNs) were first used by Yuhas in 1989 for lip reading recognition. Goldschen used Hidden Markov Models (HMM) in 1993, but in 1994, Bregler combined both ANNs and HMM for lip recognition [10].

In 2010, researchers at the Institute of Technology Karlsruhe in Germany developed a mechanism for transferring lip movement into a sound, which implied that smart phones would be able to read lips. By using the mouth muscle measurements, the telephone will be able to know what is being said. Using this information, it can then transmit this data down the line.

In 2013, Microsoft launched Kinect2 which could read lips. Kinect2 accurate motion capturing function, captures the depth information of the lip by the depth sensor to receive recognition of the lip to be identified. This summarizes the history of the development of lip reading technology. The accumulated work done in the eighties and nineties helped in the breakthrough required for lip-reading technology in the twenty-first century. Lip reading technology is currently in a hot area of research, not only in theoretical terms, but also practical. There are many new software and hardware related to lip reading, that is becoming more mainstream [8].

For humans, lip reading helps to drastically improve the accuracy of speech recognition in a noisy environment when compared to just relying on the audio. Lip reading is also useful for those who are deaf as their main means of communication to understand what is being said. Lip reading technology would be very useful to automate such situations for those who don't know how to read the lips. The ability to extract the lips from facial features is critical, it would also need to recognize different varieties of lips, so this is an essential part in enhancing the recognition rate of speech recognition frameworks, or for improving the communication abilities of those who are mute [11].

The ability to use lip reading to enhance human speech recognition is extremely valuable [12]. As mentioned earlier, lip reading technology would be useful for use in noisy environments, or where audio cannot be heard, either by those who are deaf, or with special needs [12].

The complete lip reading system can include lip position, lip tracking, lip movement extraction and lip reading. From the face image, the lip position is the primary and critical part, and its accuracy can affect the performance of the following lip tracking, lip movement extraction and lip reading. In previous systems, it was predominant to use the camera to capture only the lips region, or manually marking the lips' area. However, this is not practical and cann't be applied automaticly. In addition, it increases the limitations and difficulties for developing lip reading applications. In more modern systems, the lips position is usually determined based on face detection.

There have been some ways to locate the lips region after facial detection. One is to roughly position the mouth region (region of interest) based on the distribution of the mouth in the face area. The technique, for the most part, takes a large portion of the width of the face. This portion should include the width of the mouth area and the lower third of the face that contains the mouth. The advantage of this method is that it is a simple and fast mouth positioning method. However, it loses the effectiveness of those images with different head positions and lips shape.

Another way is to calculate the gray projection [13]. In this method, the image is projected onto the horizontal and vertical axes, and the mouth area is defined by the valleys of the horizontal and vertical curves. In this method, the mouth region can be easily defined; however, the accuracy can be also be easily affected by poor lighting conditions, low discrimination of the lip and skin color, and the beard. Lips tracking is based on the precise lips position. Historically, there are two primary approaches to track from the image sequences of lips. The main technique depends on the color strategy. In this strategy, distinctive color spaces are proposed, for example, RGB [12], HSV (Hue Saturation Value). Red exclusion is a very good way to extract lips in which the mouth is tracked by using the 0 and B color components, but it is only useful for white people. Another color based strategy [14] was proposed by analyzing the color distribution of the lips and the

skin to extract the lips area. But the disadvantage is that it is only useful for a particular color, yellow skin or white skin, and does not consider beard and teeth. The second sort is known as a model-based approach.

The lip model is portrayed by an arrangement of parameters, for example, the tallness and width of the lips. These parameters are ascertained from the cost function minimization handle on the captured picture that fits the model of the lip. The dynamic form show, the deformable geometric model, and the dynamic shape model are examples of such a strategy that is broadly utilized as a part of lip following and highlight extraction. The benefit of this technique is that the lip shape can be effortlessly portrayed by a low request measurement, and it is invariant under pivot, interpretation or scaling. In any case, the strategy requires a precise model instatement to guarantee that the model refresh prepare joins [14].

1.1. Aim of Study

The purpose of this research is to investigate lip recognition using openCV.

1.2. Overview of Thesis

Chapter 1 discuss the general review of the topic, the aim and thesis organization

Chapter 2 presents the related literatures of lip recognition

Chapter 3 gives the theoretical framework of the thesis

Chapter 4 and 5 present the experimental results along with their discussion

Chapter 6 presents the conclusions and final recommendations of the thesis

2. RELATED WORK

Recently, there has been an increasing demand for tracking and locating human lip systems [15]. The human lip has more data than other facial elements, so the lip data can be utilized for picture coding [15]. With a specific end goal to enhance the execution of speech recognition, lip data and sound signs are utilized together [16]. The data is likewise connected to the realistic movement framework, which obliges it to recognize the lips of the speaker [17]. The inclination based strategy [18] for edge recognition of the lip gets frequently stuck because of the difference between the lip and the encompassing skin region. For the utilization of shading data for the lip shapes to develop parametric deformable models, these strategies require improvement methods to refine the evaluations of the human lip form the view [19].

Many papers describe the application of active contour model (snake) to lip boundary detection [20]. Snake methods can analyze fine outline details, but it is difficult to add constraints of the shape while deploying them. Over the past 15 years, a variety of lip positioning methods have been described in the literature. The popular method divides the lips from the rest of the face based on the color and intensity thresholds. The lips are usually positioned by fusing the shape model around the split mouth, where many techniques are studied. Another popular method is to use snake joint mouth feature detection [21].

In addition, the shape template has been used to position the lip profile. Another method is to classify the regions in the image according to the horizontal and vertical intensity distributions, taking into account the different projections of the shadows in the oral area [22]. There are several publications devoted to real-time lip tracking. They often use the same method as above, perhaps as a simplified or accelerated variant. Color-based segmentation methods often lack robustness to changes in lighting and lips shape, and especially for facial hair. Petajan and Graf [23] presented an interesting solution where the opening of the nostril is used to determine the approximate mouth position and to estimate the facial hair. Yang et al. [24] proposed a simpler method for searching only six feature points with characteristic corner features on the lips.

In a recent paper, Jang et al. [25] proposed the use of Gaussian Mixture Model (GMM) as an alternative to GLDM. Although the overall detection quality was only slightly improved, the placement of the inner lip profile was significantly improved by this average. Lips feature extraction or lip tracking becomes complicated due to the same problems encountered by facial detection, such as changes between humans, changes in lighting, and the like. However, lip feature extraction is often more sensitive to adverse conditions. For example, a mustache can easily be confused as an upper lip [25].

The lack of sharp contrast between the teeth, tongue, lips and face can further complicate the extraction of the lip features. Recent techniques use knowledge about the color or shape of the lips to identify and track the lips. In fact, color differentiation is an effective technique for locating lips [19]. The study shows that the hue component provides a high degree of discrimination in the hue saturation color space. Thus, the lips can be found by isolating the connection area with the same lip color.

Obviously, color discrimination technology does not apply to grayscale images. Techniques for using information about lip shapes include the active contour model, the shape model, and the active appearance model [26]. Unfortunately, these technologies also require a lot of storage, which from the hardware point of view is not attractive. In the fifth chapter, a lip feature extraction technique that utilizes the contrast at the lip contour is proposed. This technique works well on grayscale images and can be easily implemented on hardware.

As of late in the field of PC vision, lip recognition has brought on a great deal of consideration. This developing interest emerges from an extensive variety of uses, where visual data is an essential piece of the general framework or can enhance the general framework execution and vigor. These applications incorporate varying media speech recognition, lips synchronization, manufactured discussion facial and outward appearance examination.

Nonetheless, exact and hearty lip recognition is a non-insignificant undertaking because of the high changeability of the lip, the distinctive lip tones, the lighting conditions, the presence of the teeth and tongue, the nearness of facial hair, et cetera. In the previous couple

of years, various procedures have been proposed to accomplish lip recognition. Edge data is one of the principal techniques used to accomplish lip section [27]. At the point when there is no shape or smooth imperative, the division may for the most part be harsh, and the edge of the lip limit can be low and overpowered by solid false edges. Known as model in light of vast scale innovation, the model of building lips and its arrangement is portrayed by an arrangement of model parameters. These procedures incorporate snake [28], dynamic form show [29], deformable format [30] and a few other parametric models [31].

The upside of these procedures is the way that imperative components are spoken to in low-dimensional parametric spaces. Likewise, they are invariant for turn, scaling and lighting. In any case, the development of these models is normally exceptionally difficult and requires an extensive preparing set to cover the high fluctuation of the lips. What's more, parameter alterations are frequently exceptionally hard to actualize, and a large portion of them should be physically chosen and instated.

The shading gives extra data to demonstrate which is helpful for the assignment of lip recognition and has been generally utilized. Dissimilar to most lips recognition frameworks that force certain limitations on the client [32]; for example, wearing a headset [31], spreading the subject's lips or operating in a very controlled condition. In this way, disposing of the down to earth application is the best approach to stay away from these requirements. The main prerequisite this study's framework is to approach the issue of lip fracture. That is, the lips locale must be recognizable from whatever remains of the skin range, accepting it is the more prominent redness along the face. Red artifacts in little zones don't influence division comes about and are naturally dispensed with [31].

Yao et al. [33] proposed to first identify the position of the eye and then split the lips picture as indicated by the relative position of the eye and the mouth. The strategy can position the lip around the position, yet the upper and lower parts of the lip are ineffectively situated.

Rao et al. [34] extricated the element by utilizing the lips edge, particularly the clear flat edge include [35], to figure the dark incentive from the lines and segments to decide the lip territory. Be that as it may, this strategy is influenced by light, shadows and facial hair.

7

Pera et al. [36] acquired the lips district by disposing of the red shade of the skin. Lievin et al. [37] utilized shading contrasts between the skin and the lips, first transforming the Red Green Blue (RGB) space into the lips locale with a significantly enhanced space, contrasted with the split lip district. Jun [38] proposed a versatile lip position calculation in view of chromaticity differentiation, which first uses the HSV shading model to separate chromaticity and illuminance. Then, as indicated by the diverse Chroma of the skin and lip, consolidated with the R/G channel strategy to find the lip area. This strategy in view of shading space and shading model is constrained by lighting changes and amplifier hues [37].

Moving item location can rapidly remove moving articles in the picture arrangement, while in the lip picture succession, the lips are the fundamental moving target [39]. Pao [40] distinguished lip locale identification in light of movement data. Da Silveira et al. [41] proposed a technique in view of casing contrast strategy and moving item recognition association frame.

Also, Timothy Jordan proposed a lip-construct division calculation situated in light of shading and profundity pictures [42]. This system utilized the Otsu versatile edge division technique, from the grayscale luminance difference shading space segment. This technique uses the *k*-means grouping calculation of lip pixels and different pixels, successfully enhance the shade of the tongue shading model of the order impact. In spite of the fact that lip identification and situating methods have been developed, there are still issues worth contemplating. For instance, there are a few issues in this present reality that should be tended to, for example, light changes, speaker developments, distinctive angles, diverse head pose identification issues, redundant data reduction, enhancing handling velocity and so on [42].

3. DEVELOPMENT OF LIP RECOGNITION

A collection of lip constraint strategies has portrayed in the composition of lip recognition technology through the late 15 years. Conspicuous philosophies based on shading and power thresholding to partition the lips from whatever left from the face were proposed in [43-44]. By and large, the lips are then arranged by fitting a shape model around the isolated mouth, where various procedures are investigated. Another surely understood procedure is the usage of snakes in conjunction with mouth corner part identification [45].

Furthermore, shape designs have been used as a piece of solicitation to restrain lip frames. Another procedure is to organize the zones in a photo according to the level and vertical power profiles, with unprecedented contemplations of the assorted tossing of shadows in the mouth zone [46]. There are a couple of studies that especially focused on consistent lip taking after. Every now and then, they use the same strategies as indicated above, perhaps as unraveled or quicken varieties.

3.1. Features Used in Lip Segmentation

The features utilized as a part of lip segmentation can be broadly separated into two primary classifications: appearance based components, and shape-based elements [47].

