Is he Chinese, Korean or Japanese?

Yu Yan
Stanford University
450 Serra Mall, Stanford, CA 94305
yuyan1@stanford.edu

Abstract

This is an attempt to build a classifier which can distinguish between Chinese, Korean and Japanese faces. 2 new dataset was constructed for this task and different variations of convolutional neural networks were applied on the dataset. SVM and KNN classifiers were trained for comparison.

I. Introduction

Studies have shown that race is subconsciously one of the first information retrieved by a human brain upon perceiving a face[11]. Yet, a simple search online will tell you that research in face classification by race lags behind research in face classification by age or gender. In other words, computers today are comparatively "color-blind" when their programmers are the exact opposite of that. This project will tackle one of the most difficult(and perhaps unsolvable) problems in race classification. It is in a subset of race classification called ethnic classification. It is colloquially known as the "Is he Chinese, Korean or Japanese?"(shorthand CKJ) problem.

A. Problem Scope

Below is a rewording of a hand-wavy definition(blame [6]) of race for computer vision research:

Definition 1. Race in Computer Vision[6]

A set of humans is called a race if all of the following are satisfied:

1) Its members have a common place of origin. (this depends on how far back you want to look. By most theories everyone came from the same place at some point, but I'll just assume that we're reasonable here)
2) There exist a feature representation of the facial images of its members such that they form a cluster by themselves. (emphasis is on "themselves" meaning no one else can get in the mix)
3) Its size is greater than s(s corresponds to how finely you want to classify people and is your choice)

Claim 2. Chinese, Koreans and Japanese are distinct races

The problem of proving the above claim(i.e. finding a feature representation to separate these groups) is the focus of this project and will be refered to as the CKJ problem.

B. Overall Plan

The project is carried out as follows:

1) Dataset Construction
2) Data Preprocessing
3) Classifier training(CNN, KNN, SVM etc.)
4) Comparison of classifers

II. Related Work

A. Traditional Approaches

Traditional approaches in face recognition generally emphasizes dimension reduction/feature extraction[10](they're the same thing). These will be discussed below.

1) Dimension Reduction: The need for dimension reduction arises from the fact that faces occupy only a subset of an image. This is known as the "curse of dimensionality[10]". While this condition holds for many recognition problems, what is special about face recognition is that given the constrained nature(as you control for pose and lighting etc) of most face images, this subset region is consistent across different faces. Hence, general dimension reduction techniques are possible.

There are two types of dimension reduction techniques, geometry preserving and non-geometry preserving.

Geometry preserving dimension reduction techniques take images and output images. Examples[10] include:

- Face cropping
- Local geometric feature cropping(eyes, nose etc)

Non geometry preserving dimension reduction techniques take images and output a reduced representation of the images which are not images themselves. Examples[10] include:

- PCA based techniques(Eigenface, Fisherface etc)
- Gabor Wavelets
- Local Feature Analysis(LFA)

2) Relevance of the Traditional Approach: Given the "end-to-end" nature of CNN, the non geometry preserving dimension reduction techniques will not be attempted in this project. The implementation of these traditional techniques will be used for the CS231A project report.

B. Related Experiments

The table below summarizes the classification accuracies obtained by other researchers in related problems.

Problem	Caucasian/Asian	Southeast Asian/East Asian
Accuracy[%]	FRGC 2.0	80.37
Database	80.37	440 Face Images
Source	[1]	[8]

As of 2015/3/10 there is no published paper on the CKJ problem.

III. Experiments

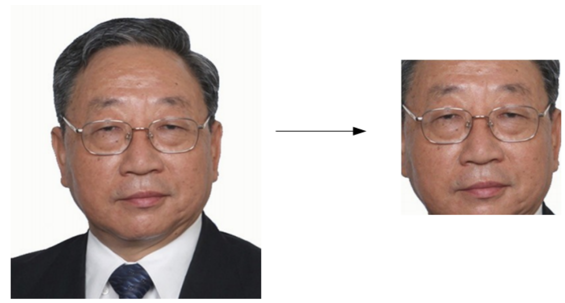
A. Data

Lack of an Asian face dataset is a common problem when dealing with face recognition related to this group[6]. Of the datasets that are available, I was not granted access to them after email requests, and had to resort to construction of a dataset myself.

I have generated my own dataset of the face shots of government officials in all 3 countries taken from the websites of the National People's Congress of China[2], National Assembly of Korea[3] and the House of Councilors of Japan[4]. This dataset will be referred to as the GOV dataset. There are about 500 pictures from each country. On the other hand, I have also taken the faces of the alllooksame quiz[5] from alllooksame.com. This dataset of 18 images will be called the ALS dataset and will serve as the ultimate test set. According to [5], the average human accuracy on the ALS dataset is 38%.

B. Data Preprocessing

1) Face Cropping: The Haar Cascade Classifier method was used to detect and crop faces from the photographs. In particular, this project makes use of an implementation of the method in OpenCV[7]. All pictures from all databases used in this project had gone through this processing stage.



