Team Members: Hunter Glass, James Hanselman, Alice Kuang, Mackenzie Meier, Evan Powell

Professor Sumaiya Shomaji

EECS 700

7 December 2023

Project: Analyzing and Improving a Basic Face Biometric System

**Part I: (25 points)**

You build the basic face biometric system in this part using the given data “ProbeSet.rar” and “GallerySet.rar” from HW3. Complete the following:

(a) Consider the “normalized correlation coefficient” function (see Appendix) as your similarity measure and perform the comparison between each probe template with all the gallery templates to generate a score matrix. Your score matrix A will be a 200x100 matrix where A[i,j] denotes the match score generated by comparing the i-th probe with the j-th gallery data. Provide a snippet of the A[0:9,0:9] matrix. (5 points)

(b) Plot the genuine and impostor score distributions in a reasonably comparable window. (5 points)

(c) Report the decidability index value. (5 points)

(d) Plot the Receiver Operating Curve (FAR vs. FRR). (5 points)

(e) What is the EER? (5 points)

**Part II.**

Part I represents a very basic face biometric system which considers the raw pixel intensity values as features as well as a normalized correlation coefficient as comparator. As mentioned in the Objective, in Part II you will propose and implement a method for improving/attempting-to-improve the performance of the above basic face biometric system. [Hint: You may consider different feature representations, pre-processing techniques, matching algorithms, etc. or combination of them.]

Now answer the following:

1. Briefly introduce your proposed method. Clearly define the differences between proposed vs. basic face biometric system. Also, discuss the justification of your choice of the proposed method. *(2.5 + 2.5 + 5 = 10 points)*
2. Generate a new score matrix (*A\_new*) for your proposed solution. Your new score matrix *A\_new* will also be a 200x100 matrix where *A\_new[i,j]* denotes the match score generated by comparing the *i*-th probe with the *j*-th gallery data. Provide a snippet of *A\_new[0:9,0:9]. (5 points)*
3. Plot the genuine and impostor score distributions in a reasonably comparable window. Discuss if the distribution is better than that of the basic system. *(5 + 5 = 10 points)*
4. Report the decidability index value. Discuss if the separation is better than that of the basic system. *(5 + 5 = 10 points)*
5. Plot the Receiver Operating Curve (FAR vs. FRR). Discuss if the verification performance is better than that of the basic system. *(5+ 5 = 10 points)*
6. What is the EER of the proposed system? Discuss if the verification performance is better than that of the basic system. *(5+ 5 = 10 points)*
7. The goal of Part II is to revise the basic system in such a way so that it can provide better recognition performance. Quantify the amount of improvement by the metric called “*Improvement Factor (IF)*” considering the given formula. *(20 points)*

*IF = round (delta\_ d’, 2)* where *delta\_d’ = d’ of your proposed system – d’ of basic system*

A questionnaire with black text

Description automatically generated

\*

**General Comment:** For this project, you will receive credits for how much improvement you could earn (if any), thoroughness of your work, justification of implementations, and your analytical ability while analyzing as well as comparing the performance of your proposed system vs. the basic system.

**\***

**Part III (Code).**

# PART I **==========================================================================**

import patoolib

import cv2

import math

import numpy as np

from numpy import \*

import matplotlib.pyplot as plt

import statistics

from shapely.geometry import LineString

import seaborn as sns

sns.set\_theme()

patoolib.extract\_archive("GallerySet.rar", outdir="GallerySet")

patoolib.extract\_archive("ProbeSet.rar", outdir="ProbeSet")

# Given formula from appendix; use as similarity measure.

def normalizedCorrelationCoefficient(img1: np.ndarray, img2: np.ndarray):

x = img1.reshape((-1,1))

y = img2.reshape((-1,1))

xn = x - np.mean(x)

yn = y - np.mean(y)

r = (np.sum(xn \* yn)) / (np.sqrt(np.sum(xn\*\*2)) \* np.sqrt(np.sum(yn\*\*2)))

return r

score\_matrix = np.zeros([200,100])

