Face Detection and Recognition

25/05/2017

Yinbo Yang (U5457700, contribution: 50%)

Yuanxin Wei (U5703881, contribution: 50%)

The Australian National University

ABSTRACT

This article is mainly focus on solving the face detection and recognition from a real classroom image, the input of the system should be a class photo and an individual person photo, the output should be the number of students in the class photo and add bounding box around their face and meanwhile, find if the give student in this class. This project can be separated into two functional part. The first task is face detection that counts and extracts faces by implementing a cascaded Viola-Jones face detector which boosts up accuracy compared with original method. The second task is face recognition, in order to recognize the face, we adopted Principal Component Analysis plus K-nearest neighbor [3][4] at classification and Discrete Wavelet Transform [6] at image preprocessing.

I. INTRODUCTION

There is a real-life problem that lecturers want to statistic the attendance of their lectures by some pictures taken from podium in different views and use it as the reference of evaluation of students. Our project is designed to deal with this problem. The two main tasks of our project are face detection and face recognition

1. Face detection: for a given class image, the face detector should find and count all faces and then extract them as the inputs of second task. The goal of this task is to get high accuracy as well as high recall rate, however, there are some problems we might encounter that extreme illumination conditions, variety of human poses and overlapped faces. In this report, we propose a cascaded Viola-Jones face detector based on Viola-Jones face detector [5] which increases significantly on accuracy rate. Because we have been provided 414 “ANU faces” images as our training set, test set includes faces that are detected and extracted by my cascaded Viola-Jones face detector, so we put training set into my cascaded Viola-Jones face detector to do the pre-process that ensures training set has similar format with test set.
2. The face recognition part gives each student that detected by our face detection algorithm a label, the number of the label means which class they are assigned to (we have 414/3 = 138 classes). We preprocess all the test image and test image with Discrete Wavelet Transform algorithm. After preprocessing all the training and testing data, we applied PCA algorithm to get the eigen

face and use KNN to classify the projected test vector.

II. Related Work

In 2001, Viola and Jones proposed an object detection framework which can firstly detect object in real time with decent detection rate. Based on Viola-Jones object detection framework, they proposed a robust real-time face detection by 2002 that is even wildly used in today's world because it is open-sourced on OpenCV and Matlab and it has high detection rate. Apart from Viola-Jones face detector, face detectors based on Convolutional Neuron Network (CNN) [2] have achieved remarkable performance. For example, Kaipeng Zhang [8] etc. proposed a new framework using cascaded CNNs to deal with face detection and alignment. The idea of this framework is that candidate face sub-windows are refined throughout layers (CNNs). This framework achieved very good performance around 95% accuracy rate and outperformed the state-of-art methods in Face Detection Data Set and Benchmark (FDDB), WIDER FACE [7] in 2015.

There are several common methods of face recognition. First one is support vector machine(SVM), the basic idea of SVM is convert the low-dimensional linear inseparable problem to high-dimensional linear separable problem. The feature of SVM are long training time and complex algorithm. Second method is convolutional neural network: Do convolution operation on the give training image to get the features, this method needs large numbers of training image. Another one is Principal Component Analysis： Based on KL transform, PCA formed a training set matrix and do the svd brake down on the covariance matrix, get the top-k largest component based on the eigen value. The accuracy of PCA is higher than 90% on yale face database.

III. Approaches

Face detection:

- *Cascaded Viola-Jones detector*

Inspired by the idea of Kaipeng Zhang etc., we try to propose a new framework called cascaded Viola-Jones detector. We try to ensure the candidate face detected by Viola-Jones detector are real faces, so the idea of our improved detector is that we apply Viola-Jones detector to detect facial parts on those candidate faces to eliminate non-facial parts.

The procedure of our cascade Viola-Jones detector is:

Firstly, we apply Viola-Jones detector to detect candidate faces and extract them on "ANU face class".

