CALIFORNIA STATE UNIVERSITY, NORTHRIDGE

Face Recognition using Fuzzy Logic

A graduate project submitted in partial fulfillment of the requirements

For the degree of Master of Science in Electrical Engineering

By

Suchitra Tapaskumar Basak

August 2017

The graduate project of Suchitra Basak is approved:

Dr. Xiyi Hang Date

Dr. Shahnam Mirzaei Date

Dr. Ruting Jia, Chair Date

California State University, Northridge

Acknowledgement

I would like to express my deep gratitude to my mentor Dr. Ruting Jia for motivating me and giving me this opportunity which has been a very academically fulfilling experience.

I appreciate the support I received from my parents, sister and Vivek. Thank you for believing in me and my ideas.

Table of Contents

Abstract

Face Recognition using Fuzzy Logic

By

Suchitra Tapaskumar Basak

Master of Science in Electrical Engineering

This project demonstrates how Fuzzy Logic can be used for recognizing human faces. Accuracy of upto about 80% can be obtained with this method on the BioID Database. This method is robust since images in the database are taken in uncontrolled environments. Feature extraction is based on the location of Landmark points. The degree of uncertainty that fuzzy logic presents is used to classify faces. Thus, same people having different expressions in different pictures are recognized correctly. This shows that with the help of the right features extracted, Fuzzy Logic can be used as a successful Facial Recognition technique in controlled as well as uncontrolled environments.

Chapter 1: Introduction

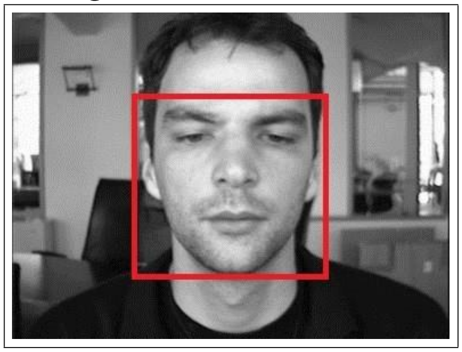
Facial Recognition is the ability of a system to study an image or a video and identify the participant/s by studying their faces present in the image or videos. It is the process of labeling faces with their respective individuals. Along with fingerprints, voice *etc.* the face is a unique biometric feature. It is sometimes the most accessible one- for example, in an image or a surveillance video clip.

Face Recognition systems, in most cases, need to possess at least some amount of robustness. Since in practical situations, all the images available may or may not be in the most desirable form which includes perfect lighting, angles or perfect facial expressions.

To human beings, face recognition is intuitive. The human brain is constantly predicting objects in its surroundings and making corrections if there is an error. The amount of training data that human beings are exposed to is massive. Since human beings start to obtain “training data” since a very young age. Artificial neural networks, which mimic the functions of the human brain, can be used in a similar way for recognizing faces [6].

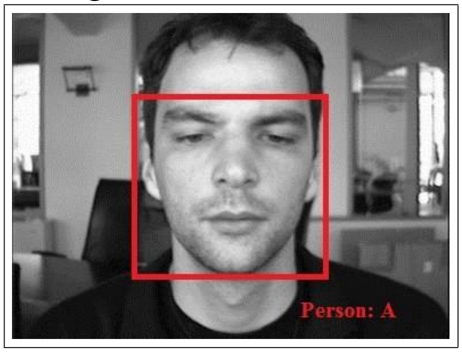
Face Detection vs. Face Recognition Face detection is the ability of a system to convey the presence of a human face anywhere in an image. It is a subset of object detection. The input to a face detection program would be an image and the output would be binary- whether there is a face present of no. The coordinates of the object detected can be obtained too. An example of the input and output of a system that performs face detection is shown in figure 1.2

Figure 1.1: Face Detection



Face Recognition on the other hand is finding out the person the face belongs to. The input of a recognition system would be the face of a person and the output would be the name of the person. This is shown in figure 1.1. Detection is the first step of any recognition system.

Figure 1.2: Face Detection



1.1 Design of a Face Recognition System

Any recognition system can be divided into two main parts:

1. Feature Extraction
2. Matching

Feature Extraction It is the process of obtaining the required information from data. The data, is the input. It may contain image or speech signals. And the information that is extracted would be something that represents the input data in the best form. For example, consider a speech recognition system. The input is a spoken sentence, i.e., a continuous speech signal. It is then broken down into the constituent words. Every word has a specific set of phonemes. A phoneme is the smallest element of sound which depends on the pronunciation of that word. Thus every word has a unique set of phonemes. These phonemes have to be represented mathematically.

Similarly, one of the methods used in facial recognition is dividing the face into its constituent parts- like eye brows, eyes, lips, nose, *etc*. These parts define a face and are unique to every person.

The areas of every part are counted. A set of areas, unique for every part is assigned to every image. Images of the same person are grouped together.

Matching: Once the mathematical representation of the features is obtained, it is stored in a database. Whenever an external input is taken, the closest match is searched in the database. The database may also be used to train a mathematical model. This model can be used to recognize images or faces that are not in the database.