3.1.1. Appearance-based Features

These can be basically gotten after face detection. A few strategies are utilized to discover facial components, for example, eyes and nostrils, in light of their relative position on the face. An illustration of the eyes confinement calculations can be found in [48]. Another sample of the appearance - based element incorporates pixel shading data or force esteem. Shading data gives advantages in either removing highlights, for example, lips, or smothering the undesirable ones. The appearance-based elements are ordinarily extricated from the Region of Interest (ROI) utilizing picture changes. For example, changes to various shading space segments, where pixel estimations of the face/mouth ROI are utilized [47].

3.1.2. Shape-based Features

These are by and large separated into geometric, parametric and measurable models of the lip shape, and are extricated utilizing systems, such as snakes [49], template models [50, 51], appearance model, and active shape and model [52]. This component expects that a large portion of the data is an accurate representation of the speaker's lips. Geometric elements, for example, tallness, width, and the border of mouth, can be promptly separated from the ROI. Then again, model-based components are acquired in conjunction with parametric or factual element extraction calculation [52].

3.2. Technologies Used in Lip Localization

The following are fundamental techniques used in lip localization.

3.2.1. Region-based Technique

The locale based technique mainly uses the local measurement qualities to acknowledge lip tracking. Ordinary samples incorporate Deformable Template (DT), region based ASM (Active Shape Model), ACM (Active Contour Model), and AAM (Active Appearance Model) [53]. The DT figuring uses a local cost, a lip picture divided by the lip and non-lip ranges by a method for a parametric format, and it addresses the mouth area fittingly. Yuille introduced the initiative effort that demonstrates a lip format controlled by a template of parameters [54], these parameters are changed by a method for a noteworthy minimizing taking care of so that the lip layout can encourage beyond what many would consider possible prices. A lip region is shown in Figure 3.1.
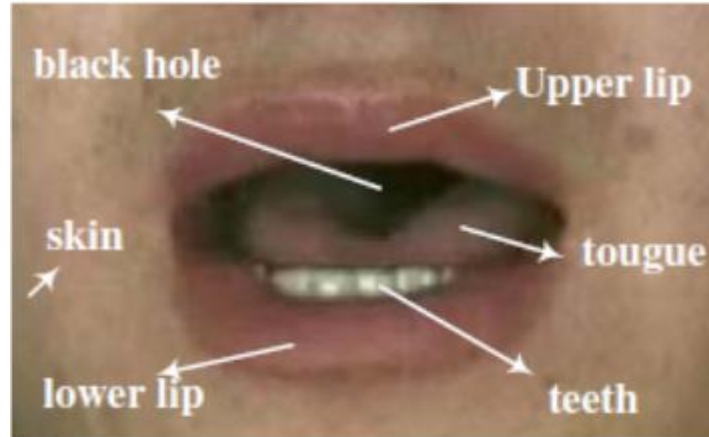
Figure 3.1 A lip region consolidated the appearance of teeth, tongue and dark gaps in the oral depression [53]

Region based technique can be divided into three main classes:

- The deterministic technique which depends on shading appropriation modeling and threading operations.

- The classification technique (managed or non-administered) which considers lip division as a pixel class issue, for the most part, between skin pixel and lip pixel.

- The statistical technique which depends on mouth shapes and appearance.

As every one of these techniques is an area-based, the exact lip division around the lip shapes is not generally accomplished.

3.2.1.1. Deterministic Technique

In this strategy, no earlier information and no former models about the measurements of the lip and skin shading dissemination are used. The lip division is performed with a thresholding venture on the luminance or specific chromatic segment. Programmed calculation of the powerful edges in different lighting conditions is the main test and constraint of this technique [55]. An approach for determining the edge appears in [56], With an example picture information set, the histogram records the different chrominance parts for each set of colour values against the maximum intensity for each individual colour component, which is known as the maximum normal intensity normalisation.

11

3.2.1.2. Classification Technique

With face detection as a preprocessing step, the lip division can likewise be seen as a grouping issue with two classes: the lip class and the skin class. Using distinctive characteristics characterizing every class, the grouping technique used in face detection between the skin and non-skin class can likewise be connected to lip division [47]. The most commonly used strategy includes factual techniques (estimation hypothesis, Bayesian choice hypothesis, and Markov random field), neural systems, bolster vector machine and fluffy C-mean. They can be ordered into a regulated and unsupervised methodology. Administered strategies use earlier information about the distinctive classes. It involves the development of the training information set that covers an extensive variety of classes and ecological conditions.

3.2.1.3. Statistical Technique

Another administered technique is the statistical shape models in which the training set is incorporated to depict the lip shape or appearance variety instead of the lip shading circulations. A shape model is found out from training a set of clarified pictures. The Principle Component Analysis (PCA) produces a little arrangement of parameters to drive the model. To achieve the correct results, appropriate parameter values are chosen. This strategy is known as ASM [51]. AAM was introduced by Cootes and Taylor to include dim level appearance information in the training. Reference [51] is discussing a sample of ASM and AAM.

3.2.2. Contour-based Technique

It is a standout amongst the essential strategies for applications of human-machine, for instance, lip-acknowledgment and outward appearance examination [56]. Incidentally, it may be unnecessary to find a strong and exact extraction methodology because it can produce significant image errors. This may happen because the speakers can have different skin tones which are light and dark, and the tongue and teeth may appear which changes the color of the mouth between the lips. There can be little difference in contrast between the facial skin and the lips, as well as the lip shapes, which may be differ among people. By the earlier decade, different procedures needed to be suggested to complete the lip

structure [56]. These are divided into two classes: model-based methodology and edge-based strategy.

3.2.2.1. Edge-based strategy

It mainly uses the low-level spatial prompts, for example, edges and shading, to fulfill lip confinement and eradication. The execution of such a system frequently rots when there is a poor many-sided quality amidst lip and encompassing skin regions [57].

3.2.2.2. Model-based methodology

This model forms a lip model and a little course of action of lip-parameters that usually produces better results than previous [56].

The conventional ACM grants an underlying shape to turn by minimizing a specific overall imperative ability to convey the pined for the division. Reference [56] is presenting the accomplishment of this methodology in its application area. This framework may delineate the exact parameters, uneven enlightenment, and teeth sway. Moreover, while things have heterogeneous estimations, it is found that the limited locale based ASM can overall perform a satisfactory division result while the routine ACM misses the mark [58]. From the LACM (Localized Active Contour Model), the advancing bend parts in those close-by neighborhoods under the area inside the range and neighborhood outside the district independently. As need be, the limited essentialness for advancing and removing can be figured. In any case, less than ideal parameters [59], for instance, broad reach or distance away advancing twist in LACM can provoke the incorrect extricating results as shown in Figuer 3.2 and Figuer 3.3.
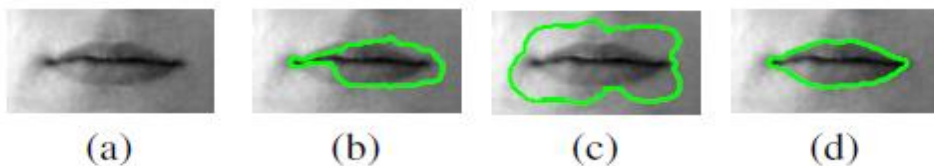

Figure 3.2 (a) Mouth. (b) Extricated by ACM. (c) Extracted by LACM. (d) Separated by the LACM with the best possible parameters [59]
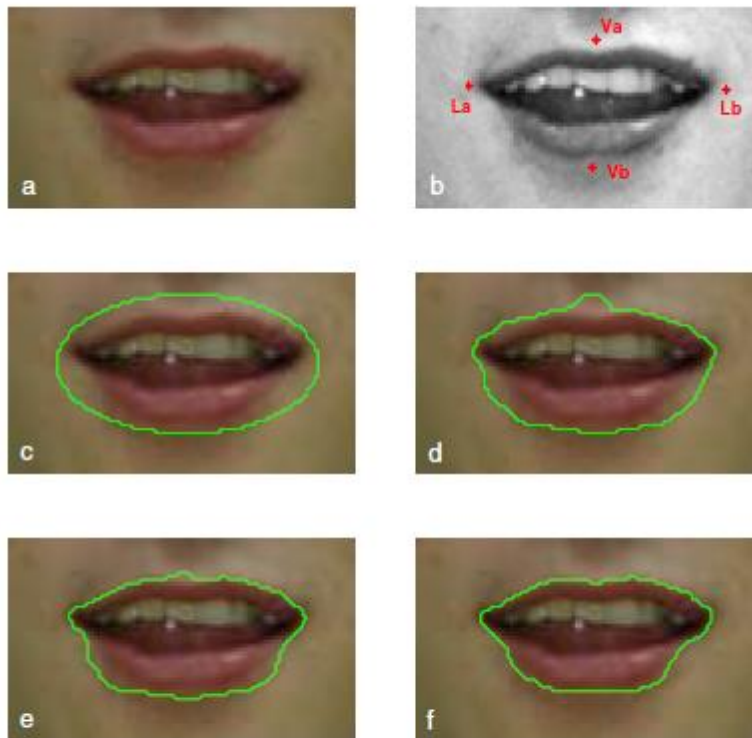
13

Figure 3.3 The technique of lip form: (a) size 52 X 90 of the lips, (b) corner dots and lips, (c) ellipse of least-bounding (d)- (f)- the separated results after 20, 40 and 60 cycles, individually [59]

3.3. Applications of Lip Recognition Technology

- Aid Signals Acknowledgment

Motions and lip-acknowledgment are required. When they watch developments, they need a portable amplifier calms watch and moving shape and likewise facial appearance and lip moving. The conducted experiment on mouth shape movement acknowledgment can upgrade clarity of development acknowledgment. Its planned framework is the same as that of the discourse acknowledgment.

- Aid Discourse Acknowledgment

It is a principal usage of the lip-acknowledgment to join discourse acknowledgment. In the world of different situations and speakers, it can improve the rate of discourse acknowledgment to join the lip-acknowledgment and discourse acknowledgment.

- Human-PC Interaction

Speech-recognition development ability to be related in the range of somatotype communication will get fresh method for human-computer interaction, for example, lip information methodology.

- Mouth-coding and Mouth Combination

As one might expect, the most important emotional indicator is the sentiment words, also called opinion words. These are words that are normally used to express positive or negative sentiments. For instance, great, awesome, and stunning are certain positive sentiment words, and awful, poor, and unpleasant are negative sentiment words. Aside from individual words, there are additional expressions and sayings, e.g., cost somebody dearly. Sentiment words and expressions are instrumental in sentiment analysis for evident reasons. A rundown of such words and expressions is known as a sentiment dictionary (or opinion vocabulary). Throughout the years, analysts have composed various calculations to assemble such vocabularies.

- Authentication and Security

Biometric studies with the aid of the lip were investigated in [27]. In [28], a demonstration of a part extraction technique that is based on the quadratic presentation framework for interior lip shapes. Here, and the pictures were changed over into shading

15

area to color area which is used to diminish the influences of human's teeth. In [29], sound and apparent modalities were essentially used for discourse acknowledgment. It demonstrated another cross breed approach too that oversees lip limitation and taking after. In [30], to beat the officially communicated lighting issues, different strategies with the iterative system were proposed. Basically, these strategies depend on the conventional of the structure. They upgraded the dynamic structure model to see the structure.

- Deaf-quiet Guide Instruction their Discourse Capacity

It is astoundingly chief for us to add to a mouth shape and the talk structure for the educators of needing a listening device calm school to demonstrate practically hard of hearing calm understudies.

