# Iris Recognition on Big Data Report

W4121 extra credit project

Name: Xiaohui Guo

UNI: xg2225

## Abstract:

Iris recognition technology is popular nowadays due to iris's uniqueness and the increasing requirement of personal security. Some cell phone companies already implemented the iris recognition technique in the product, such as Samsung Galaxy S8. More and more iris information will be collected for security, personalization and other purposes. The iris images collected are not always with high quality, such as when camera lens is dirty or the person moves his or her eyes when taking the iris pictures, all of these cases will produce low quality iris images with noise. This project studied how the iris recognition algorithms works when implemented on large scale data with different noise density.

## 1. Introduction

The current machine learning algorithms, such as linear discriminative analysis and PCA, can work well on iris recognition when the training and testing images' quality are very high. This project studies how the algorithms perform when the iris's image quality are not that good, also this project studied how the algorithm works on scaled dataset.

In related work that has been done, such as in the paper "Personal Identification Based on Iris Texture Analysis" written Li Ma, Tieniu Tan and others in 2003, they implemented the linear discriminative analysis method on 213 irises' images to recognize irises. The size of the database is small and the quality of the images are very high. They didn't study whether this algorithm still works well on low quality images.

In my project, I studied how the algorithms work on scaled dataset and tested whether it still works well with low quality iris images. The project is relevant to our course since it is about project on scaled data and also Spark machine learning is used for the iris matching step.

Due to the size of current available iris database is small, which is 108 eyes and 756 images of iris (take 7 pictures of each eye), I will scale up the dataset by multiply it with 128 to make the dataset 128 times bigger. I will also add noise of different density on the scaled dataset. The noise density is from 0 to 0.02 with the increment of 0.001. Then compare the classification result among the data set with 27 different noise densities.

## 2. Hypothesis

The more noise in the images, the more important iris pattern information will be lost. Thus the hypothesis is that the algorithm will perform worse and worse linearly as the noise density increases.

## 3. Audience and Needs

Currently iris recognition techniques are already used in products in market, such as Samsung Galaxy S8. Similar recognition techniques are widely used in the market, such as the face recognition in the newest Iphone. The problem is how secure the iris and face recognition is. In the IphoneX launch conference, it is really awkward to see that IphoneX failed to recognize the face. The industry need to test how well the algorithms works in iris, face recognition or other recognitions, such as finger or voice recognition. Among all the personal identity techniques, iris recognition require higher quality images since it is very easily to get blur iris images due to subject's movement or the dust on camera lens. Also when a subject wears dirty glasses or when the subject's palpebral covers a significant part of the iris, all of these cases can make the iris recognition fail. The industry need to know how the noise density influence the performance of iris recognition and come up with techniques to make sure iris recognition works well even in extreme bad situations.

The audiences for this project could be the phone company, safety inspection bureau, police station or other companies or organizations that need to use iris recognition. They would know how the noise density in the iris image influence the performance of iris recognition reading this paper.

# 4. Approach

**Dataset:**

756 iris images of 108 subject. Each iris has 7 images, 3 for training and 4 for testing. After scale the dataset, there would be data of 96768 images, which mean scale up to be 128 times bigger.