Secondly, we apply Viola-Jones detector to detect facial parts like left eye, right eye, both eyes, nose and mouth on candidate faces. Here we need to balance the accuracy and time-consuming. There is need to detect all parts on candidate faces because candidate faces have variate pose that some parts cannot be detected, so we only need to detect two parts and ensure at least one is detected on candidate face.

Thirdly, we treat candidate faces that are detected at least one part as faces and the rest are not faces.

The architecture is shown in Figure 2.

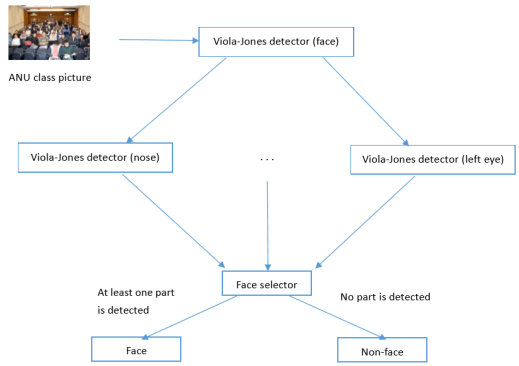


Figure 2: Architecture of cascaced Viola-Jones detector

- *Apply face detection method on training set:*

  In order to uniform the training, set and the test set, we apply same detect method on both training set (the 414 given students photo) and the test set (49-52 students photo detected from class image) After this process, the uniform level of two sets are higher than before.

  As the following figure 1 shows, the format of new training image applied the face detection algorithm is much closer to the test image.

Original training image	new training image	test image
		

Figure 1: comparison of three image

Face recognition

- *1: Discrete Wavelet Transform:*

  The basic idea of Discrete Wavelet Transform is divide the image into different frequency part.

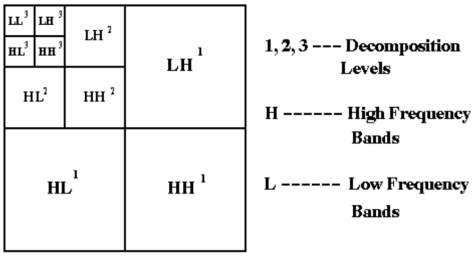


Figure 3: 2D-DWT with 3-Level decomposition [6]

Figure 3 shows the 3-level decomposition of discrete wavelet transform.

The LH^1 is horizontal fluctuation which reflects the horizontal feature of the original image, HL^1 is vertical fluctuation which reflects the vertical feature of the original image, HH^1 is first diagonal fluctuation which reflects the diagonal feature of original image.

That should be an LL^1 which shows the original picture in the lower dimension. According to the figure 3, the LL^1 part can also be break down to the lower dimension and get LL^2, LH^2, HL^2 and HH^2 and we can do same operation on LL^2 as well[6].

Considering that the given test ANU faces test, all the pictures are in the different light condition and contain different kinds of noise.

I adopt Discrete Wavelet Transform to preprocessing the given training pictures in order to lower the influence of the light condition and removing noise.

Here, I just use 2D-DWT to all the training image with the first-level decomposition and combine the LL, LH, HL and HH part with the following weights:

$$NEW\_img = LL + 0.5HL + 0.5HL$$

The reason that I remove the HH part is, the diagonal information for a human face is much

useless than the vertical and horizontal information.

The result of 2D-DWT process like following:

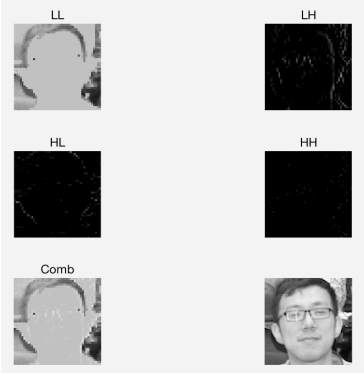


figure 4: 2d-DWT results

- *2: Principal Component Analysis*

  Based on the background knowledge on c-lab4 we apply PCA algorithm in this project to build the training combination matrix A, size of A is 65536*414.