4. ARCHITECTURAL DESIGN OF LIP RECOGNITION

The lip reading using OpenCV described in this research, as shown in Figure 4.1, can be divided into the following four stages:

- The face is detected from the video sequences of the speaker using the Harr cascade.
- The lower part and mouth are detected also by using Harr cascade.
- The lips are detected using image-based methods (RGB color space, HSV color system, YCbCr color system).
- The classification is done using machine learning algorithms such as k-Nearest Neighbor (k-NN), Naïve Bayes, Support Vector Machine (SVM) and Levenberg-Marquardt. The output of the system is a speech detection. The architectural design is shown in Figure 4.1.
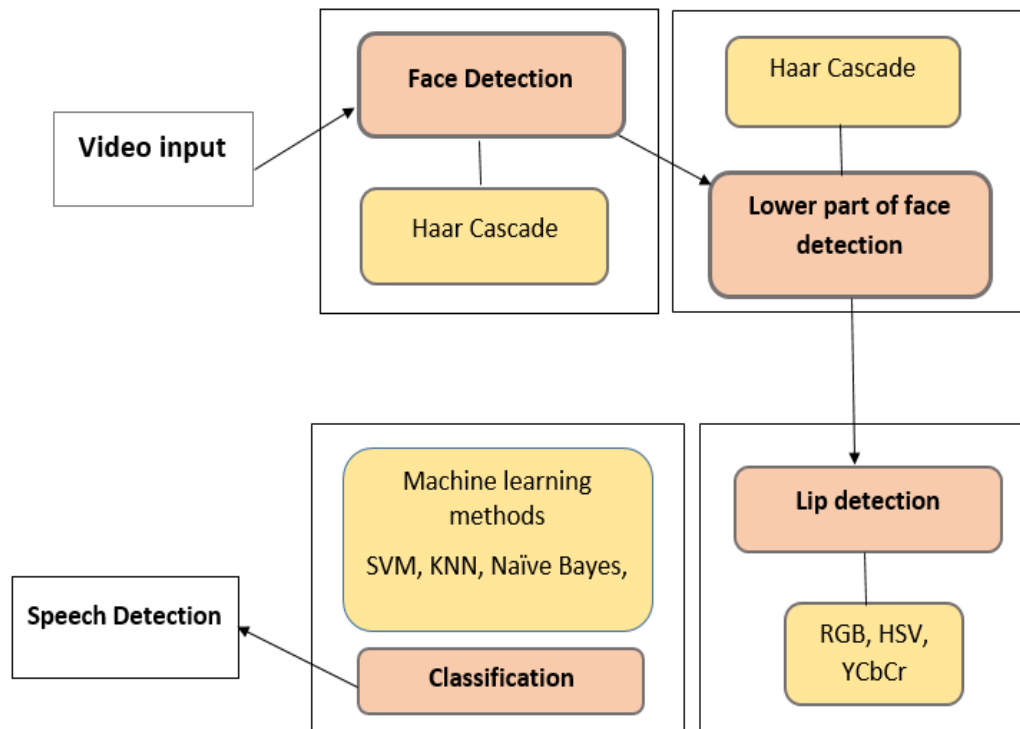

Figure 4.1 System design steps

4.1. Detection of Face and Lower Part

Haar classifier is used for the adaboost cascade in OpenCV in this thesis. The face shape and the nose position on the face are important input values for recognizing the identification of a person. As the nose is relatively stable when the mouth shape changes, one can use the place of nose on the face to estimate the position of the mouth. Since the mount is moveable, a contour of lips should be defined at the first stage. The position of the nose and the distance between lips' contour and tip of the nose are key factors to recognize the person's identification [60].

Haar feature-based cascade classifier is an object recognition method that can be used for different kinds of biometric features. Haar classifier is used for lip recognition in this work. At the beginning, many images of lips during a speech are needed. Some Haar features of images have to be defined. Each feature is a single value obtained by subtracting the sum of pixels under a white rectangle from the sum of pixels under a black rectangle. All possible sizes and locations of each kernel is used to calculate plenty of features. For each feature calculation, the sum of pixels under white and black rectangles need to be found. It is suggested that the integral images should be considered to solve black and white rectangles [60].

Speaker images from the video are extracted. Face and mouth region are detected by using Haar cascade. One of the most important process in a lip recognition system is the detection phase; if this step is accurately set, then the recognition step is not very complicated [61]. After applying the face detection process, the mouth region is cropped by mouth detection process. There are a lot of techniques for extracting the mouth region, some of which were discussed in previous sections. The machine learning (Viola Jones object recognizer) [62] is employed for distinguishing between the lower and upper parts of the face and for isolating the mouth area in the lower part of the face. The procedure requires two steps, the first is detecting the facial region of the person, and the second is using the fact that the area of mouth is in the lower part of the face.

4.2. Lip Detection

Lip detection is an important part of lip recognition and facial feature extraction. The different methods for lip localization/detection can be classified as image-based and model-based, image-based techniques are used for the system of this study, which were suggested by [63].

The following three steps are required to detect the lip from original face image:

- Lip regions are cropped from the original image.
- Lip features are extracted.
- Lip is recognized by using machine learning algorithms.

4.2.1. Image-based Lip Detection Methods

These methods are based on the difference between face colors around the lip and lip colors. The color of lip and skin are differentiated using a conversion based on hue of HSV color, RGB, and YCbCr color system. The important information that is focused on includes the hue of the HSV, red and green in the RGB, and red and blue in the YCbCr.

4.2.1.1. RGB

In RGB, lip and skin pixels are different in color components. They both have red color components, but the skin has more green than the blue color components. In addition, the skin appears more yellow than the lips [31]. The approach presented in [31] works as follows: after detecting the face area firstly, the lip ROI is defined as the region in the lowest part of the face, then, the R-component in RGB color system is excluded, next, B-component and G-component are used as the Fisher change vector to improve the lip image. At last, an adaptive threshold is used to split up the lip color and the skin color according to the normal distribution of the grey value histogram.

4.2.1.2. HSV

In HSV approach, 'H' means Hue (the dominant color perceived by someone). 'S' represents the saturation, and 'V' means intensity or brightness. Some specific equations

have to convert from HSV to RGB and vice-versa. Hue can be used for differentiating between face and lip as hue value of the lip is less than the face's one [64].

4.2.1.3. YCbCr

YCbCr is used in digital videos, where the luminous component is Y, Cb is the blue difference and Cr is red-difference Chroma components. Lips have low Cb, high Cr values and more red than face. YCbCr color system is based on this fact. Also by using some specific equations, the Cb is minimized and Cr is maximized. After that, any unnecessary information is removed using edge detection, then the output of the image is threshold, and the mouth area is found [64].

4.3. Machine Learning

In machine learning, a machine is basically a storage and a predictor too, which has training and output data. The training data is the teacher of the machine, which is represented in this development as an array of doubles standing for the pixels of the image. The learner is an object of the teacher class [65]. The methods that have been used in this implementation are:

1. SVM
2. K-NN
3. Naïve Bayes (NB)
4. Neural Network (NN)

Each of the above methods have different ways of dealing with the training data and inputs. They could be all run in parallel or individually. They learn the training data after the user has selected which machine to run. Selecting the algorithms are toggled using a 4-bit selector as shown in Table 4.1.

Table 4.1. Selectors of algorithms

Selector 4 Bit				KNN	SVM	NB	Neural Network
0	0	0	0	OFF	OFF	OFF	OFF
0	0	0	1	ON	OFF	OFF	OFF
0	0	1	0	OFF	ON	OFF	OFF
0	0	1	1	ON	ON	OFF	OFF
0	1	0	0	OFF	OFF	ON	OFF
0	1	0	1	ON	OFF	ON	OFF
0	1	1	0	OFF	ON	ON	OFF
0	1	1	1	ON	ON	ON	OFF
1	0	0	0	OFF	OFF	OFF	ON
1	0	0	1	ON	OFF	OFF	ON
1	0	1	0	OFF	ON	OFF	ON
1	0	1	1	ON	ON	OFF	ON
1	1	0	0	OFF	OFF	ON	ON
1	1	0	1	ON	OFF	ON	ON
1	1	1	0	OFF	ON	ON	ON
1	1	1	1	ON	ON	ON	ON

The above selector is filtered using AND operation to see which machines to operate. The order of machines which run are followed, as it has been listed in columns of the table. Algorithm chooser is shown in Figure 4.2.
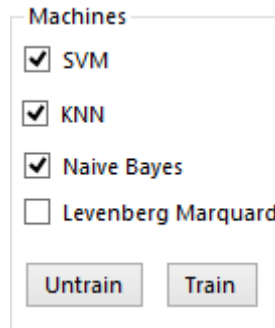
Figure 4.2 Algorithm chooser

4.3.1. Static Machines

The static science is, by some means, helpful to the machines and it creates models which could be used in prediction. This approach follows the distance functions closely. Those algorithms are often employed in computing systems. What they have in common, as the main factor, is the differences between each sample and their distance from the center of data.

4.3.1.1. K-Nearest Neighbor

This mechanism uses a traditional static approach on the training data. Its linear view on data is occasionally predicted in a high accuracy. K-NN learning system creates an N dimension environment where the N is the number of attributes. So, each sample is given a specific location to sit in this N dimension world. The new incomes, are also given a location, and upon their distances to the closest neighbor, they are classified as what their neighbors are. In this project, (16x16) attributes are created to save samples and forms of the trained data [66]. A general K-NN model is shown in Figure 4.3.
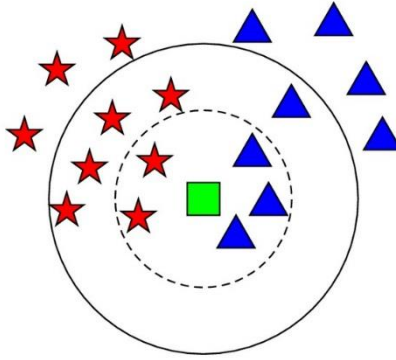
Figure 4.3 K-NN model [66]

4.3.1.2. Support Vector Machine

This algorithm uses the same N dimension world to represent the data. However, the difference between SVMs and K-NN is that the SVM uses some of the samples as the border. The border samples, which are at the end edge of the set, are called support vectors. Those would decide for the boundary of each class. For instance, what have been implemented in this study system is a multiclass SVM using the Leningrad kernel function. The function creates a border straight or covered line to separate classes from each other [67]. A general SVM model is shown in Figure 4.4.
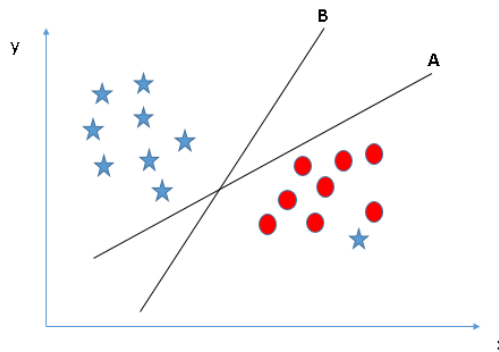

Figure 4.4 SVM model [67]

4.3.1.3. Naïve Bayes

Naive Bayes is a simple method for producing classifiers. Those classifiers are models that map class labels to the problem samples. The class labels here are represented as arrays of feature values, pixels in our case, where the class labels are taken from the data set. This

is a probabilistic algorithm which applies Bayes's theory on naïve assumption of attributes. The Naïve Bayes programmed in our system uses a byte value to represent the data. Those bytes are saved in a vector of integers; after the training, the model will be created [65].

4.3.1.4. Neural Network

The neural network mimics the human brain's method to recognize and predict values. It creates a network of neurons "virtually" which are activated when they receive the right input. Those type of networks are trained in different ways in static machines. Neurons should be inserted to the network in a large number and the training is done in a number of iterations. The trainer class is run until it has the least error rate. This process has been programmed in developed system as an object of the teacher class. The input is a matrix of attributes (1 and 0) to the output representing the image frames received from the camera [68]. A general K-NN model is shown in Figure 4.5.
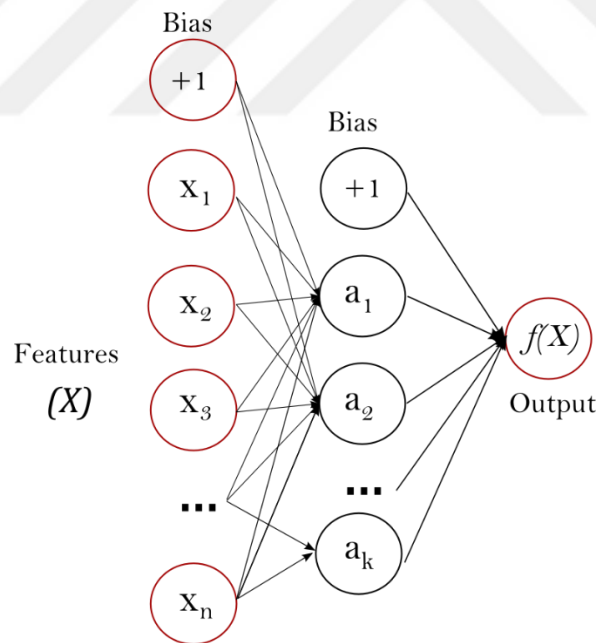
Figure 4.5 Neural network model [68]

The trainer uses a loss function. In the proposed system [68], Levenberg-Marquardt algorithm, also known as the damped least-squares method, is used. This algorithm takes the sum of squared errors after generating the matrix, the matrix itself is calculated as in the famous Jacobian matrix of loss function. This method runs as in equation 4.1.

$$J_{i,j} f(w) = de_i/dw_j \ (i = 1,\dots,m \ \& \ j = 1,\dots,n) \tag{4.1}$$

Equation 4.2 defines the parameters improvement process with the Levenberg-Marquardt algorithm for weights calculation.

$$w_{i+1} = w_i - (J_i^T \cdot J_i + \lambda_i I) - 1 \cdot (2 J_i^T \cdot e_i), \ \ i = 0,1,\dots \tag{4.2}$$

5. IMPLEMENTATION OF THE SYSTEM

5.1. Introduction

In this chapter the practical parts of the project are described and explained. The reader will find out what materials, including software, hardware or virtual system that have been employed. Furthermore, the training data is defined, stating how and why the data has been collected. The main aspect of the actual work is addressed in the context of this chapter of research. It also shows a brief view on the challenges and difficulties faced while creating the implementation. At the end, proper solutions are proposed to the faced problems.