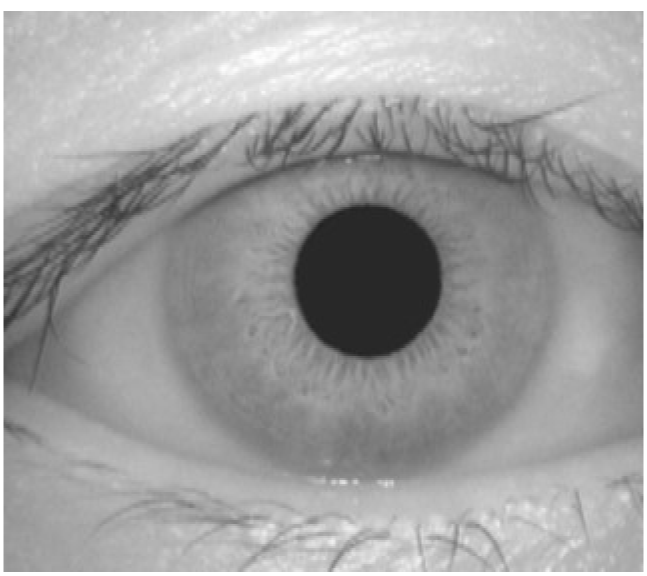


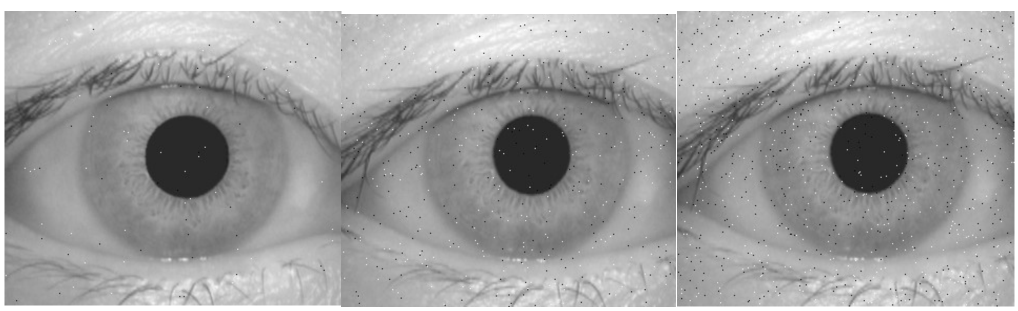
Figure 1: One of the images in the database.

**Steps:**

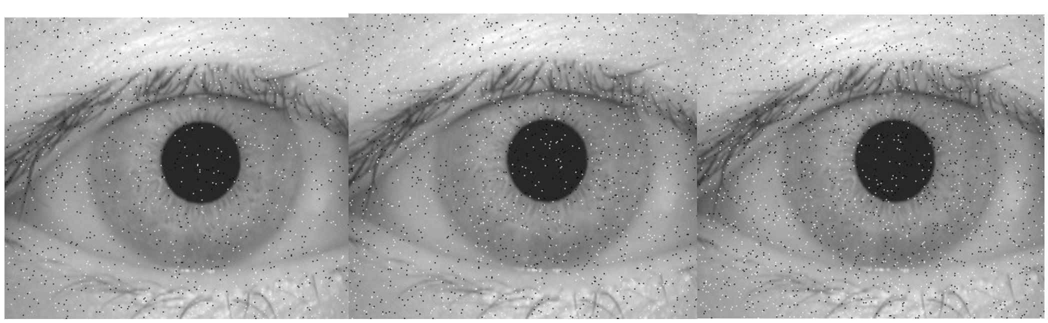
- Add Noise with Different Noise Density.

  In the original images, add salt and pepper noise with different noise density. The noise density I used in this project is 0 (no noise), 0.001, 0.002, 0.003, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.011, 0.012, 0.013, 0.014, 0.015, 0.016, 0.017, 0.018, 0.019, 0.02, 0.021, 0.022, 0.023, 0.024, 0.025, 0.026. This step is used to mimicking the situations that the images are of low quality due to dust on the camera lens, subjects wears dirty glasses or other reasons that make the image quality lower. In later steps, we will compare the prediction for iris recognition on these 27 datasets.

Figure 2: Add salt and pepper noise with different noise density:



Noise density: 0.001 Noise Density: 0.005 Noise Density: 0.010



Noise density: 0.015 Noise density: 0.020 Noise density: 0.025

- Image Processing

## A) Iris Localization

Implemented edge detection technique to find the pupil center and pupil radius. Remove the pupil and do edge detection again to find the center of iris and radius or iris.

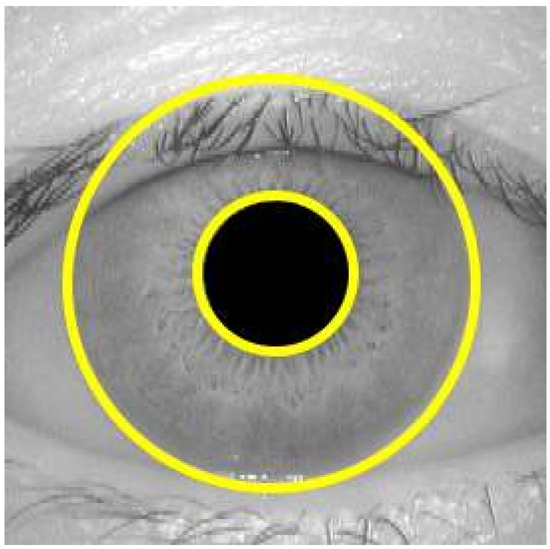


Figure 3: localize the iris

## B) Iris Normalization

Due to the different size of irises or different size of the iris images of the same iris because of the illumination variations, it is necessary to normalize the iris to achieve more accurate recognition results. Project the original iris in a cartesian coordinate system into a pseudopolar coordinate system. Then irises of different sizes are normalized to the same size. Use the Dugman's normalization method here.