# record scores to correlation matrix

pi = 2

for i in range(0, 200):

if i%2==0:

pi = 2

else:

pi = 3 # probe idx

img1\_path = "ProbeSet/subject" + str(math.ceil((i+1) / 2)) + "\_img" + str(pi) + ".pgm"

img1 = cv2.imread(img1\_path, cv2.IMREAD\_GRAYSCALE)

for j in range(0, 100):

img2\_path = "GallerySet/subject" + str(j+1) + "\_img1.pgm"

img2 = cv2.imread(img2\_path, cv2.IMREAD\_GRAYSCALE)

score\_matrix[i][j] = normalizedCorrelationCoefficient(img1, img2)

print('10x10 matrix')

for i in range(10):

print('[', end = '')

for j in range(10):

print(score\_matrix[i][j], end = ' ')

print(']')

# filter by idx i = j to get arrays containing two distributions.

def separate\_distributions(arr, genuine, imposter):

for i in range(200):

for j in range(100):

if math.floor(i/2) == j:

genuine.append(arr[i][j])

else:

imposter.append(arr[i][j])

# plot w/ Seaborn distplot

def plotNormalizedDist(genuine, imposter, title):

sns.histplot(data=genuine, bins=35, kde=True, stat='density')

sns.histplot(data=imposter, bins=25, kde=True, stat='density')

plt.xlabel('Threshold')

plt.ylabel('Probability Density')

plt.title(title + ' Normalized Distribution')

# initialize two arrays to hold genuine and imposter scores

genuine=[]

imposter=[]

separate\_distributions(score\_matrix, genuine, imposter)

plotNormalizedDist(genuine, imposter, 'Face Matching Score Distribution')

#(c) Report the decidability index value. (10 points)

def d\_value\_calc(genuine, imposter):

m1 = statistics.mean(genuine)

s1 = statistics.stdev(genuine)

m0 = statistics.mean(imposter)

s0 = statistics.stdev(imposter)

return math.sqrt(2) \* abs(m1 - m0) / math.sqrt(s1\*s1 + s0\*s0)

print('Decidability value: ', d\_value\_calc(genuine, imposter))

#(d) Plot the Receiver Operating Curve (FAR vs. FRR).

# FAR

dimensions = np.shape(imposter)

imp\_rows = dimensions[0]

FAR=np.zeros([11])

for i in range(imp\_rows):

score = imposter[i]

for j in range(11):

if score >= j/10: # threshold

FAR[j] +=1

for i in range(11):

FAR[i] = 100 \* FAR[i] / imp\_rows

# FRR

dimensions = np.shape(genuine)

gen\_rows = dimensions[0]

FRR=np.zeros(11)

for i in range(gen\_rows):

score = genuine[i]

for j in range(11):

if score < j/10: # threshold

FRR[j] +=1

for i in range(11):

FRR[i] = 100 \* FRR[i] / gen\_rows

# ROC curve

plt.plot(FAR, FRR)

plt.title("ROC Curve")

plt.xlabel("FAR (%)")

plt.ylabel("FRR (%)")

plt.xlim(0, 100)

plt.ylim(0, 100)

plt.rcParams["figure.autolayout"] = True

#(e) Threshold for EER and value of EER at threshold

threshold = [0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]

plt.plot(threshold, FAR, 'r')

plt.plot(threshold, FRR, 'b')

plt.title("Finding Threshold from FAR and FRR curves")

plt.xlabel("Threshold")

plt.ylabel("Identification Rate (%)")

plt.legend(["FAR", "FRR"], loc=0, frameon=True)

#get intersection

line1 = LineString(np.column\_stack((threshold, FAR)))

line2 = LineString(np.column\_stack((threshold, FRR)))

intersection = line1.intersection(line2)

x, y = intersection.xy

plt.plot(\*intersection.xy, 'o')

plt.show()

print("Threshold value: ", x[0])

print("Equal Error Rate: ", y[0])

# ===============================================================================