5.1.1. Terminology

The practical part of the research is what most of developers do focus on. The primary display of production depends on the practical tasks which define the result. In this research, programming and developing were the major components of the production. However, it could not have been done without studying each part of the research. Although, in this developed application, high level languages cover most of coding areas, some functions are designed in low level coding grounds. What attracts the users most, is usually a decent graphical user interface. The interface is designed for both programmers as well as the end user.

5.1.2. Tools and Materials

The main language of the developed program is C# (C Sharp). The interface has been made for Windows operating system and it can run on all Windows frameworks. The selected framework for C# is Version 3.5 to support most of used libraries. OpenCV as well as EMGU have been employed in this project, which are open source binaries. The OpenCV comes with a number of other connected libraries which help the application to detect capturing devices. For instance, DirectShow and ZGraph extensions which aid the project by their utility pre-programmed libraries.

For machine learning and NN algorithms, they have benefited from Accord.NET and AForge.NET open source libraries.These are the official container and wrappers for

machine learning algorithms. In these libraries, there are classes for static and kernel learners which could be activated in the system. For capturing training or test data, video clips in this case, any sort of 2D camera would be applicable. There is also flexibility that video clips from different sources can be imported and worked on. From these tools and materials, a very friendly window is designed for the lip recognition application, as shown as in Figure 5.1.
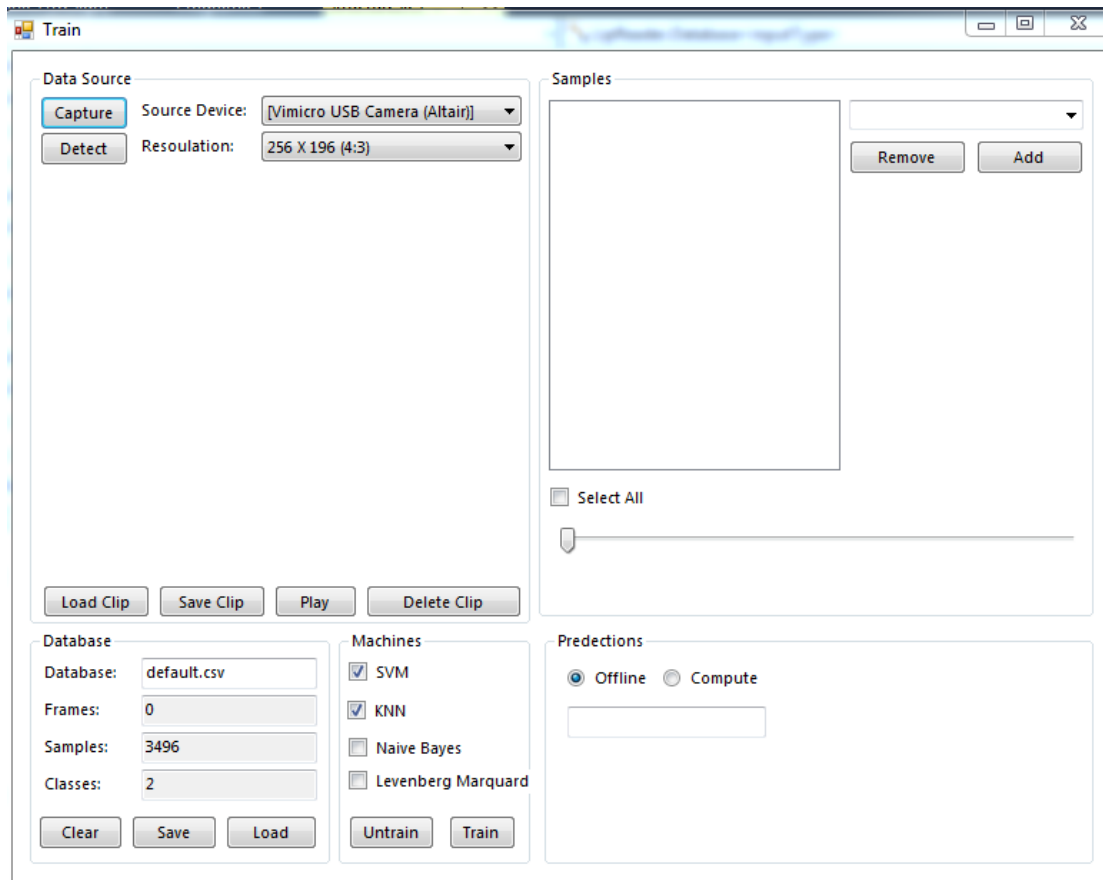
Figure 5.1 Interface of the lip recognition application

5.1.3. Data Definition

It is a fact that an artificial intelligence algorithm needs training data in supervised learning. For this propose, a number of video clip samples has been collected from various styles of lip movements. The range of selected people for data ranges from 18-35 years old for both male and female. Furthermore, they were captured to have different style of speaking and obviously different shapes of lips. Those clips were recorded with no voice over but labeled instead. The clips were recorded and injected to a preprocessing phase to

27

be used as real trainer. This system is explained in the next sections of this chapter. An example of the captured samples is shown in Figure 5.2.

Figure 5.2 Captured samples

5.2. Preprocessing

Preprocessing is an essential task for machine learning and neural network. This increases the reliability and correctness of the data. In this section, a number of steps are carried out where used data is passed through. The preprocessing tool has been programmed and it is shown in Figure 5.3.
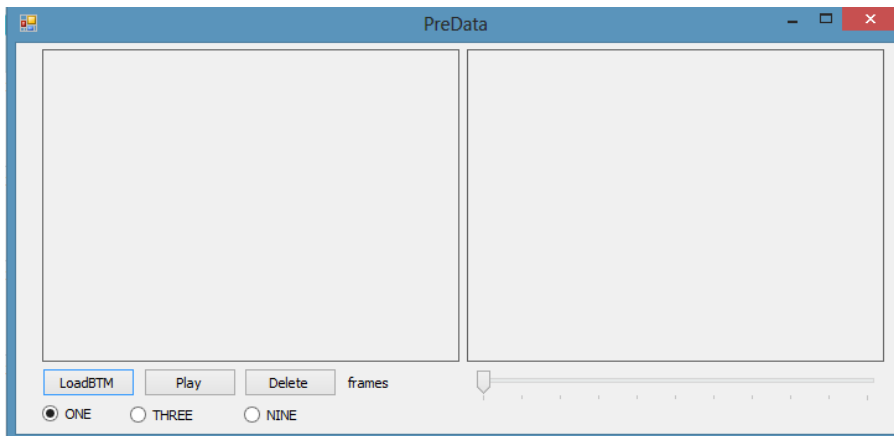

Figure 5.3 Preprocessing Utility Tool

5.2.1. Data labeling

Supervised learning requires labels for each sample and this should be done before any learning takes place. The labels in the developed application are not string words but indexes. This would definitely help the trainer to focus on an index pointer rather than a form of a string word or number of letters. The indexes are listed in another attached table mapped to each word as the video clip suggests and says.

5.2.2. Data Portability

All programming objects are required to be portable and reusable in other platforms. Data is an object of the programming matter and it should be reusable anytime in the platform later. The origin of the data is video clips from a camera and its extension is not usable on every machine. Therefore, the first step is to reconstruct the videos in MP4 and AVI formats. The second step is to reformat them to RGB so they could be read in the application. For importing those videos, a Windows Form is designed to read and import those files; this form is existing in the project folder.

5.2.3. Videos to Vectors

The prepared Windows Form also converts these videos into vectors. The system uses same detecting algorithms that have been used to read lip movements. The question is why should the same system be used? To answer that, it should be noticed that train and test data should have the same attributes. So, to this matter, the same protocol converts those images to grayscale and then detects the face, and finally the lips. Those steps are be discussed in following sections in this chapter. Those vectors are being saved in data tables in the memory to be imported into for the next steps. Each field of this vector is an RGB system following 0-255 integer value.

5.2.4. Noise Removing

One common preprocessing techniques is removing the noise. In machine learning, attributes with no change are counted as noise. Those are called constant columns which have no effect in the learning rate but a waste of memory. So, those columns are removed and the memory is returned. At the end of preprocessing, the first version of training vectors

is ready to be tested. This still lacks a data verification stage which is explained in the following sub-section in details. Those techniques and extensions are all programmed in the attached application which is in the folder of the project.The Database is saved as form of CSV file shown in following below Table 5.1.

Table 5.1. Data samples

label	pixcel0	pixcel1	pixcel2	pixcel3	pixcel4	pixcel5	pixcel6	pixcel7	pixcel8
ONE	149	139	131	118	107	109	120	144	147
THREE	131	124	117	96	123	139	145	130	105
ONE	140	134	118	112	108	116	146	139	130
THREE	130	123	115	96	123	139	144	128	106
ONE	140	133	117	103	102	139	135	135	112
THREE	130	122	117	96	123	140	145	128	106
ONE	142	133	115	103	104	143	133	132	111
THREE	130	122	117	96	125	140	147	128	104
ONE	140	142	124	113	101	115	142	134	130
THREE	140	127	122	99	98	141	149	148	120
ONE	137	147	129	110	98	119	141	127	119
ONE	142	140	126	115	106	109	141	137	135
ONE	135	132	117	109	107	140	137	147	118
ONE	142	135	126	113	107	114	146	145	143
ONE	138	133	115	109	108	146	128	135	109
THREE	132	124	122	107	119	134	147	129	105
ONE	136	126	116	111	114	139	143	155	108
ONE	135	129	112	108	112	141	134	145	108
THREE	133	125	120	94	125	141	148	130	106
ONE	142	135	124	112	108	118	147	141	137

5.3. Data Verification

In real world, no data is 100% perfect and error free. The first version of our data has deficiently mislabeling and null reading problems. To remedy this issue, cross validation steps are proposed to overcome the null hypothesis occasions. There are also better techniques such as filtering and removing false negatives along with false positives. Hence, these steps would make our training data shrink. Therefore, the training and testing data would also occupy less memory and it will increase the performance of our machines. In the following sections, those techniques are fully described and analyzed for higher accuracy and boosting up.

5.3.1. Cross Validation

The cross validation is one of the most machine learning data filtering methods that are widely used when there is lack of testing data. In the proposed application, the training data goes through an automatic validation system. The system has been programed and attached with the main application. The cross validation basically divides the training data into the training set and the testing set. These divisions are measured with a static percentage such as 80 to 20 or 95 to 5, meaning that 95% of the labeled data is used as training set and the rest is used as testing set.