  The 65536 = 256*256 is all the pixels of one image and the 414 reflects the number of training image.

  After the svd braking down, we choose top k =10 components and the energy is 99.5913% as figure 5 shows:

```
persentage =

    99.5913
```

figure 5: energy percentage

The ten eigenfaces:

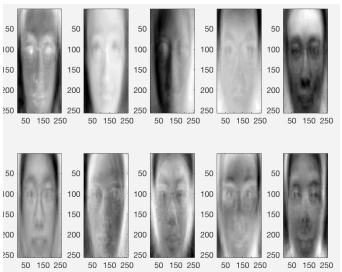


figure 6: ten eigenfaces

- *3: K-nearest neighbor:*

  Finally, we apply KNN to classify the projected testing vector, we compute all the distance from test vector to training vector and let k=50.

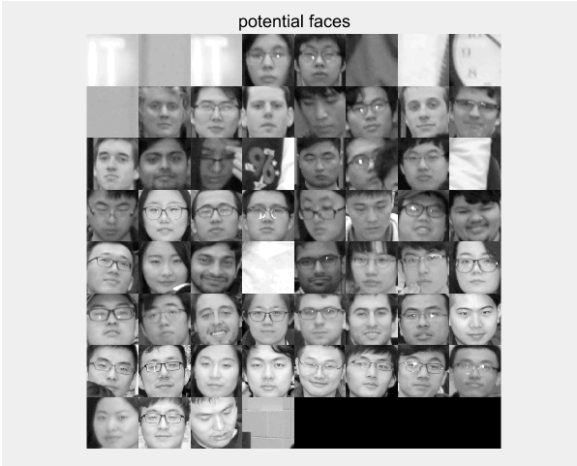
Then gives all the testing image a label.

IV. IMPLEMENTATION AND RESULTS

Face detection:

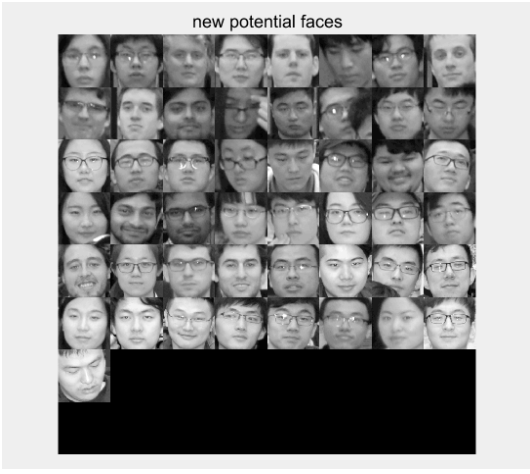
- Pure Viola-Jones face detector tested on "anuclass01.jpg"

This figure shows that the result of candidate faces processed from Viola-Jones face detector. The precision rate is 49/60, and the recall rate is 49/59.



- Cascaded Viola-Jones face detector (left eye + nose) tested on "anuclass01.jpg"

This figure shows that the result of candidate faces processed from cascaded Viola-Jones face detector. We detect left eye and nose by using Viola-Jones face detector on candidate faces. The precision rate is 49/49 that is improved significantly, and the recall rate is 49/59.



Face recognition:

- PCA accuracy = 2/49, nearly zero.

- PCA plus KNN accuracy = 12/49

- LBP preprocessing, knn and pca accuarcy = 6/49

This result is unexpected, we still trying to find the reason of that, according to the article [1] [9]the result after LBP processing should be higher than just take PCA.

- 2D-DWT , PCA plus KNN accuracy = 13/49

The result here is a little bit higher than PCA plus KNN.

V. Discussions

Face detection:

The advantages of our cascaded Viola-Jones face detector:

1. It significantly improves the precision rate compared with original Viola-Jones detector.
2. It is easy to implement because there is open-sourced pre-trained model on Matlab and OpenCV, we do not need to train new models.
3. It achieves relatively high recall and precision rate on our test images.