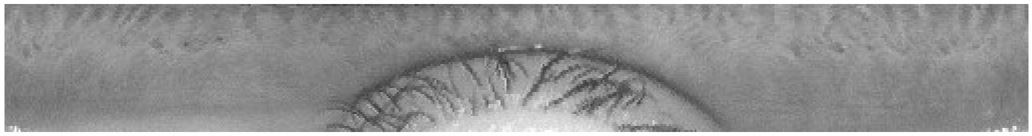


Figure 4: Normalize the iris.

C) Image Enhancement

Due to the low contrast of the normalized iris images, or the nonuniform brightness because of the position of light sources, image enhancement is needed to obtain a better distrusted texture image. Calculate the mean of each 16 y 16 small block to estimate the background illumination. Use the bicubic interpolation to expand the background illumination to the same size of the original image. Subtracted the expanded background illumination from the normalized images. Then enhance the lighting corrected image using the means of histogram equalization in small regions in the image. Then the iris texture characteristics are enhanced.

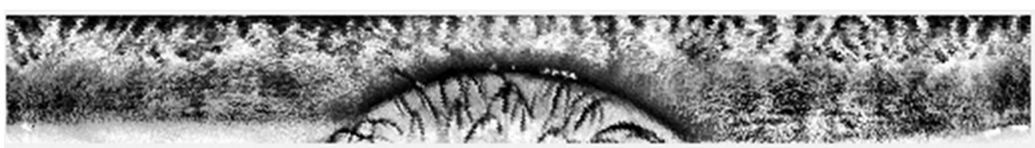


Figure 6: Image Enhancement.

D) Feature Extraction

Extract information of image using Gabor filters. Calculate the mean and variance of each 8 by 8 small block. Append all the mean and variance of each 8 by 8 small block to be the vector used for the classification part.

- ## Scale the dataset

Scale the dataset to be 128 times bigger. This step is to mimic the big data dataset. Since there is no iris database bigger enough online for this study, I scaled it manually. If there is a much bigger real world iris database, it would be definitely better for this study.

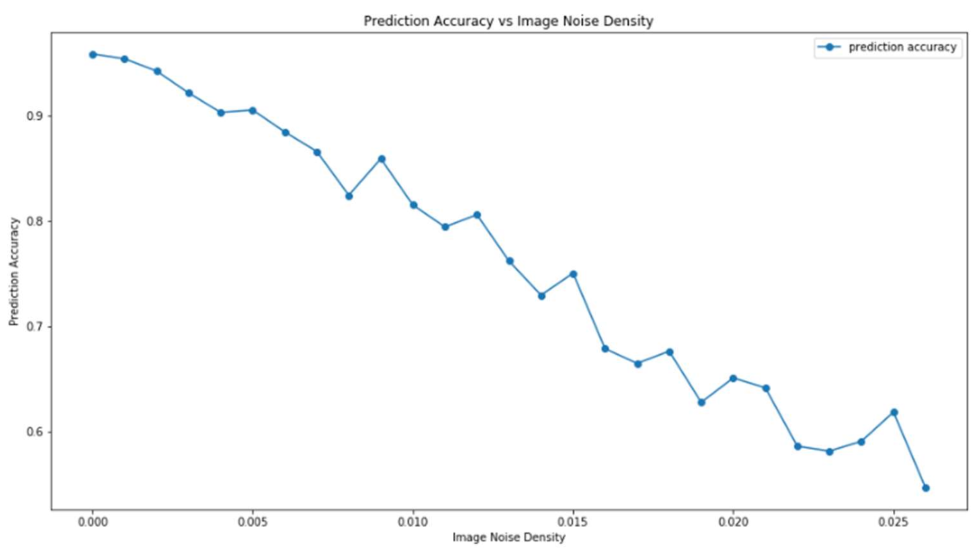
- ## Machine learning classification

  Since there are 1536 features in the dataset, first implemented feature reduction techniques, such as linear discriminative analysis to reduce the feature dimensions. Train the LDA model on the reduced dimension training dataset and predict the reduced dimension test dataset.