For this purpose, the original labeled data is imported into our developed utility application which becomes the validator main set. Afterwards, the training and testing sets are injected to an SVM object. The machine learns from the training data and it tests the testing data, then it tests the training data itself. After the tests, the results are scored against the true positives and the false positives. The scored table is used to remove all false decisions and renew the labeled data. This action is done in a repeated fashion until the data is at its maximum correctness. Cross validation on training set is shown in Figure 5.4.
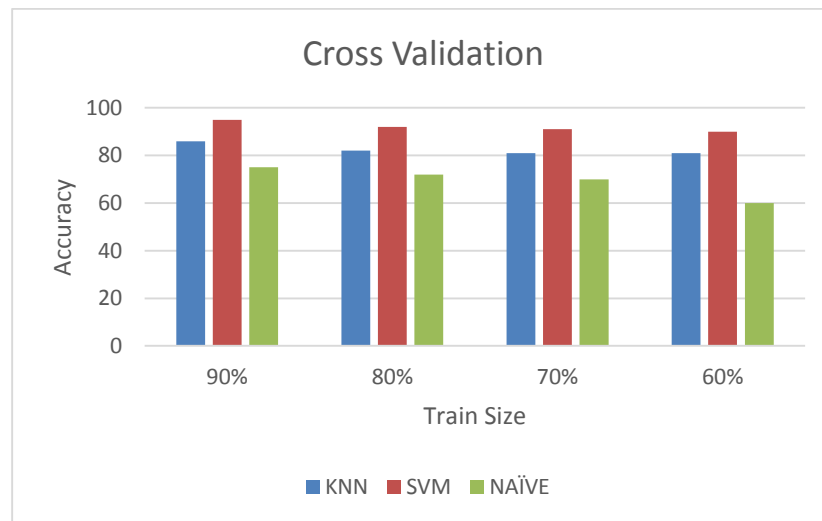

Figure 5.4 Cross validation on training set

5.3.2. Data Filtering

Data filtering comes after cross validation steps. This is done to ensure that the support vectors are significant and reliable. The filters are programed manually and dynamically for each set of data to decrease the error rate. The filters basically remove unsuitable samples which are occasionally duplicates and generate null memory. After both cross validation and data filtering stages, the training data is a smaller in size, but higher in accuracy. Furthermore, the sets are ready for the lip and face detectors.

5.4. Detections

The detection is carried out using image processing techniques from the OpenCV library and frameworks. There are color and size algorithms being used for this action to improve the detection process.

5.4.1. Face Detection

The action of face detection requires image processing tools and algorithms. One of the most common algorithms for this action is Haar Cascade. Haar is a machine learning based method in which a cascade function is trained from many positive and negative images saved in an XML database. Afterwards, it uses the database to decide whether an input sample is positive or negative. This database is imported into our system externally and is being run by Haar algorithm to detect the face of the speaker. The result of Haar algorithm is an array of faces where the first one is used. An example of the detected face is shown in Figure 5.5.
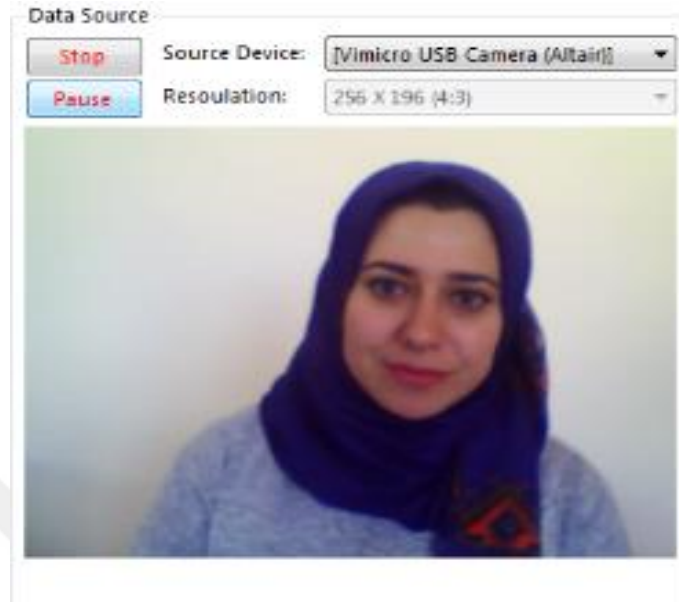
Figure 5.5 An example of the detected face

5.4.2. Lip Detection

The input for the lip detection is taken from the output of the detected face from the Haar algorithm. As mentioned before, the output is a single image even if the number of faces are multiple in the captured clip. Therefore, the closest person is being read and classed with the prepared algorithm. In the second step, the lips are detected using four different ways:

1. Haar cascade: this has been already identified and explained. However, in this section, the algorithm does not detect faces, but lips. This is also done using a dictionary in order to detect the result of a given input. The problem in this case is singularity because obviously there is only one mouth per person.

2. Color of lips: The lips can be identified using their natural color. For coloring, YCbCr, RGB and HSV have been employed and their average is the result of detection.

3. Color of face: Another way to detect lips is to detect the color of face and remove it from the image. By using this way, the skin color is being set as 0 and the lips are the result of the difference.

All coloring methods seem promising; however, they have a few challenges, including human races, color of skin, the reflection of light and types of cameras. For these issues,

mostly the average of all coloring systems are being focused on. After detecting the lips, the image is sent to the classification process to be read and predicated. An example of lips detecting is shown in Figure 5.6.
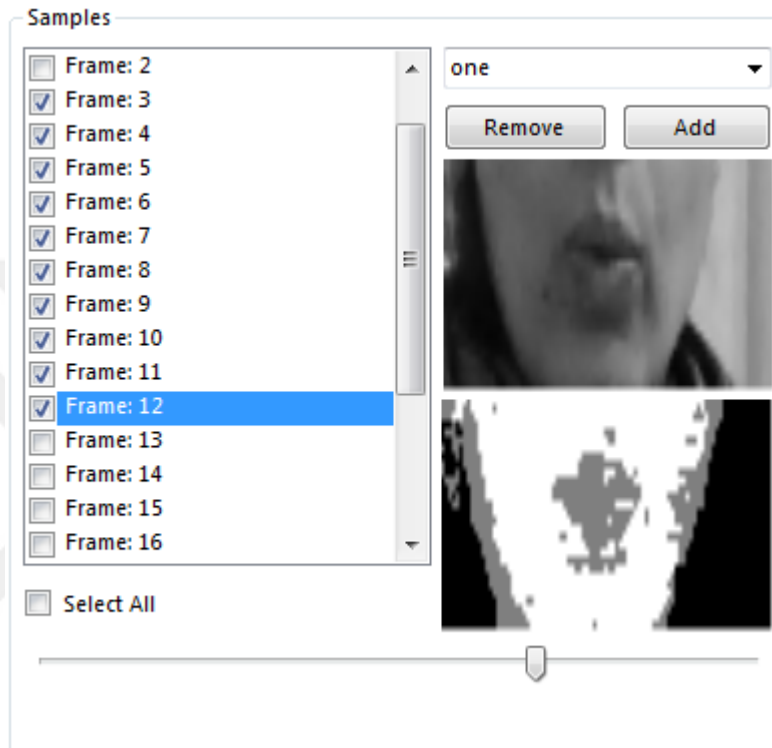

Figure 5.6 An example of lips detection

4. Locating the lips: by logic, the mouth of a human face is located in the lower part of the captured image. Therefore, an easy approach could be cropping the image to the lower half of the face and then zooming it. Color based detection are shown in Figure 5.7.

Figure 5.7 Color based detection

5.5. Classification

In this thesis, K-NN, SVM, Naïve Bayes and Levenberg-Marquardt are the four algorithms used for classification purpose. SVM is the algorithm used in computer vision. This method is working with higher dimensional space, by combining the features the new dimension is creating. K-NN is also used for classification, and the class with the highest frequency using the K-most algorithm is selected as the output for each instance. Also Naïve Bayes is a simple method for producing classifiers. The neural network mimics the human brain's method to recognize and predict values.. For all these classification methods, each instance is counted as a vote for their class, and the class with highest votes wins. At first, face and facial part are detected, then separated, which retrieves some information about the upper and lower parts. In further steps, the classification of a given sequence of images in speechless and speech modes is conducted, as shown in the results presented in Figure 5.8.

Figure 5.8 Sequence of images speechless and speech

5.6. Prediction

The most important part of the application is the prediction section, where all the trainer machine learning algorithms give their assumption towards an inputted sample. In this implementation, the prediction function is run by an independent thread which gives the ability to capture, detect, and predict at the same time, less than or equal to the time of each frame. This also depends on the response ability of processor and hardware types. The machine learning algorithms give their assumptions separately and their means are calculated to obtain the best result. There are three different ways each machine learning algorithm obtains the outcome:
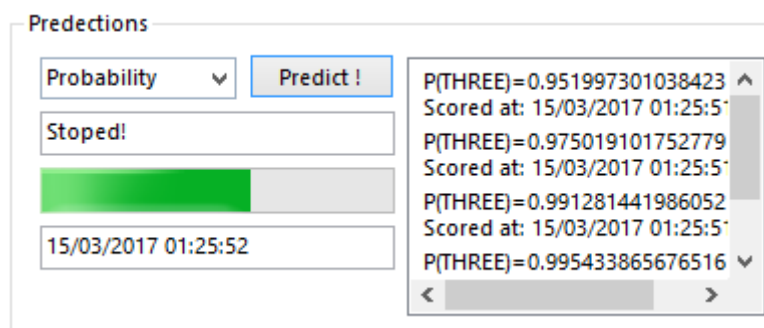
1. Scoring
2. Probability
3. Log Likelihood

Figure 5.9 Prediction Tools

36

5.6.1. Score Results

This method applies to all machine learning algorithm types. The score of each class obtained to each sample is often used in linear problems. The score is calculated as the square differences of the distance of the sample to the nearest prediction.

5.6.1.1. Probability

The K-NN algorithm lacks the needed probability method to score the classes. However, other methods use this way to calculate the assumption. The probability of each lip reading class is a fractional number between 0 and 1.

5.6.1.2. Log Likelihood

In statistics, maximum likelihood estimation is a method of approximating the parameters of a statistical model given observations, by finding the parameter values that maximize the likelihood of making the observations of the given values.

5.7. Programming Concepts

The application is designed using C# Windows Forms. It has a number of classes and data structures which are discussed in this section.

5.7.1. Classes

There are a number of classes working together:

5.7.1.1. LipReader.Database

A CSV (Comma Separated Values) file is basically saved in this structure which has a set of classes and a list of sample class. The default instance of the database is loaded into the application. LipReader.Database class members are shown in Figure 5.10.
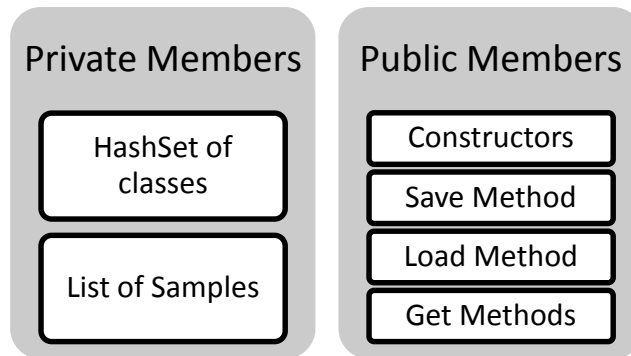
Figure 5.10 LipReader.Database class members

5.7.1.2. LipReader.Sample

Each sample of our application is created from this class. It is constructed from the vector of doubles and a string. LipReader.Sample class members are shown in Figure 5.11.
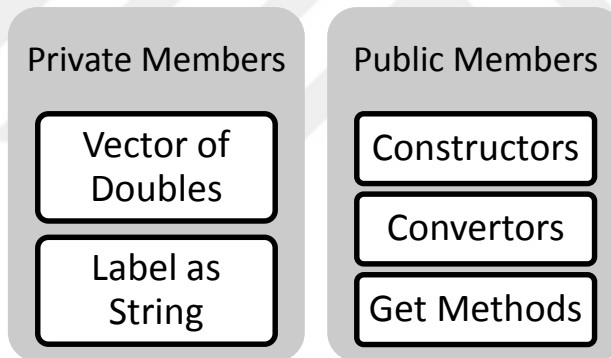

Figure 5.11 LipReader.Sample class members

5.7.1.3. LipReader.Goals

The goal structure has instances that are the result of the prediction. Each time a lip is detected with a high accuracy, an instance of this structure is created. . LipReader.Goals class members are shown in Figure 5.12.
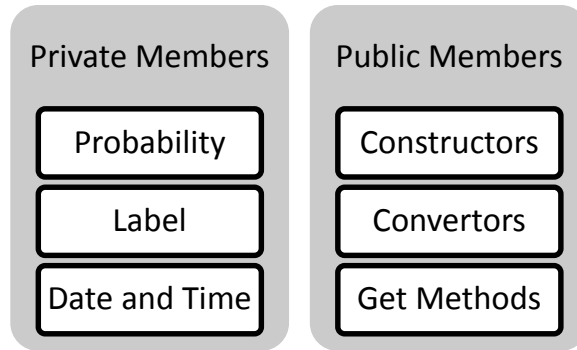
Figure 5.12 LipReader.Goals class members

5.7.1.4. LipReader. IColorSkinDetector

This is the base class for both our HSV and YCbCr methods which are used to detect color classes. Those classes detect the color of skin, and the application removes the detected skin color from the lip surroundings. LipReader.IColorSkinDetector class members are shown in Figure 5.13.
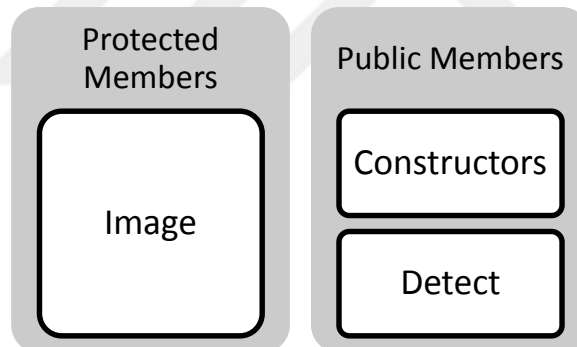

Figure 5.13 LipReader. IColorSkinDetector class members

5.7.1.5. LipReader.Info

The info structure holds information about the training data, the lip Haar files and all the information about addressing and variable paths. It also saves a log of those information:

- Total recorded frames
- Total detected frames
- Current frame detected
- Min neighbor for HAAR algorithm

- Face HAAR XML address file

- LIP HAAR XML address file

- The size of each detection and their scale.