Chapter 2: The Procedure Used

2.1 Overview

In this chapter the procedure used in this project will be discussed. The first step is Preprocessing, followed by feature extraction. Preprocessing makes feature extraction more accurate. These features are then fuzzified, i.e., they are assigned linguistic variables. These linguistic variables are saved in a database. A two-step search is then carried out to find the right match for an image from the database.

There are two procedures used- Procedure A and Procedure B. The major differences in both the procedures are data supervision and the features that are extracted. These differences will be elaborated and the steps followed in each position are given at the end of the chapter.

2.2 The BioID Database

Both the procedures used the same database- The BioID Database. This database consists of 1521 images of 23 different individuals. The number of images of every person are not constant. Images are taken in an uncontrolled environment. The distance between the subject and the camera is unequal. The lighting conditions vary. Subjects have worn different clothing and glasses in different images. Expressions vary. A subject might have closed eyes and mouth open in one image and eyes open and mouth closed in the other. Subjects might laugh in one and frown in the other. This database can be used as a testing database. But a model which gives good results on this database can be called a robust model.

All the images in the database are in the PGM format. The center of the eyes of every person are given. 20 landmark points are also manually marked for each image in the database. They are given as text files.

Feature Extraction As stated above, the image database used in this case- the BioID database has twenty points manually marked on the face, representing the center of eyes, end points of eyebrows *etc*. These points were used as landmark points and were used to find out the lengths of their respective parts. In Procedure B, the dlib library was used to identify more landmark points (68 in total) and find areas of each part with the help of those points.

Data Supervision: The BioID database has 1521 images in total of 23 different individuals. In Procedure B, each person is assigned an alphabet. An incoming image is matched to a particular person. For this reason, we have a profile for every person in Procedure B. On the contrary, in Procedure A, the image closest to the input image is matched. Whose image it is does not matter. Hence, Procedure B involves supervised matching and Procedure B has the opposite.

2.3 Procedure A

The first procedure is relatively less complex. There are lesser steps involved. They can be elaborated as followed:

1. Feature extraction, which involved measuring distances between points which was done using the coordinate geometry distance formula.

2. Processing, which helped cancel the *zooming effect* since all the images were not equidistant from the camera.

3. A 2 Step matching processing- involving Fuzzy Logic.

2.4 Procedure B

The second procedure can be elaborated as followed:

1. First all the images were supervised. Every image in the database was classified into 23 sets. Each set involved images of one of the 23 people pictured.

2. Preprocessing data, which first involved detecting faces, which makes sure that we get only one set of measurements from every image involved canceling off the *zooming effect*.

3. Feature extraction, which involved finding areas of the eyes, lips, eyebrows *etc*.

4. A search using Fuzzy Logic

5. A matching technique which includes calculating scores to find the best match.

Chapter 3: Feature Extraction from Images

Extracting the features correctly helps successful matching. The aim of Feature Extraction is to find a way to best represent the input data. A method that takes into account the similarities and differences between two images. These features can be used for finding the right match.

For example, for a face, we first break it down into its respective parts- like eyes, lips *etc.* Each part will be represented by its area. This ensures that even if the shape has been changed, the area of the lips will stay the same.

3.1 Feature Extraction for Procedure A

3.1.1 Markup Scheme

The markup scheme given with the BioID Database is given in table 3.1 and shown in figure 3.2

Table 3.1: Markup Scheme used in Procedure A

Point #	Description
0	right eye pupil
1	left eye pupil
2	right mouth corner
3	left mouth corner

4	outer end of right eye brow
5	inner end of right eye brow
6	inner end of left eye brow
7	outer end of left eye brow
8	right temple
9	outer corner of right eye
10	inner corner of right eye
11	inner corner of left eye
12	outer corner of left eye
13	left temple
14	tip of nose
15	right nostril
16	left nostril
17	center point on outer edge of upper lip
18	center point on outer edge of lower lip
19	tip of chin

Figure 3.1: Markup Scheme for Procedure A

3.1.2 Finding the Length of Every Face Part

The following equation was used to find out the distance in each case:

$$d = \sqrt{(y_2 - y_1)^2 + (x_2 - x_1)^2} \tag{3.1}$$

3.2 Feature Extraction for Procedure B

The biggest difference in Procedures A and B is that the former uses lengths and the latter uses areas of face parts as the feature extracted.