However, the drawbacks are also obvious:

1. Viola-Jones face detector is not state-of-art method in face detection; CNN could achieve better performance than it. Our face detector is sensitive to extreme illumination and variety poses.
2. Our face detector sacrifices time to improve precision rate.
3. It cannot improve recall rate and may eliminate some real face candidates.

Face recognition:

First of all, the reason that we choose PCA is we have done series of validation to prove that PCA will work in this project.

In order to do the validation for PCA algorithm, we train the yale face training set and test with yale face testing set, the accuracy is higher than 90%, we also tried to use the first two images in the anu faces set (276 images) as the PCA training set and choose 20 images in the third image and the accuracy is

10/20 = 50%. When we change the test set to the detected faces in the class the accuracy is nearly to zero.

According to the above validation, we find that the uniform level between training set and test set and the length of training set (how many images for each person) will influence the accuracy of PCA method.

After find this issue, we mainly focus our attention on image preprocessing in order to improve the quality of the training set and testing set, so we tried the LBP and 2d-DWT and finally decide to choose 2d-DWT to preprocess the image.

Secondly, in the sample PCA method, we just assign the test vector to the nearest vector in the training set, apply KNN classifier here should improve the classification obviously, at least, KNN will not decrease the accuracy of PCA.

VI. CONCLUSIONS

1.We propose a cascaded Viola-Jones face detector in face detection and achieves relatively good results that precision rates are over 94% and the recall rates are around 85% on two test images.

2.We have tried to setup deep learning framework, but some errors happened in CAFFE setup, we do not have enough time to waste on environment setting, so we choose to use other methods to deal with this problem. In our future work, we will implement CNN method to process both face detection and recognition.

3.Base on the ANU face recognition problem, use 2D-DWT to preprocess the image can remove the light condition influence and remove noise at the same time, after applying the 2D-DWT on both training set and test set, apply PCA plus KNN method to do the face recognition, the result shows that the accuracy will be higher than just singly use one of each.

VII. LEARNING OUTCOMES

1.We understand and explore different algorithms of face detection such as Viola-Jones face detector, deportable part-based models and convolutional neuron networks. And we analyze the pros and cons of above algorithms.

2.We propose an improved cascaded face detector based on Viola-jones face detector.

3.The length of training set (how many images in a same class, in this project length of training set = 3 because every student has three photos in the training set) influence the accuracy of both PCA and KNN algorithm, based on our test on yale face database, if we enlarge the length of training set form 3-10, the accuracy of both PCA and KNN will increase significantly.

4.The uniform level of training set and test set can influence the accuracy of PCA as well, the higher uniform level they are, a higher accuracy will be. (based on the comparison of test on yale database test and real class image test.)

VIII. REFERENCES

[1] Ahonen, T., Hadid, A., & Pietikainen, M. (2006). Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence,28(12), 2037-2041.

[2] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks.," *Advances in neural information processing systems,* p. 1097–1105, 2012.

[3] Dudani, S. A. (1976). The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics, (4), 325-327.

[4] Peterson, L. E. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883.

[5] P. VIOLA and M. J. JONES, "Robust Real-Time Face Detection," *IEEE International Conference on Computer Vision,* 2004.

[6] Mohideen, S. K., Perumal, S. A., & Sathik, M. M. (2008). Image de-noising using discrete wavelet transform. International Journal of Computer Science and Network Security, 8(1), 213-216.

[7] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A Face Detection Benchmark". arXiv preprint arXiv:1511.06523.

[8] Z. Kaipeng, Z. Zhanpeng, L. Zhifeng and Q. Yu, "Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Networks," *Computer Vision and Pattern Recognition (cs.CV),* 2016.

[9] Zhang, G., Huang, X., Li, S. Z., Wang, Y., & Wu, X. (2004). Boosting local binary pattern (LBP)-based face recognition. In Advances in biometric person authentication (pp. 179-186). Springer Berlin Heidelberg.