- Detection types

5.7.1.6. LipReader.Video_Device

This structure finds the number of installed cameras on the computer.

5.7.1.7. LipReader.Teacher

The main class which creates the machine learning algorithms and trains them to become ready to predict. There will be only one object created from this class per run. LipReader.Teature class members are shown in Figure 5.14.

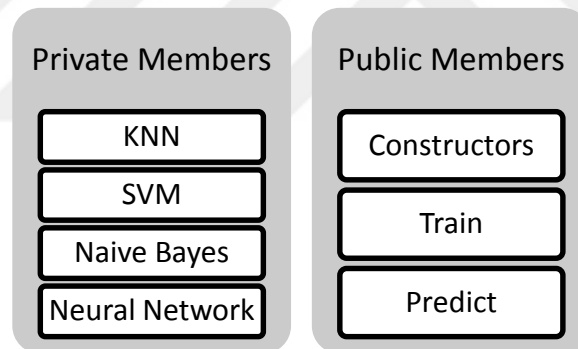**Figure 5.14** LipReader.Teature class members

5.7.1.8. LipReader.Program

Any C# program has the main program class which runs at the beginning. Creating a platform that Windows application can run on it.

5.7.1.9. LipReader.PreData

This is a utility window form creates the training data out of the recorded videos. This application is only run by the developer, so the user cannot see it.

5.7.1.10. LipReader.Train

The main application form which has the training set included and a test mechanism programmed along with it.

5.7.2. Global Memory

The global memory which has been used in the training form is divided on numbers of structure types. The global memory is shown in Figure 5.15.
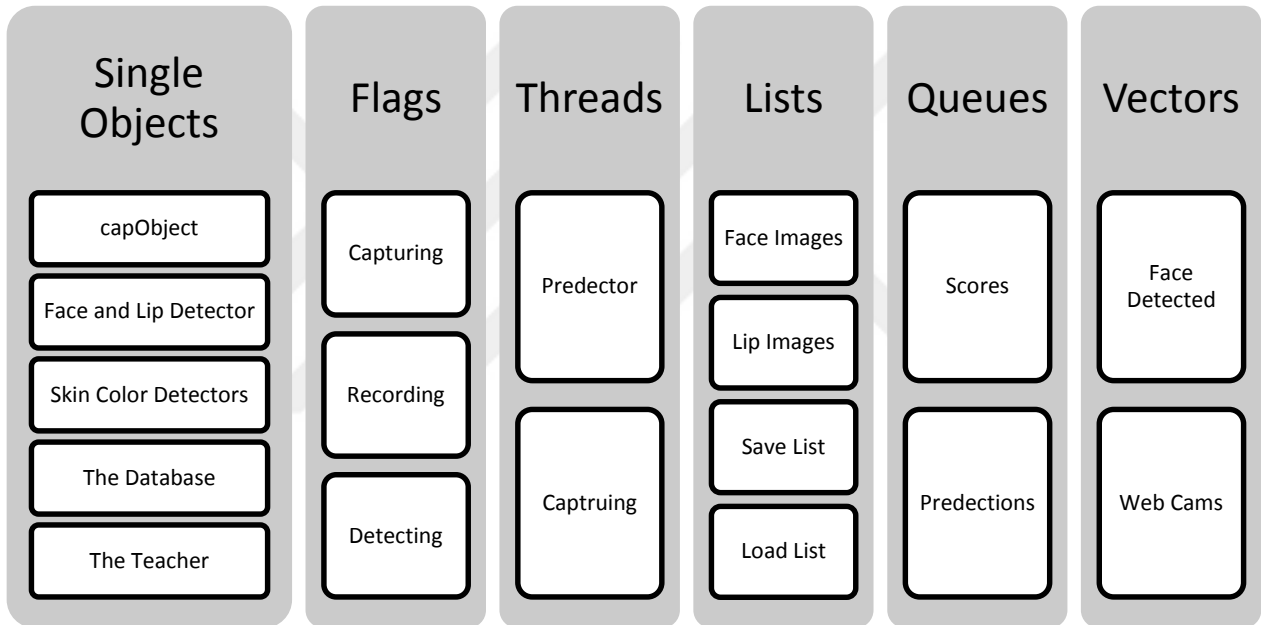
Figure 5.15 Global memory

5.7.2.1. Single Objects

These objects are created out of the mentioned classes and each of them has a specific task.

- capObject is the OpenCV object to capture images from camera.
- Face and Lip Detector are made of OpenCV Haar classifiers to detect the faces.
- Skin Color detector is created out of the skinColor detector class.
- The database is the CSV file loaded into the database structure.
- The Teacher is the object created out of the teacher class.

5.7.2.2. Flags

The flags are the identifiers of the methods which can show each thread status.

5.7.2.3. Threads

The threads run in parallel with the main method. They run for the following different tasks:

- Predictor runs the queue of received image and predict them using the teacher object. This runs only if there are items
- The capturing thread captures images from the camera device.

5.7.2.4. Lists

The list structure is used when the limited array size has no use to our implementation.

5.8. Implementation Problems

There have been a number of difficulties with detecting the boundaries of lips due to its variation among candidates and people. Since we locate it from the down part of the face, it is based on the knowledge of its location. The size also matters, there is a variation in size; however, there is set of attributes to be modified to the rate of the face in order to overcome this problem.

Frame accuracy: there is always a chance of increasing the frame reading accuracy; however, this will increase the data size too. It is recommended to have a balance between the quality and performance, while both work against each other. For this issue, there is a quality checker option, which allows the user to modify it to meet his/her device capability limit.

Continuous Training: the developed machine learning algorithms require training and it could be unlimited. The new data is a good example for training the machine learning algorithms, it must all be true positives. The data is checked manually in case the developing flag is set on. The new training data is added to the database by the semi-developer, the frontend user who has a developing ability.

5.9. Algorithms Testing Using Cross Validation Method

K-NN Test:

K-NN always shows a particular high results on the cross validation test. The test has been accumulated on four different distance functions shown in Figure 5.16 charts. The best results are achieved with Chebyshev distance function. The Manhattan and Euclidian distance functions shows no significant change by altering the number of neighbors. On the other hand, Hellinger's accuracy decreases by increasing the number of neighbors. The complete K-NN test result is shown in Figure 5.16.

Figure 5.16 KNN test

SVM Test:

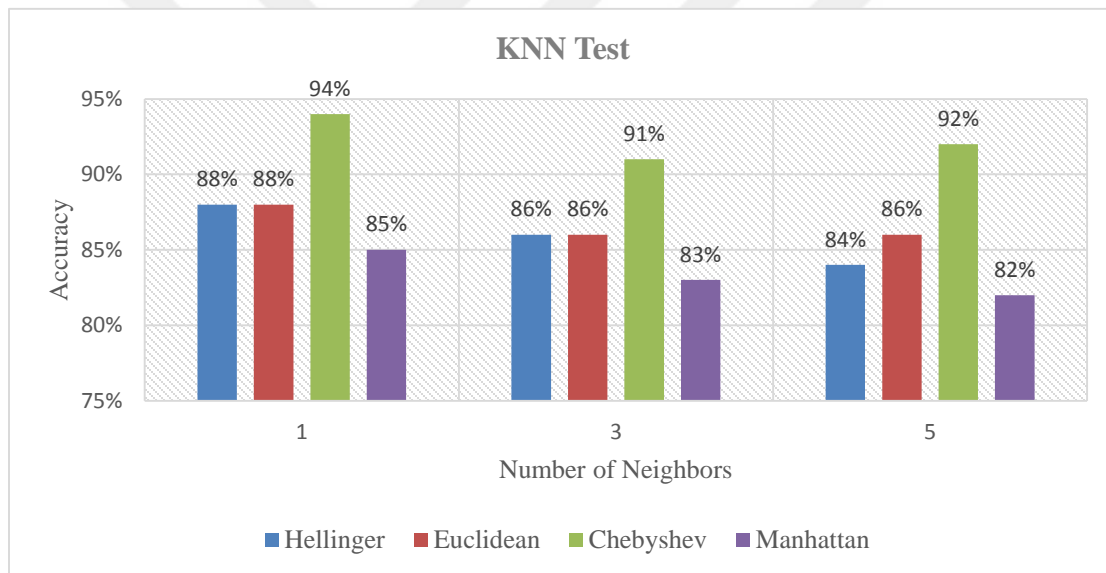The SVM always shows the best result in binary classification, which will not apply to the developed project. For testing this machine learning algorithm, four different kernel functions have been used. All four functions show higher accuracy when the train size is increased. The best algorithm to be used regarding the lip reading in the SVM is ChiSquare. SVM test result is shown in Figure 5.17.

43
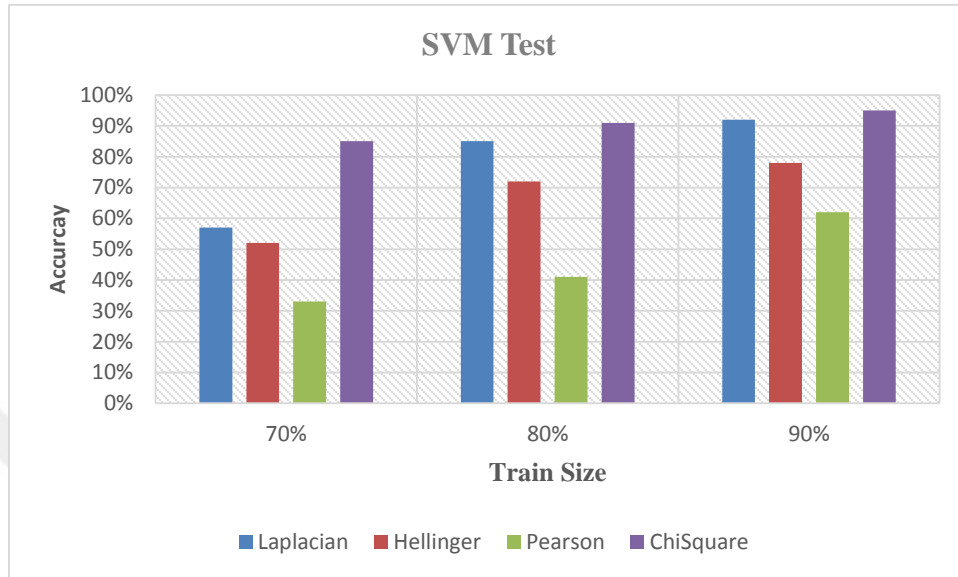
Figure 5.17 SVM test

Neural Network Test:

The Back Propagation and Levenberg Marquardt are two learning methods that have been tested and obtained significant results in this project. The simple neural network is not the best option according to the results of the test. Although in some cases it has a significant accuracy. The best configuration regarding the test development is when the learner Levenberg Marquardt is run for 20-25 iterations on 2 layers. Neural network test result is shown in Figure 5.18.