Table 3.2: Distances used in Procedure A

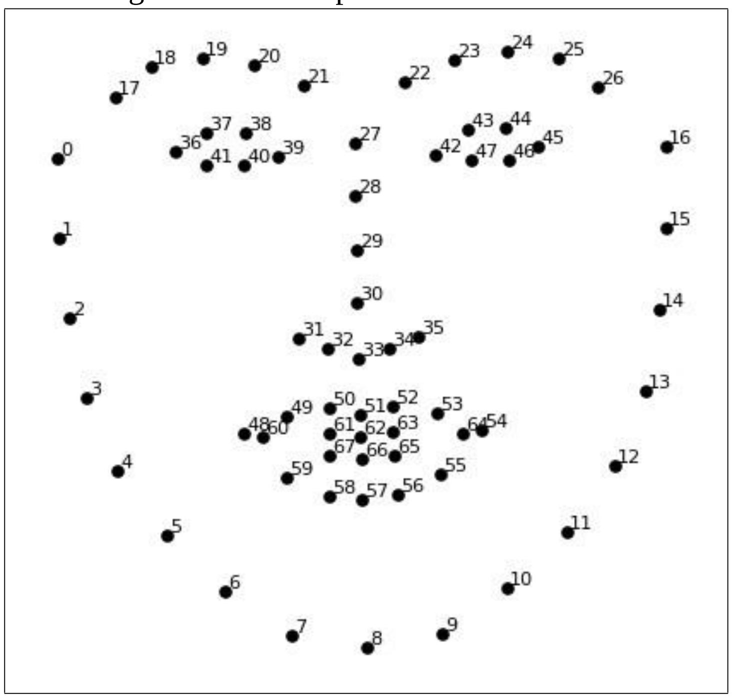
Face Part	Points
right eyebrow	4,5
left eyebrow	6,7
right eye	8,10
left eye	11,12
length of lip	2,3
height of lip	17,18
temple to temple distance	8,13
length of nose	14,17
height of nose	15,16

3.3 Markup Scheme used in Procedure B

Procedure B involves the use of 68 landmark points detected on the human face. It gives us form details than the 20 points used in Procedure 1. Here, all the points along the jaw, eyebrows and nose are labeled. The eyes and lips can be estimated as polygons since their edges are labeled. Thus, rather then just finding out lengths, we can get the curves describing eyebrows and jaws. In this way, their shapes are also considered. Areas of lips or eyebrows ensure that even though there is a difference in expressions the area will be same. For example, if a person is smiling or pouting, the area of the lips stays the same.

The markup scheme used is as shown in figure 3.2

Figure 3.2: Markup Scheme for Procedure B



3.4 Finding Measurements of Every Face Part

In this section, every face part whose area was calculated will be described in detail. The parts include:

1. eyes
2. nose
3. lips
4. eyebrows

5. jawline

Eyes

Points 36 to 41 and 42 to 47 represent the left and right eyes respectively as shown in figure 3.3.

Figure 3.3: Approximating Face Parts

The shoe lace formula is used to find the area of eyes. The shoe lace formula is shown in equation 3.2

$$\begin{aligned}\mathbf{A} &= \frac{1}{2}\Big|\sum_{i=1}^{n-1} x_i y_{i+1} + x_n y_1 - \sum_{i=1}^{n-1} x_{i+1} y_i - x_1 y_n\Big| \\ &= \frac{1}{2}|x_1 y_2 + x_2 y_3 + \cdots + x_{n-1} y_n + x_n y_1 - x_2 y_1 - x_3 y_2 - \cdots - x_n y_{n-1} - x_1 y_n|\end{aligned} \tag{3.2}$$

Lips

The area of the lips was calculated by subtracting the area of inner lip polygon from the outer lip polygon. The outer polygon is denoted by points 48 to 59. The inner polygon is denoted by points 60 to 68.

Area of lips = Area of outer polygon - Area of Inner Polygon

Area are found out with the help of the shoe lace Formula given in equation 3.2

Nose

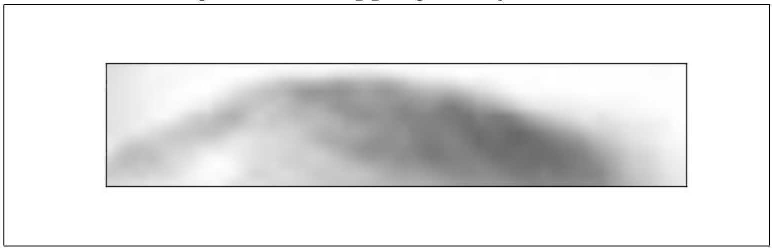
The approximation of the nose is shown in figure 3.3. The shoelace formula is used here to find the area too.

Eyebrows

A slightly different approach was used for eyebrows. eyebrows of all the subjects stood out, i.e., they were easily recognizable compared to their skin tones. The approach can be described as follows:

1. First the eyebrows were detected and landmark points were plotted. These are points 17 to 21 and 22 to 26.
2. These points were used to crop the rest of the image so that only the parts important to usethe eyebrows can be the only image under study

Figure 3.4: Clipping the eyebrows

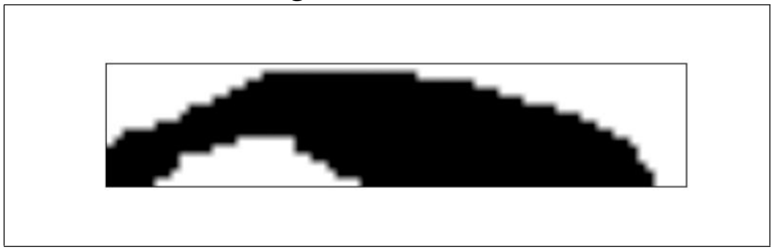


3. Adaptive Thresholding was carried out. The adaptive approach was selected since the pictures are taken in an uncontrolled environment. So having one threshold will not give best results.

