Forehead Wrinkles Image Matching for Authentication (Result Analysis)

Manav Mehta 2021AAPS0636P

1 Introduction

In this report, the results of image matching for forehead wrinkles is analysed. Authentication using facial features becomes a task with the ongoing pandemic which requires us to wear a mask. The technique used in this experiment does not require expensive scans, rather makes use of images captured from a standard hand held mobile phone.

2 Database

The data was collected by the students enrolled in the course CS-F425 Deep Learning (BITS Pilani). Data was collected in 2 sessions with a minimum of 12 hour gap between the sessions. Each session captured 10 pictures, 5 close up and 5 far off. From the total data collected, 336 images were used for the experiment. For 28 subjects, there were 3 close up images and 3 far off images per session.

3 System Pipeline

The general recognition pipeline consists of RoI (Region of Interest) extraction, image enhancement, feature extraction and image matching. Various feature extraction algorithms are used and the score obtained from them are used to plot the genuine/imposter histogram. A ROC (receiver operating characteristic curve) is plotted at varying classification thresholds. At the corresponding threshold, the EER (equal error rate) and CRR (correct recognition rate) is calculated.

3.1 Image pre-processing (RoI extraction, Image enhancement)

The region of interest was extracted manually by letting the the data provider align their head such a way that only the forehead wrinkles were in the image. The images were converted to grayscale and Contrast Limited Adaptive Histogram Equalization (CLAHE) was applied. CLAHE is better than the general histogram equilization. CLAHE is a variant of Adaptive histogram equalization (AHE) which takes care of over-amplification of the contrast. CLAHE operates on small regions in the image, called tiles, rather than the entire image. The neighboring tiles are then combined using bilinear interpolation to remove the artificial boundaries.

(a) Before

(b) After

Figure 1: Converting to grayscale and applying CLAHE

3.2 Feature Extraction and Score Calculation

The feature extractors used in this experiment are SIFT, SURF, ORB, AKAZE, BRISK, ArcFace and SSIM. The first four algorithms are used to get the local image features for the input image. To perform image comparison, the extracted feature vectors are compared for the given two images. These matches are computed using the Brute Force algorithm which compares the feature descriptors of the two images under consideration.

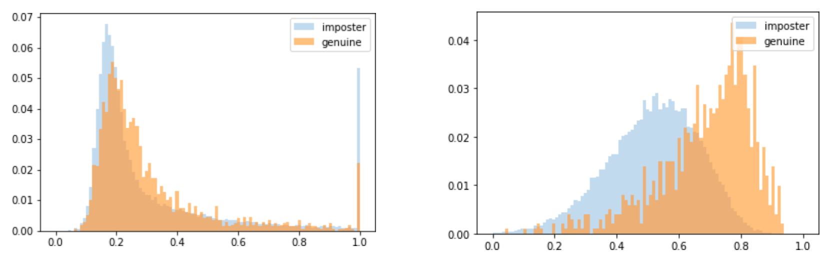
Lowe’s ratio test is performed to remove the ambiguous matches. Each keypoint of the first image is matched with a number of keypoints from the second image. The 2 best matches for each keypoint (best matches is the ones with the smallest distance measurement) are kept. Lowe’s test checks that the two distances are sufficiently different. If they are not, then the keypoint is eliminated and will not be used for further calculations.

ArcFace is a neural network classification model. It uses a similarity learning mechanism that allows distance metric learning to be solved in the classification task by introducing Angular Margin Loss to replace Softmax Loss. The distance between faces is calculated using cosine distance, which is calculated by the inner product of two normalized vectors.

SSIM stands for Structural Similarity Index. It is a metric that quantifies the structural difference between two images. This metric depends on the image data and not the features present in the image.

3.3 Genuine/Imposter Histogram

A match is said to be genuine if the subject of the two images compared are the same. The images compared are across different sessions. If the two subjects of the compared images are different, then it is an imposter match. For an authentication system,



(a) SIFT (b) ArcFace

Figure 2: Genuine/Imposter histogram

fewer number of imposter matches are ideal. The genuine/imposter histogram is plotted for score vs probability of occurrence of that score. The frequency is not plotted since the number of genuine matches are far fewer compared to imposter matches in $n \times n$ matching. For the algorithms that gave a matching score greater than 1, it was clipped. The frequency of occurrence of those large scores were less.

3.4 Calculation FAR and FRR

False Acceptance Rate (FAR) refers to the rate of false acceptance over the number of imposter attempts. False Rejection Rate (FRR) refers to the rate of false rejection over the number of genuine attempts by the user.

$$\text{FAR} = \frac{FP}{FP + TN} = \frac{\text{No. of incorrect matches recognized}}{\text{Total no. of imposter attempts}} \quad (1)$$

$$\text{FRR} = \frac{FN}{FN + TP} = \frac{\text{No. of correct matches recognized}}{\text{Total no. of genuine attempts}} \quad (2)$$

where FP refers to false positive, FN refers to false negative, TP refers to true positive, TN refers to true negative.