Hidden Layers	Learning Iterating			Back Propagation Learning			Hidden Layers
	20	30	40				
1	33%	57%	53%	51%	52%	52%	3
2	66%	55%	35%	39%	40%	41%	5
3	41%	54%	56%	53%	54%	60%	10
Levenberg Marquardt Learning				40	50	60	
				Learning Iterating			

Figure 5.18 Neural network test

All Algorithms Test:

Testing all four machine learning algorithms together is a technique to find the best combination of them. The accuracy is not the only factor that is why a timer is added to check how much time is it required to use a combination. As it is shown in Figure 5.19, when neural network is added to the recipe, the performance will get slower. The accuracies of SVM and K-NN score the highest, while the worst and the fastest classification algorithm is Naïve Bayes. All classification algorithms test result is shown in Figure 5.19.
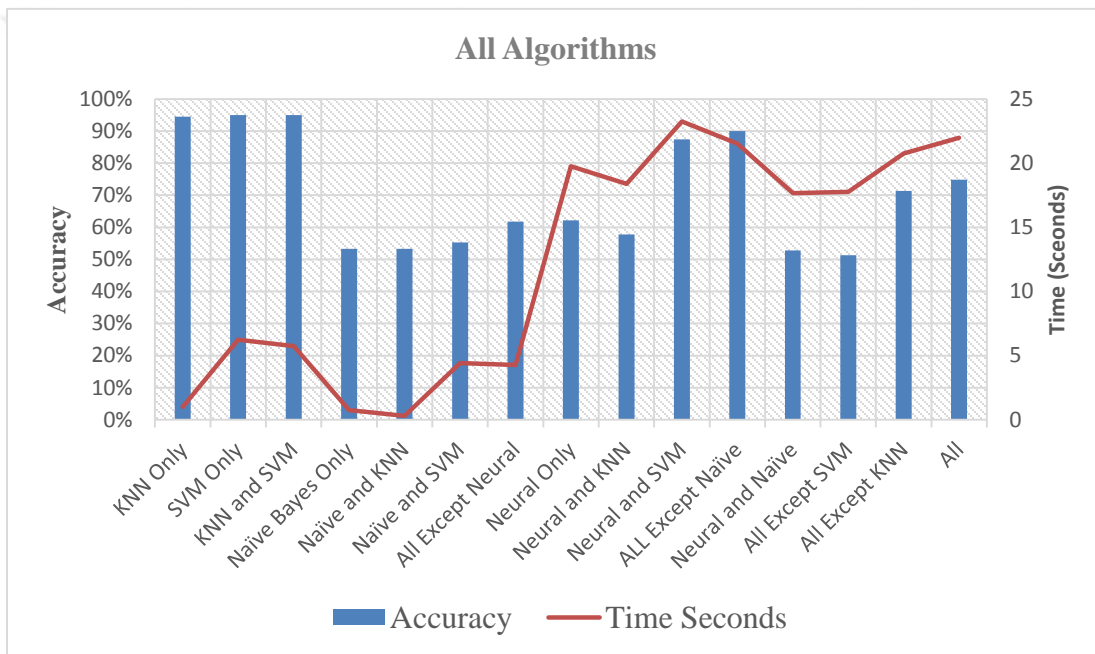
Figure 5.19 All algorithms test

Male and Female Reading Test:

The results of reading different genders are significant. It is illustrated in Figure 5.20 that male lip readings have been way more significant than female readings. The only time, female lip detection and reading are showing a better performance, is when the three classification algorithms SVM, KNN and Naïve Bayes try to predict the results. This combinaïon also shows a close match when the neural network is added to the program. Meanwhile, neural network in a solo classification algorithm does not show any significant results on the female or male readings. Male and Female reading accuracy is shown in Figure 5.20.

Figure 5.20 Male and Female reading accuracy

6. CONCLUSIONS

In this thesis, lip recognition using OpenCV system was presented. This approach used the Haar cascade and four classification algorithms from machine learning which are: Support Vector Machine (SVM), K-Nearest Neighbor (K-NN), Naïve Bayes and Levenberg-Marquardt. The Haar cascade was used to find the s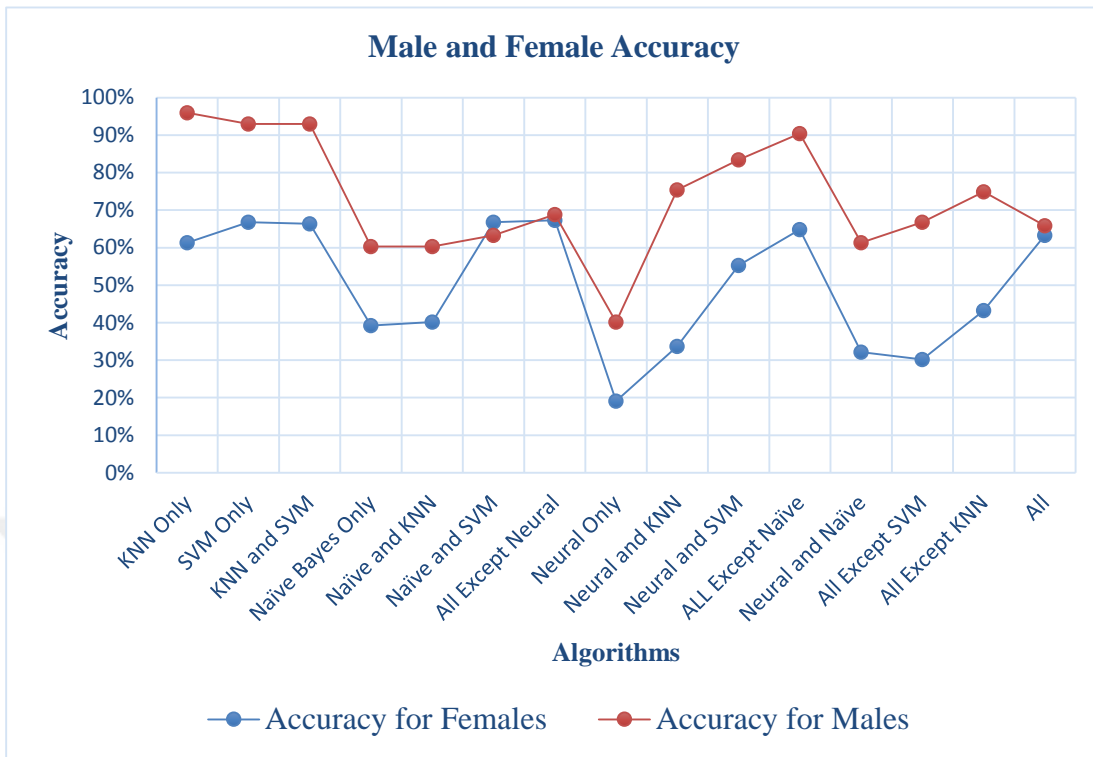peaker's face and then the mouth region of the face. SVM, KNN, Naïve Bayes and Levenberg-Marquardt were used to classify the data, however, for less than ten shapes, K-NN worked better on its own. The developed approach is able to detect and recognize lips effectively and accurately compared to some methods found in the literature. From the results of analysis, it is clearly found that the method is superior because it allows the user to adapt the system for different shapes of faces. The lips were found automatically in the lower part of the mouth. The frame rate was adjustable according to the capabilities of the machine. A higher frame rate means much more data to process and thus it will be too heavy for a slower machine. The implemented system has the ability to accept more training data, which means that additional face shapes other than the encountered ones can be used. This makes it more adaptive than other techniques. Although the proposed method has demonstrated a superior performance, further validation and practical development are needed to ensure the feasibility of the proposed system. As further development is needed, the system should be re-evaluated and modified in order to obtain higher accuracy and better performance.

7. REFERENCES

[1]   **Skodras, E. and Fakotakis, N.,** 2011, May. An unconstrained method for lip detection in color images. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (pp. 1013-1016). IEEE.

[2]   **Zhang, X. and Mersereau, R.M.,** 2000. Lip feature extraction towards an automatic speechreading system. In *Image Processing, 2000. Proceedings. 2000 International Conference on* (Vol. 3, pp. 226-229). IEEE.

[3]   **Liew, A.W.C., Leung, S.H. and Lau, W.H.,** 2002. Lip contour extraction from color images using a deformable model. *Pattern Recognition*, *35*(12), pp.2949-2962.

[4]   **Delmas, P., Coulon, P.Y. and Fristot, V.,** 1999, March. Automatic snakes for robust lip boundaries extraction. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings. 1999 IEEE International Conference on* (Vol. 6, pp. 3069-3072). IEEE.

[5]   **Liu, X., Cheung, Y.M., Li, M. and Liu, H.,** 2010, August. A lip contour extraction method using localized active contour model with automatic parameter selection. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 4332-4335). IEEE.

[6]   **Eveno, N., Caplier, A. and Coulon, P.Y.,** 2002, December. A parametric model for realistic lip segmentation. In *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on* (Vol. 3, pp. 1426-1431). IEEE.

[7]   **Meier, U., Stiefelhagen, R., Yang, J. and Waibel, A.,** 2000. Towards unrestricted lip reading. *International Journal of Pattern Recognition and Artificial Intelligence*, *14*(05), pp.571-585.

[8]   **Zheng, G.L., Zhu, M. and Feng, L**., 2014, December. Review of Lip-Reading Recognition. In *Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium on* (Vol. 1, pp. 293-298). IEEE.

[9]   **Petajan, E.D.,** 1984. *Automatic lipreading to enhance speech recognition (speech reading)* (Doctoral dissertation, University of Illinois at Urbana-Champaign).

[10]   **Bregler, C. and Omohundro, S.M.,** 1994. *Surface learning with applications to lipreading*. International Computer Science Institute.

[11]   **Pingxian, Y., Rong, G., Peng, G. and Zhaoju, F.,** 2011, December. Research on lip detection based on Opencv. In *Transportation, Mechanical, and Electrical Engineering (TMEE), 2011 International Conference on* (pp. 1465-1468). IEEE.

[12]   **Lewis, T.W. and Powers, D.M.,** 2000, December. Lip feature extraction using red exclusion. In *Selected papers from the Pan-Sydney workshop on Visualisation-Volume 2* (pp. 61-67). Australian Computer Society, Inc.

[13]   **WenJuan, Y., YaLing, L. and MingHui, D.,** 2010, August. A real-time lip localization and tacking for lip reading. In *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on* (Vol. 6, pp. V6-363). IEEE.

[14]   **Zhang, Z.W. and Shen, H.B.,** 2008. Lip detecting algorithm based on chroma distribution diversity [J]. *Journal of Zhejiang University (Engineering Science)*, *8*, p.013.

[15]   **Kuo, P., Hillman, P. and Hannah, J.,** 2005. Improved facial feature extraction for model-based multimedia. In *Proceedings 2nd IEE European Conference on Visual Media Production* (pp. 137-146).

[16]   **Sadeghi, M., Kittler, J. and Messer, K.,** 2001, September. Real time segmentation of lip pixels for lip tracker initialization. In *International Conference on Computer Analysis of Images and Patterns* (pp. 317-324). Springer Berlin Heidelberg.

[17]   **Waibel, A., Stiefelhagen, R. and Yang, J.,** 1996. A modelbased gaze trackeing system. In *IEEE International Joint Symposia on Intelligence and Systems*.

[18]   **Eveno, N., Caplier, A. and Coulon, P.Y.,** 2004. Accurate and quasi-automatic lip tracking. *IEEE Transactions on Circuits and Systems for Video technology*, *14*(5), pp.706-715.

[19]   **Neti, C., Potamianos, G., Luettin, J., Matthews, I., Glotin, H. and Vergyri, D.,** 2001. Large-vocabulary audio-visual speech recognition: A summary of the Johns Hopkins Summer 2000 Workshop. In *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on* (pp. 619-624). IEEE.

[20]   **Potamianos, G., Neti, C., Gravier, G., Garg, A. and Senior, A.W**., 2003. Recent advances in the automatic recognition of audiovisual speech. *Proceedings of the IEEE*, *91*(9), pp.1306-1326.

[21]   **Nefian, A.V., Liang, L., Pi, X., Liu, X. and Murphy, K.,** 2002. Dynamic Bayesian networks for audio-visual speech recognition. *EURASIP Journal on Advances in Signal Processing*, *2002*(11), p.783042.