Figure 3.5: Thresholding the eyebrows

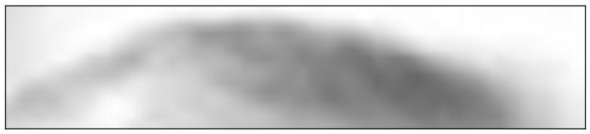
4. Erosion is then carried out. It makes the bright area brighter. The smaller noisy details are erased off.

Figure 3.6: Erosion



5. Now that we have the overall shape of the eyebrow, Dilation is carried out to enhance the shape.

Figure 3.7: Dilation

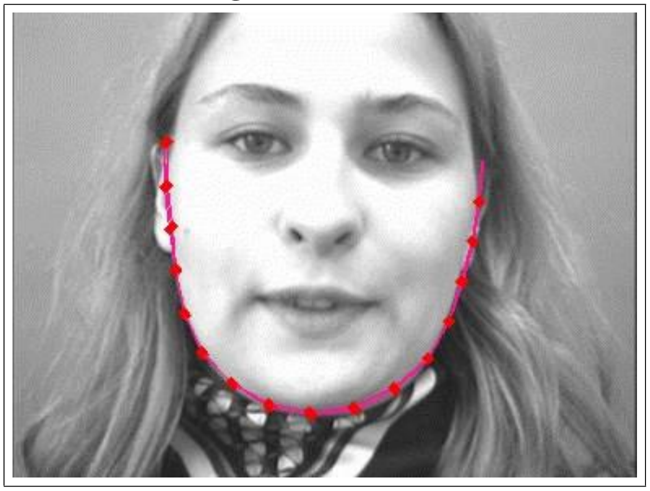


6. The area is calculated by taking the ratio of the number of black pixels to the total numberof pixels.

Jawline

Here, the words ‘jawline’ refer to the edge of the face. Points representing the jawline are points 0 to 16. These points can be considered a part of the curve. A curve is fit in those points. The coefficients of a polynomial of degree 4 representing the jawline are calculated and saved.

Figure 3.8: Jawline



3.5 Normalization and Scaling

The images are taken in an uncontrolled environment. The distance between the camera and the subject varies. Due to this measurements are non-uniform. To get rid of this ‘zooming’ effect, all the measurements are divided by their respective interpupilary distance. The interpupilary distance is the distance between the centers of the two eyes. These two points are also manually already marked on the BioID databases. Sample eye positions are shown in figure 9.9.

Figure 3.9: Eye Positions

Finally, all the normalized distances are scaled between 0 and 1 for uniformity.

Chapter 4: Fuzzy Logic and Matching

Fuzzy Logic is used to find out the face from the database that is most similar to the face of person in the input. Thus this concept can also be used to find the person's identity. Fuzzy Logic is used to classify measurements. The way Fuzzy Logic is used in Procedure A and B differs. But their use is the same.

Fuzzy Logic is used in its most common manner. It is used to classify every measurement (area or distance) to its respective linguistic variable. This is carried out with the help of membership functions.

4.1 Use of Fuzzy Logic in Procedure A

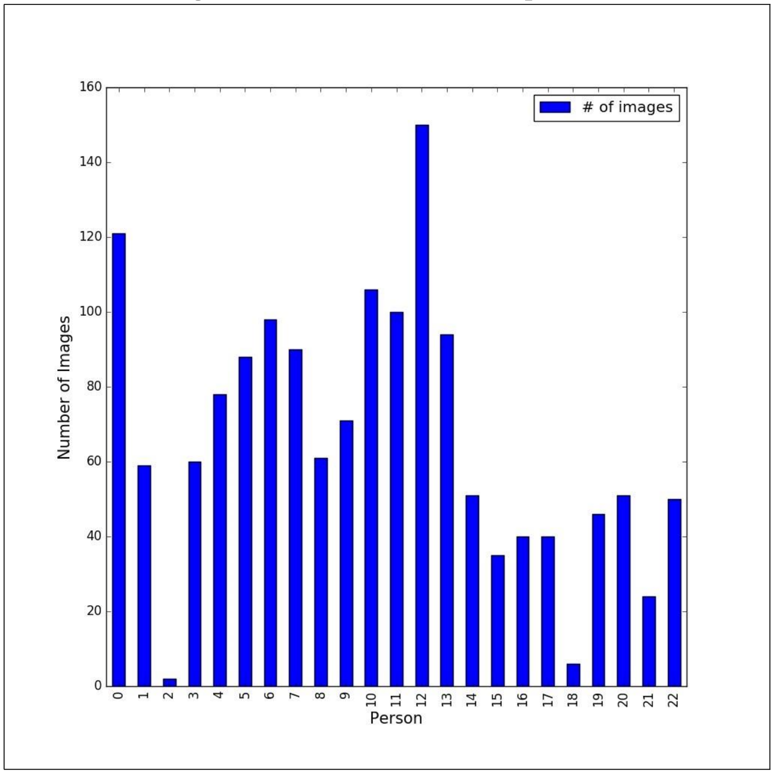
In Procedure A, the measurements of all face parts were classified by the same membership functions. Triangular membership functions were used. Gaussian memberships increased the execution time and using 5 instead of 3 linguistic variables did not make a big difference to the results.