- ## Calculate the prediction accuracy

  Calculate the percentage of the correct prediction.

# 5. Best Case Impact




From the prediction accuracy vs noise density plot we can see as the image noise density increase, the prediction accuracy decreases linearly. When there is no noise, the prediction accuracy is the highest, which is 0.958.

The following are the noise density and its corresponding prediction accuracy:

| Index | Noise Density | Prediction Accuracy |
|---|---|---|
| 1 | 0 | 0.9583 |
| 2 | 0.001 | 0.9537 |
| 3 | 0.002 | 0.9421 |
| 4 | 0.003 | 0.9212 |
| 5 | 0.004 | 0.9028 |
| 6 | 0.005 | 0.9051 |
| 7 | 0.006 | 0.8843 |
| 8 | 0.007 | 08657 |
| 9 | 0.008 | 0.8241 |
| 10 | 0.009 | 0.8588 |
| 11 | 0.010 | 0.8148 |
| 12 | 0.011 | 0.7940 |
| 13 | 0.012 | 0.8056 |
| 14 | 0.013 | 0.7616 |
| 15 | 0.014 | 0.7292 |
| 16 | 0.015 | 0.75 |
| 17 | 0.016 | 0.6782 |
| 18 | 0.017 | 0.6644 |
| 19 | 0.018 | 0.6759 |
| 20 | 0.019 | 0.6273 |
| 21 | 0.020 | 0.6505 |
| 22 | 0.021 | 0.6412 |
| 23 | 0.022 | 0.5856 |
| 24 | 0.023 | 0.5810 |
| 25 | 0.024 | 0.5903 |
| 26 | 0.025 | 0.6181 |

| | | |
|---|---|---|
| 27 | 0.026 | 0.5463 |

## 6. Milestones

- Find iris database from Internet.
- Add salt and pepper noise with different noise density on the images. This step could be removed if I can find larger iris database which contains low quality iris images online.
- Localize the iris
- Normalize the iris
- Enhance the iris image
- Extract features from the enhanced iris image to create dataset for machine learning part
- Choose first three images per iris to be the training dataset. Choose the rest four images for each iris to be the training dataset.
- Scale the dataset to be 128 times bigger. This step will be removed if I could find bigger iris database from the Internet.

## 7. Obstacles

Major Obstacles:

The original dataset is small in terms of a big data project. There is not big iris database available online yet. I found one lager database but still not big enough for the desired data size. In addition, I didn’t get permission to download the dataset from that website. So need to manually scale the dataset for the project.

Minor obstacles:

After scaling the dataset, it is around 100GB, need to make sure my laptop has that much storage space. I didn’t notice the data could be that big so my laptop works pretty slow and died frequently. After I found that there is only 3.5 G storage left for my laptop, I cleaned my laptop and problem solved. Better to do the experiment on the google cloud VM.

## 8. Additional Resources

I need a iris database with more than 80000 iris images. There should a portion of the images are low quality images, such as blur images or images with dust on it.

I also need a VM, such as google cloud or other platform to do the experiment. The data set is very big, it is risky to work on locally in personal laptop. Laptop could work super slow if there is not enough storage space left.

## 9. Literature Review

“Personal Identification Based on Iris texture Analysis” by Li Ma, Tieniu Tan and more.

This paper discussed the method of preprocessing the image and it used an iris database of 213 subjects. It didn’t work on low quality images but on pretty high quality images.

## 10. Define Success

Successfully completed all the steps for the experiment and get the results of the experiment. The hypothesis is correct. But in terms of publishing the paper, I had better find a real large iris dataset and do the experiment again. Even though I found that the noise density is negatively linearly related with the classification prediction accuracy, but I manually scaled the dataset. It would be more convincing if I could implement the experiment on a real big iris data.

## Reference


Personal Identification Based on Iris texture Analysis by LiMa 2003