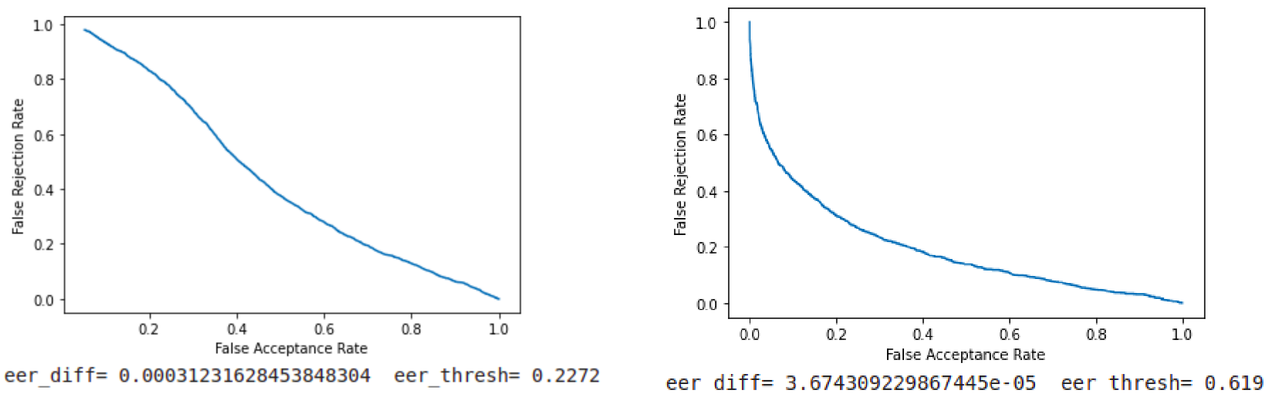
3.5 Plotting ROC and calculation EER

The ROC curve is plotted by taking the FRR on y-axis and FAR on x-axis.

The point where the difference between the FAR and FRR is minimum is EER. Theoretically, both FAR and FRR should be equal at EER. EER is calculated while varying the threshold.

CRR is the number of actual matches that are obtained at rank one recognition.

$$\text{CRR} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{\text{No. of matches correctly recognized}}{\text{Total no. of matches}} \quad (3)$$



(a) SIFT (b) ArcFace

Figure 3: ROC curve

4 Results

After performing the image preprocessing and enhancement and running various feature extracting algorithms, the results of the experiments are tabulated in Table 1.

Feature Extractor/ Image Matching Algorithms	EER	CRR
SIFT	0.2272	0.553766298185941
SURF	0.3572	0.5374858276643991
ArcFace	0.619	0.7401147959183674
ORB	0.2395	0.5282193670150988
AKAZE	0.2501	0.5486291486291486
BRISK	0.2069	0.5163198913198913
SSIM	0.1588	0.6439023526077098

Table 1: Summary of results of feature extractors used

ArcFace performs the best with the given data. The results are good since it used pre-trained weights and parameters to extract facial features from the image. The similarity scores calculated were 1-(cosine distance between two images). SSIM also performed relatively well due to the similarity in position and lighting of the images comapared. Since SSIM does not extract the features, rather compares the image data the way it is, it is not reliable with varying camera angles, lighting, contrast, etc. Since we performed appropriate image enhancements, the lighting and contrast of the compared images were similar. The reason for the large number of 0s in the histograms of ORB, AKAZE and BRISK is due to the lack of features extracted by the feature extractor. This is due to the presence of blurry images in the dataset. The Contrast Limited Adaptive Histogram Equalization gave marginally better results for all the feature extractors than regular equilization. The ROC curves of the extractors used in the experiment are plotted in Figure 4 for comparison.

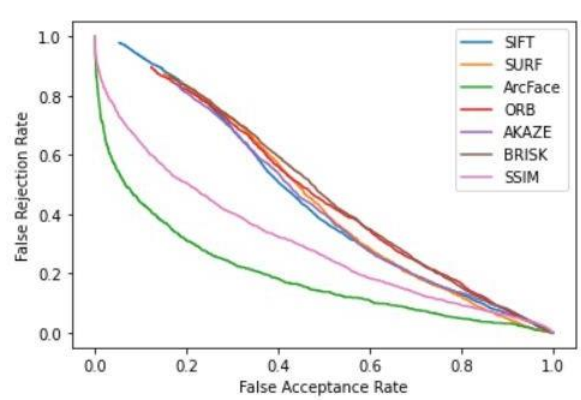


Figure 4: ROC curves for feature extractors used