4.2 Use of Fuzzy Logic in Procedure B

The subject in every image was given a name to identify that person with. All the names were assigned alphabetically for convenience. The number of images for each person were not the same.

A bar graph denoting their names and number of images is shown in figure 4.2

Figure 4.1: Pictures of each person



All the images belonging to one person were stored together.

Profile: The stored set of images of each person were further used to create a *profile* specific to every person. This profile has the name, mean and standard deviations of all the images in the group.

Profile Name: C

Images in the group: 2

Mean: 0.5693

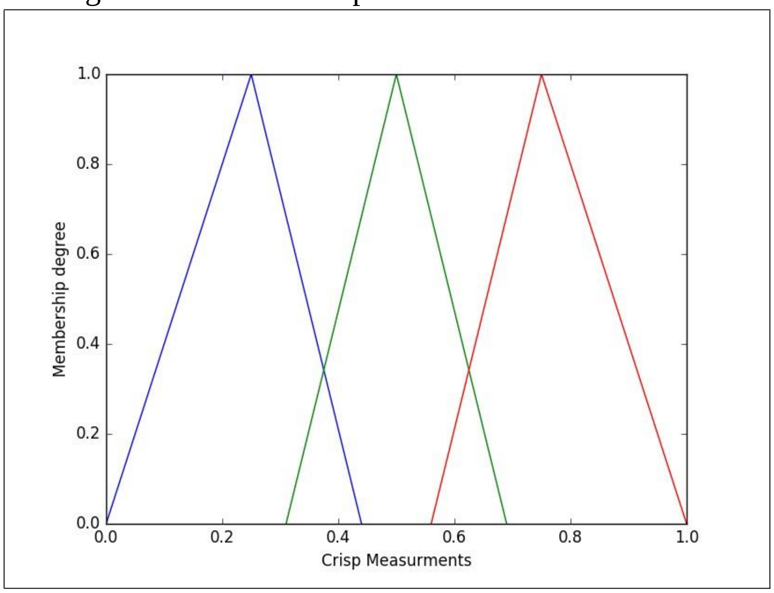
Std Dev: 0.0124

4.3 Membership Functions in Procedure B

All the membership functions were specific to the respective face part. These were made considering the mean and standard deviation of each part.

The center of the green (middle, medium) triangle is the mean of all the measurements of that part. Its spread is 0.5 times the standard deviation on both sides. The blue (left, small) triangle is centered at mean minus 0.33 times the standard deviation. The red (right, large) triangle is centered at mean plus 0.33 times the standard deviation.

Figure 4.2: Membership Fucntions for Procedure B



4.4 Matching

Matching is the final step in the process of image of recognition. The matching is a two-step process in Procedure A as well as B. The first step process is finding the right group. The second is finding the right match within the group. Finding the right group involves fuzzy variable and finding the right match within the group requires crisp values. This process will be elaborated in the further section.

4.4.1 Matching in Procedure A

In Procedure A, all the images were classified using the same membership functions. The first fifteen entries from the database used in Procedure is shown in table 4.1

Table 4.1: Part of the database used in Procedure A

	t2t	eyes	lips	Nose	filename
0	medium	large	large	Large	BioID 0000.jpeg
1	medium	medium	medium	Small	BioID 0001.jpeg
2	small	small	small	medium	BioID 0002.jpeg
3	large	small	small	Small	BioID 0003.jpeg
4	large	medium	medium	medium	BioID 0004.jpeg
5	large	medium	medium	medium	BioID 0005.jpeg
6	large	medium	medium	Small	BioID 0006.jpeg
7	medium	medium	medium	Small	BioID 0007.jpeg
8	medium	large	large	Medium	BioID 0008.jpeg

Table 4.2: Sample input in Procedure A

t2t	eyes	lips	nose	Filename
large	medium	medium	medium	BioID 0004.jpeg

Table 4.3: Step 1 of Matching in Procedure A

t2t	eyes	lips	nose	Filename
Large	medium	medium	medium	BioID 0004.jpeg
Large	medium	medium	medium	BioID 0005.jpeg
Large	medium	medium	medium	BioID 0046.jpeg
Large	medium	medium	medium	BioID 0049.jpeg
Large	medium	medium	medium	BioID 1091.jpeg
Large	medium	medium	medium	BioID 1094.jpeg
Large	medium	medium	medium	BioID 1441.jpeg

Step 1: Using Fuzzy Values

First all the required measurements of the incoming file are taken - eyebrows, eyes, lips, nose. Those measurements are fuzzified using the membership functions mentioned previously. Once the respective membership functions are obtained, they are grouped with all the images who share the same variables.