2) Data augmentation: Due to a shortage of pictures, all 1500 pictures in the GOV database went through data augmentation. The idea is that someone's ethnicity is independent of brightness and horizontal flipping. In particular, all images were horizontally flipped and a random number is added to all images to change brightness. This step multiplied the number of images in the GOV database by four and I ended up with 6000 images.

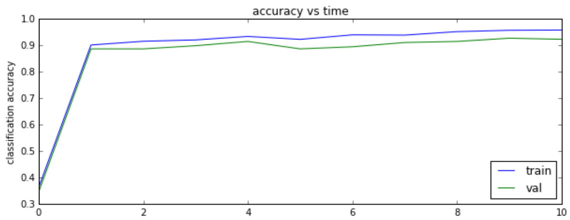
C. Classifiers

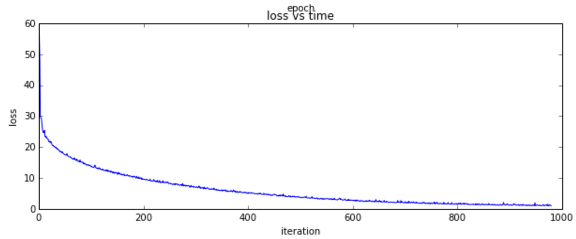
I have tried both training CNN models from scratch, using K-nearest neighbors and linear SVM classifiers. Since there is not a pretrained model that focused on faces, transfer learning is not expected to work well.

1) CNN: Two types of CNN were tried, summarized by the table below. These models made use of the CS231N assignment 3 code.

Number of Layers	Dropout Rate	Input Size	Learning Rate	Regularization
3	0.3	32x32	0.00005	0.05
5	0.6	64x64	0.00005	0.05

Below is a graph of the training of a 5-layer network.





2) PCA for dimension reduction: The experiments done with the SVM classifier operate on the first 100 principal components obtained by PCA.

D. Results

The table below summarizes the results obtained from various models. SJ Park and HC Lee are humans that were tested against both datasets. 20% of the GOV dataset and all of the ALS dataset were used for these test results.

Dataset	SVM	KNN	3-Layer CNN	5-Layer CNN	SJ Park	HC Lee
GOV	45%	80%	88%	90%	55%	58%
ALS	38%	33%	38%	33%	33%	44%

IV. Conclusions

Based on the results, there are three conclusions that can be drawn from this project, described below.

A. Drop in Performance from GOV to ALS

There is a drop in performance when one shifts from the GOV to ALS dataset for all classifiers. This can be explained by three reasons

1) ALS Quiz is biased: The ALS quiz may have been constructed in such a way as to blur the differences and encourage low score.

2) Presence of emotions: While the GOV dataset, being formal head photographs, is relatively free of emotions, the ALS dataset is filled with emotions. Hence, the classifiers trained should not be expected to generalize to ALS.

3) Variation in age: The GOV dataset is mostly consists of older people whereas the ALS dataset is consists of younger people. Just like expressions, the classifier should not be expected to be robust to variations in age.

B. Size of Dataset

The lack of face datasets labeled with ethnicity is a general problem in face recognition[6], the discussion above as well as the results point to classifiers that perform well on a limited demographic but do not generalize for factors like expressions and age. Hence, in order to build a classifier that is robust to these changes, a bigger dataset has to be constructed.

C. CKJ Problem for a limited demographic is well handled by CNN

The CNN classifiers above do well on the CKJ problem, especially considering that a related problem of Southeast/East Asian classification problem that does not make use of CNN[8] accomplishes only 80% accuracy. This points to the effectiveness of CNN at solving face recognition tasks. The drop in performance from GOV to ALS also indicate that CNN classifiers may be more prone to overfitting when used in face recognition problems.

References

[1] G. Toderici, S. M. O'Malley, G. Passalis, T. Theoharis, and I. A. Kakadiaris, "Ethnicity- and gender-based subject retrieval using 3-D face-recognition techniques," Int. J. Comput. Vis., vol. 89, no. 2, pp. 382–391, 2010. II-B
[2] http://www.npc.gov.cn/ III-A
[3] http://korea.assembly.go.kr/ III-A
[4] http://www.sangiin.go.jp/ III-A
[5] http://www.alllooksame.com III-A
[6] Fu, S., He, H., Hou, Z.: Learning race from face: A survey. IEEE Trans. on Pattern Analysis and Machine Intelligence (2014), (In Press) DOI: 10.1109/TPAMI.2014.2321570 I-A, 1, III-A, IV-B
[7] Face Detection Using Haar Cascades, "Face Detection Using Haar Cascades - OpenCV 3.0.0 - dev Documentation, Web, 08 Mar. 2015 III-B1
[8] U. Tariq, Y. Hu, and T. S. Huang, "Gender and ethnicity identification from silhouetted face profiles," in Proc. IEEE Int. Conf. Image Process., 2009, pp. 2441–2444. II-B, IV-C
[9] http://cs231n.stanford.edu/
[10] Handbook of Face Recognition, S.Z. Li and A.K. Jain, eds. Springer, 2011 II-A, II-A1
[11] T. A. Ito and B. D. Bartholow, "The neural correlates of race," Trends Cogn. Sci., vol. 13, no. 12, pp. 524–31, 2009.

I