[22]   **Lewis, T.W. and Powers, D.M.,** 2003. Audio-visual speech recognition using red exclusion and neural networks. *Journal of Research and Practice in Information Technology*, *35*(1), pp.41-64.

[23]   **Petajan, E. and Graf, H.P.,** 1996, October. Robust face feature analysis for automatic speechreading and character animation. In *Automatic Face and Gesture Recognition, 1996. Proceedings of the Second International Conference on* (pp. 357-362). IEEE.

[24]   **Yang, J., Stiefelhagen, R., Meier, U. and Waibel, A.,** 1998. Real-time face and facial feature tracking and applications. In *AVSP'98 International Conference on Auditory-Visual Speech Processing*.

[25]   **Jang, K.S., Han, S., Lee, I. and Woo, Y.W.,** 2006, December. Lip localization based on active shape model and gaussian mixture model. In *Pacific-Rim Symposium on Image and Video Technology* (pp. 1049-1058). Springer Berlin Heidelberg.

[26]   **Kaucic, R., Dalton, B. and Blake, A.,** 1996, April. Real-time lip tracking for audio-visual speech recognition applications. In *European Conference on Computer Vision* (pp. 376-387). Springer Berlin Heidelberg.

[27]   **Zhang, X. and Mersereau, R.M.,** 2000. Lip feature extraction towards an automatic speechreading system. In *Image Processing, 2000. Proceedings. 2000 International Conference on* (Vol. 3, pp. 226-229). IEEE.

[28]   **Delmas, P., Coulon, P.Y. and Fristot, V.,** 1999, March. Automatic snakes for robust lip boundaries extraction. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings. 1999 IEEE International Conference on* (Vol. 6, pp. 3069-3072). IEEE.

[29]   **Liu, X., Cheung, Y.M., Li, M. and Liu, H.,** 2010, August. A lip contour extraction method using localized active contour model with automatic parameter selection.

In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 4332-4335). IEEE.

[30]   **Liew, A.W.C., Leung, S.H. and Lau, W.H**., 2002. Lip contour extraction from color images using a deformable model. *Pattern Recognition*, *35*(12), pp.2949-2962.

[31]   **Eveno, N., Caplier, A. and Coulon, P.Y.,** 2002, December. A parametric model for realistic lip segmentation. In *Control, Automation, Robotics and Vision, 2002. ICARCV 2002. 7th International Conference on* (Vol. 3, pp. 1426-1431). IEEE.

[32]   **Meier, U., Stiefelhagen, R., Yang, J. and Waibel, A.,** 2000. Towards unrestricted lip reading. *International Journal of Pattern Recognition and Artificial Intelligence*, *14*(05), pp.571-585.

[33]   **WenJuan, Y., YaLing, L. and MingHui, D.,** 2010, August. A real-time lip localization and tacking for lip reading. In *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on* (Vol. 6, pp. V6-363). IEEE.

[34]   **Rao, R.A. and Mersereau, R.M.,** 1994. Lip modeling for visual speech recognition Signals, Systems and Computers. In *Conference Record of the Twenty-Eighth Asilomar Conference on* (Vol. 31, pp. 587-590).

[35]   **Lewis, T.W. and Powers, D.M.,** 2000, December. Lip feature extraction using red exclusion. In *Selected papers from the Pan-Sydney workshop on Visualisation-Volume 2* (pp. 61-67). Australian Computer Society, Inc.

[36]   **Pera, V., Sa, F., Afonso, P. and Ferreira, R**., 2003, December. Audio-visual speech recognition in a Portuguese language based application. In *Industrial Technology, 2003 IEEE International Conference on* (Vol. 2, pp. 688-692). IEEE.

[37]   **Liévin, M. and Luthon, F.,** 2000. A hierarchical segmentation algorithm for face analysis. Application to lipreading. In *Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on* (Vol. 2, pp. 1085-1088). IEEE.

[38]   **Jun, H. and Hua, Z.,** 2007, July. A real time lip detection method in lipreading. In *Control Conference, 2007. CCC 2007. Chinese* (pp. 516-520). IEEE.

[39]   **Lie, W.N. and Hsieh, H.C.,** 1998. Lips detection by morphological image processing. In *Signal Processing Proceedings, 1998. ICSP'98. 1998 Fourth International Conference on* (Vol. 2, pp. 1084-1087). IEEE.

[40]   **Pao, T.L. and Liao, W.Y.,** 2005, August. A motion feature approach for audio-visual recognition. In *Circuits and Systems, 2005. 48th Midwest Symposium on* (pp. 421-424). IEEE.

[41]   **Da Silveira, L.G., Facon, J. and Borges, D.L.,** 2003, October. Visual speech recognition: a solution from feature extraction to words classification. In *Computer Graphics and Image Processing, 2003. SIBGRAPI 2003. XVI Brazilian Symposium on* (pp. 399-405). IEEE.

[42]   **Jordan, T.R., Mccotter, M.V. and Thomas, S.M.,** 2000. Visual and audiovisual speech perception with color and gray-scale facial images. *Attention, Perception, & Psychophysics*, *62*(7), pp.1394-1404.

[43]   **Zhang, X.J., Montoya, H.A. and Crow, B.,** 2007, June. Finding lips in unconstrained imagery for improved automatic speech recognition. In *International Conference on Advances in Visual Information Systems* (pp. 185-192). Springer Berlin Heidelberg.

[44]   **Schneiter, F.,** 2009. Lip Contour Localization using Statistical Shape Models.

[45] **Ooi, W.C., Jeon, C., Kim, K., Han, D.K. and Ko, H**., 2008, August. Effective lip localization and tracking for achieving multimodal speech recognition. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on* (pp. 90-93). IEEE.

[46] **Li, M. and Cheung, Y.M.,** 2009. Automatic lip localization under face illumination with shadow consideration. *Signal Processing*, *89*(12), pp.2425-2434.

[47] **Husain, B.N.,** 2011. Face Detection and lip localization.

[48] **Liew, A.W.C. and Wang, S.,** 2009. Visual Speech Recognition: Lip Segmentation and.

[49] **Bouvier, C., Coulon, P.Y. and Maldague, X.,** 2007, September. Unsupervised lips segmentation based on ROI optimisation and parametric model. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on* (Vol. 4, pp. IV-301). IEEE.

[50] **Bouvier, C., Coulon, P.Y. and Maldague, X.,** 2007, September. Unsupervised lips segmentation based on ROI optimisation and parametric model. In *Image Processing, 2007. ICIP 2007. IEEE International Conference on* (Vol. 4, pp. IV-301). IEEE.

[51] **Gacon, P., Coulon, P.Y. and Bailly, G.,** 2005, March. Statistical active model for mouth components segmentation. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP'05). IEEE International Conference on* (Vol. 2, pp. ii-1021). IEEE.

[52]   **Sadeghi, M., Kittler, J. and Messer, K.,** 2002. Modelling and segmentation of lip area in face images. *IEE Proceedings-Vision, Image and Signal Processing*, *149*(3), pp.179-184.

[53]   **Cheung, Y.M., Liu, X. and You, X.,** 2012. A local region based approach to lip tracking. *Pattern Recognition*, *45*(9), pp.3336-3347.

[54]   **Yuille, A.L., Hallinan, P.W. and Cohen, D.S.,** 1992. Feature extraction from faces using deformable templates. *International journal of computer vision*, *8*(2), pp.99-111.

[55]   **Zhang, X. and Mersereau, R.M**., 2000. Lip feature extraction towards an automatic speechreading system. In *Image Processing, 2000. Proceedings. 2000 International Conference on* (Vol. 3, pp. 226-229). IEEE.

[56]   **Dargham, J.A. and Chekima, A.,** 2006, April. Lips detection in the normalised RGB colour scheme. In *Information and Communication Technologies, 2006. ICTTA'06. 2nd* (Vol. 1, pp. 1546-1551). IEEE.

[57]   **Liew, A.W.C., Leung, S.H. and Lau, W.H.,** 2002. Lip contour extraction from color images using a deformable model. *Pattern Recognition*, *35*(12), pp.2949-2962.

[58]   **Leung, S.H., Wang, S.L. and Lau, W.H.,** 2004. Lip image segmentation using fuzzy clustering incorporating an elliptic shape function. *IEEE transactions on image processing*, *13*(1), pp.51-62.

[59]   **Liu, X., Cheung, Y.M., Li, M. and Liu, H.,** 2010, August. A lip contour extraction method using localized active contour model with automatic parameter selection. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 4332-4335). IEEE.

[60]   **OpenCV, Open Source Computer Vision**,

http://docs.opencv.org/master/d7/d8b/tutorial_py_face_detection.html#gsc.tab=0,

Last accessed date: March 25, 2017.

[61]   **Haji, S. and Varol, A**., 2016, April. Real time face recognition system (RTFRS).

In *Digital Forensic and Security (ISDFS), 2016 4th International Symposium on* (pp. 107-111). IEEE.

[62]   **Viola, P. and Jones, M.,** 2001. Rapid object detection using a boosted cascade of

simple features. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (Vol. 1, pp. I-I). IEEE.

[63]   **Azad, B. and Varol**, A., 2017, January. A New Approach for Lip Recognition.

International Journal of Engineering Research & Technology (IJERT), 2017.

[64]   **Brahme, A. and Bhadade, U.,** 2016. Lip detection and lip geometric feature

extraction using constrained local model for spoken language identification using visual speech recognition. *Indian Journal of Science and Technology*, *9*(32).

[65]   **Intro to Machine Learning**,

https://www.udacity.com/course/intro-to-machine-learning--ud120, Last accessed date: March 25, 2017.

[66]   **Wu, J., Cui, Z., Sheng, V.S., Shi, Y. and Zhao, P**., 2014. Mixed pattern matching-

based traffic abnormal behavior recognition. *The Scientific World Journal*, *2014*.

[67] **Understanding Support Vector Machine algorithm from examples**,
https://www.analyticsvidhya.com/blog/2015/10/understaing-support-vector-machine-example-code/, Last accessed date: March 25, 2017.

[68] **Neural network models**,
http://scikit-learn.org/stable/modules/neural_networks_supervised.html, Last accessed date: March 25, 2017.

Bnar Azad H.Ameen

Arbil, Kurdistan, Iraq

Email: bnarazad60@gmail.com, Mobile No: +964 – 7501196732

DOB: 1987

CAREER INTERETS

Teaching professional career that is imaginative and includes several artistic medium to encourage and inspire students by innovative and creative thoughts.

EDUCATION

- **B.Sc.,** (Computer Science) 2006-2010: from *Sulaimania University,* Kurdistan, Iraq
- **Higher School,** 2005-2006: from **Rwandz Preparatory for boys**, Kurdistan, Iraq

EMPLOYMENT AND EXPERIENCE

- December 2010 – Present, Assistant Programmer, Faculty of Science and Engineering, Department of Computer Science, **Soran University**, Kurdistan, Iraq
  Duties involved:

  o Assisting the lecturers in teaching.
  o Lab Assistant.

- December 2009 – Present, Teacher, **Rwandz High School**, Rwandz, Kurdistan, Iraq
  Duties involved:

  o Teaching Mathematics and Computer Science for the high school students.

- December 2009 – Present, Trainer, **ICDL Center**, Soran, Kurdistan, Iraq
  Duties involved:
  o Working as a trainer where I teach IT concepts, Internet & Email, Database, spreadsheets, Presentation, Windows.

COMPUTING SKILLLS

- Programming skills: C++, C# .NET
- Operating Systems: Windows, Mac
- Applications: Photoshop, MS office
- Database: SQL Server 2008
- Web Technologies: HTML, PHP

PROJECTS

B.Sc., Project Title – **Jobseeker System**

Description

With the growing use of the internet as a tool in the job search process, an important field is started which was finding the job for the people by the internet and gives a large advantage in reducing the time and the cost.

Job seeker is a person who is looking for an occupation and is registered in a work office to get it.

EXTRACURRICULARS

- Holding the position of "*Manager in research and development unit*" in Department of Computer Science, Soran University.
- Successfully organized seminars in the Department of Computer Science, Soran University.

INTERESTS

- Listening to Music
- Watching movies
- Shopping
- Hang out with friends

PERSONAL DETAILS

Nationality: Iraqi

Known Languages: English, Arabic, Kurdish