For example, if the input image has variables as shown in table 4.2

Once this is obtained, all the files with the same set of linguistic variables are selected. There are 46 images with the same linguistic variables. 10 of those 46 samples are shown in table

Table 4.4: Crisp Values in Procedure A

t2t	filename	eyes	lips	nose
0.534167	BioID 0004.jpeg	0.118479	0.144209	0.467858465585
0.634805	BioID 0005.jpeg	0.347561	0.351719	0.462618716544
0.579461	BioID 0046.jpeg	0.334246	0.338349	0.498423563828
0.584774	BioID 0049.jpeg	0.472350	0.461115	0.449699480523
0.575826	BioID 1091.jpeg	0.688343	0.679380	0.663234604647
0.587565	BioID 1094.jpeg	0.600732	0.593482	0.639033468034
0.613027	BioID 1441.jpeg	0.365937	0.359400	0.68862532815

4.4.2 Step 2: Using Crisp Values

Now the crisp values of those images were considered, as shown in table 4.4.

Finally, all these values are compared with the crisp values of the input image. The crisp values of the input image are subtracted with all the other crisp values. All the differences are added. The image with the smallest difference is displayed as the closest match.

4.4.3 Matching in Procedure B

In Procedure B, a similar approach is used. The biggest difference is that in this case, the input is matched with a profile, rather than being matched with another image. The database is shown in table 4.5. All the persons are numbered

A *group* is defined as a set of all the images which share the same set of linguistic variables

Table 4.5: Database in Procedure B

person	brows	eyes	lips	nose
0	0	0	0	0
1	1	0	0	0
2	2	0	0	1
3	1	2	2	1
4	1	2	2	2
5	2	2	2	0

6	1	2	2	0
7	1	0	0	2
8	2	0	0	0

Chapter 5: Results

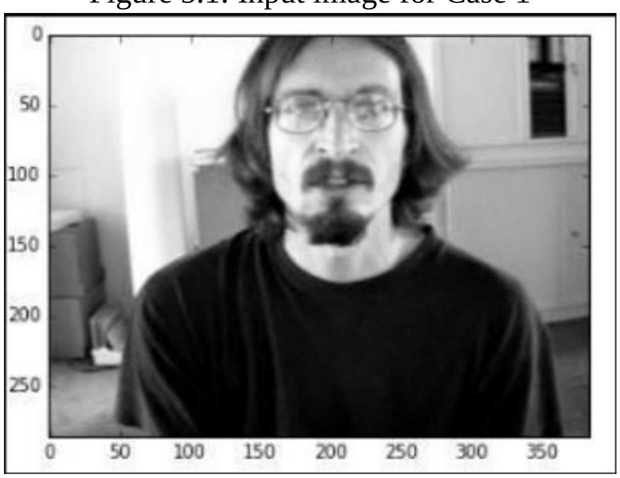
The results of Procedure A and B are explained in this chapter. A few cases of each Procedure are shown. There are two figures associated with every case. The first one is the input and the second one is a bunch of most similar results.

5.1 Results of Procedure A

5.1.1 Case 1

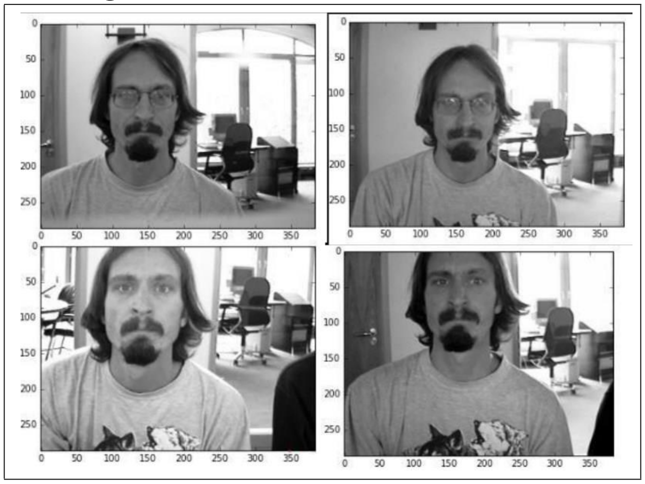
Here, the input image is selected at random from the database which has 1521 images. It should be noted that the subject is wearing glasses. Ideally, this should not affect the results. The input image is shown in figure 5.1.

Figure 5.1: Input image for Case 1



The most similar 4 images are shown as subplots in figure 5.2. All the four results correspond to the same person. The subject in the resultant image is wearing different clothes and not wearing glasses in all the images. This shows that the program works regardless of the subjects attire and image background.

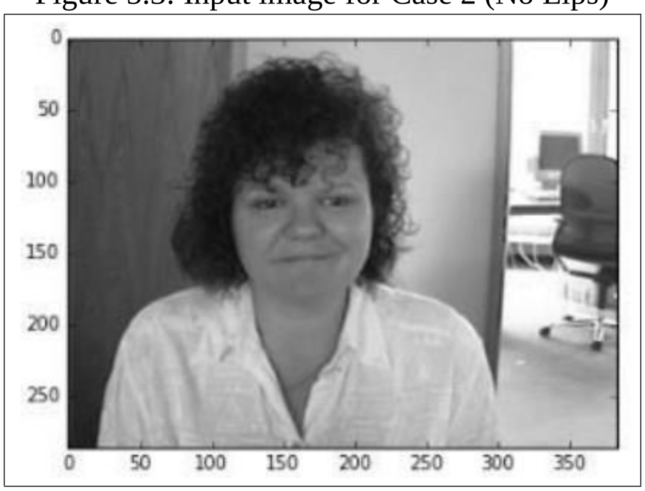
Figure 5.2: Accurate Result for Case 1



5.1.2 Case 2

The second case is similar to the input case except for the fact that the lips are not considered during classification. The parts which are considered are the eyes, eyebrows, nose and temple to temple distance. Another image is selected from random. It is shown in figure 5.3

Figure 5.3: Input image for Case 2 (No Lips)



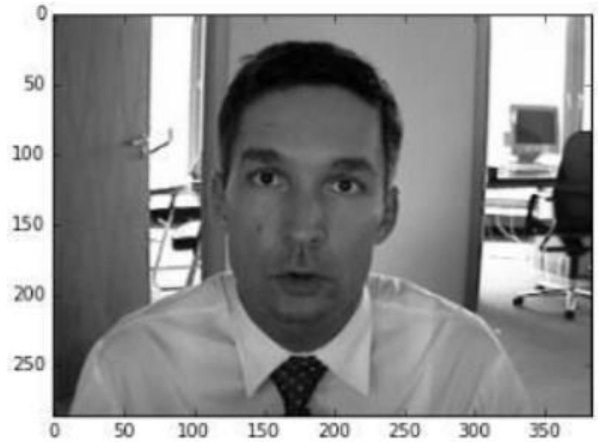
The results that are achieved are shown in figure 5.4 are incorrect. None of the four images are images of the subject in the input. But something that can be noted is that the eyes, eyebrows, nose and temple to temple distance are similar. The lips in all the images are different. This only goes to show that lips are an important feature in this recognition process.

5.1.3 Case 3

In this subsection, results of another case are shown. An input image is again selected at random. It is shown in figure 5.5. The features selected this time are eyes, eyebrows, nose and lips. The temple to temple distance - which can also be called the length of the forehead is not considered.

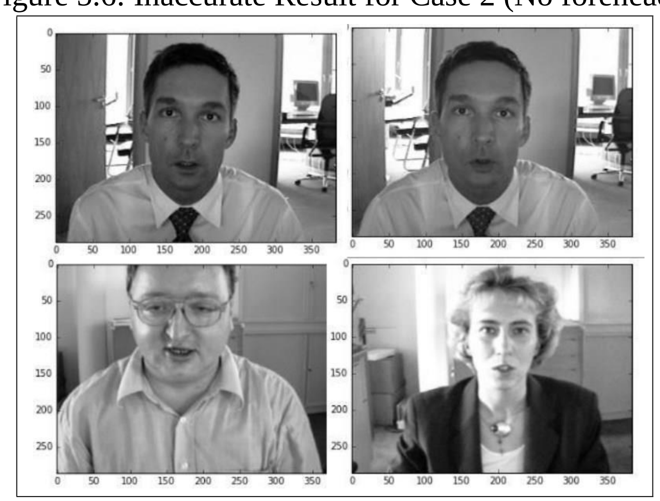
Figure 5.4: Inaccurate Result for Case 2 (No Lips)

Figure 5.5: Input image for Case 2 (No forehead)



The output is displayed as a subplot of the four closest matches as shown in figure 5.6. Here, the results are accurate. The two upper images are of the same subject. However the other two images are vastly different. This goes to show that the forehead is an important feature to consider during recognition but not as important as the lips.

Figure 5.6: Inaccurate Result for Case 2 (No forehead)



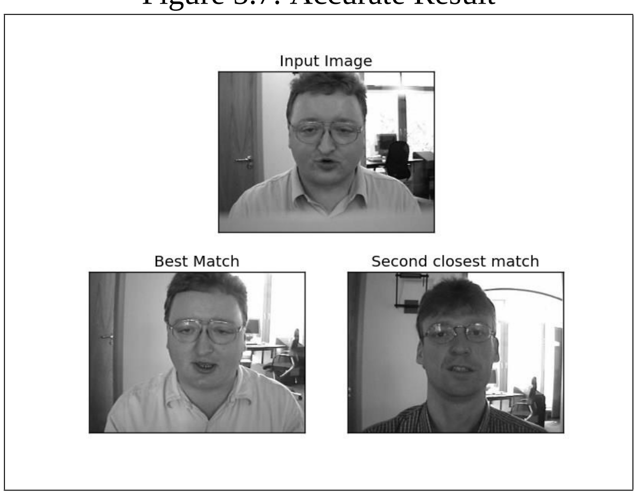
5.2 Results of Procedure B

During simulation, one profile is selected at random. An image of the person's profile is labelled as the "input image." It is the image in upper row in the output dialog box. The lower row has two images- the best match and the second closest match. The image labelled as "Best Match" should belong to the person whose picture is the input image. Two best matches which are displayed at the output also show how fuzzy logic is used to find out similarities in faces.

5.2.1 Case 1

Figure 5.7 has persons B and H. H is at the input. The best match made here is the same person H. This is a case of successful recognition, despite having different lip movements and changes in eye positions. Thus because of the use of fuzzy logic, the program was able to identify and capture these differences as similarities. The two similar faces at the output have similar sized eyes, nose and lips.

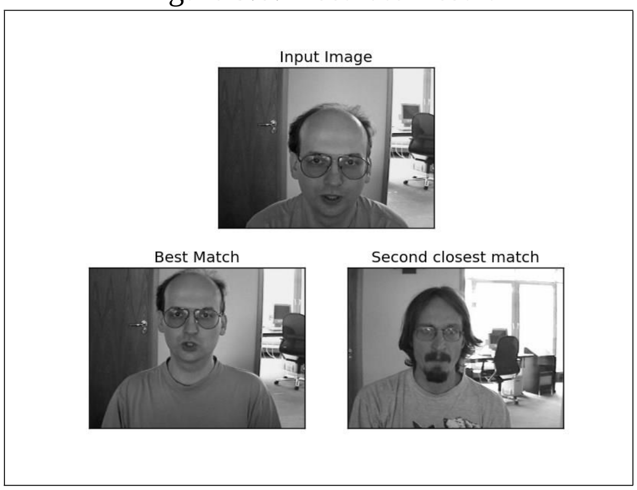
Figure 5.7: Accurate Result



5.2.2 Case 2

Figure 5.8 has person U as the input and U as the right match. The facial expressions are almost the same. Here the distance between the camera and subject changes. However despite that, due to data normalization and scaling, the program was able to correctly recognize the right face. These results also show that eyewear is not affecting results.

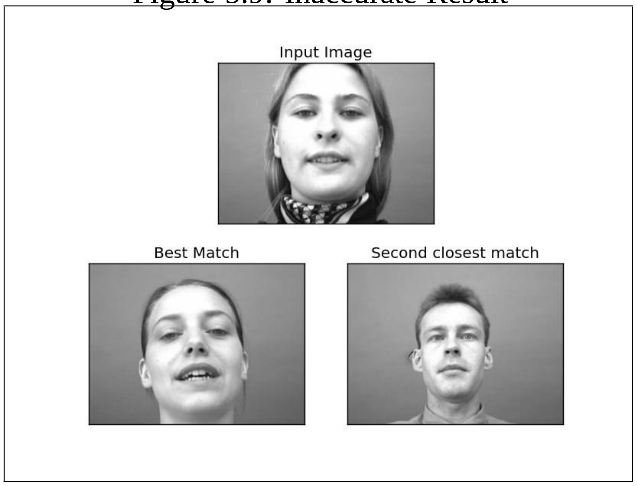
Figure 5.8: Accurate Result



5.2.3 Case 3

This first two results represent 70% of the cases where results are accurate. The next simulated result figure 5.9 is an example where a similar face is incorrectly interpreted as the best match. The program saw similar facial features without realizing that the person in the 'Best Match' image has their chin tilted towards the camera.

Figure 5.9: Inaccurate Result



Bibliography

[1] Leslie Lamport, *$\LaTeX$: a document preparation system*, Addison Wesley, Massachusetts, 2nd edition, 1994.

[2] Oliver Jesorsky, Klaus J. Kirchberg, and Robert W. Frischholz *Robust Face Detection Using the Hausdorff Distance*

[3] L. A. Zadeh *Fuzzy Sets*

[4] Javid Sadr, Izzat Jarudi, Pawan Sinha *The role of eyebrows in face recognition*

[5] Zhanpeng Zhang, Ping Luo, Chen Change Loy, and Xiaoou Tang *Facial Landmark Detection by Deep Multi-task Learning*

[6] Tadas Baltrusaitis Peter Robinson, Louis-Philippe Morency *Constrained Local Neural Fields for robust facial landmark detection in the wild*

[7] Peter N. Belhumeur, David W. Jacobs, David J. Kriegman, Neeraj Kumar *Localizing Parts of Faces Using a Consensus of Exemplars*

[8] Joshua C. Klontz, Anil K. Jain *A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects*

[9] Gregory Shakhnarovich, Baback Moghaddam *Face Recognition in Subspaces*

[10] Thomas Huang, Ziyou Xiong, Zhenqiu Zhang *Face Recognition Applications*

[11] Andrew W. Senior, Sharathchandra Pankanti *Privacy Protection and Face Recognition*

[12] Alice J. O'Toole *Face Recognition by Humans and Machines*