FEATURE IS ALL YOU NEED: A COMPARATIVE STUDY OF THE FACE RECOGNITION APPROACHES

Pinyuan Feng

School of Computer Science, University of Nottingham

ABSTRACT

The past decades have witnessed a rapid devlopment in the field of computer vision. Face recognition, one of the most representative visual computing tasks, has attracted significant interests from computer scientists due to its huge potential in various applications, such as mobile payment, information encryption, criminal investigation, multimedia interaction, and so on. While it is a promising and popular sub-area of computer vision, a number of technical challenges still remain in the research activities because face images are usually influenced by the external environment, including illumination, occlusion, rotation, etc. In this report, both traditional and modern solutions in face recognition are presented to see how researchers address or alleviate the problems. Additionally, the result of the method evaluation is discussed to make comparisons.

Index Terms— Face Recognition, Feature Extraction, One-shot Learning, Computer Vision, Deep Neural Network

1. INTRODUCTION

Can machines capture the biometric features from human faces? This question has attracted a large amount of attention in the field of computer vision for decades. A number of multitude face-based tasks have been established and performed, such as face recognition, face detection, face verification, facial tracking, and so on. Among them, face recognition is the most long-standing computer visual topic. It represents an automated process to identify or verify the identity of a human from an image or a video stream. From human perspectives, face recognition is a simple task related to cognition and perception. However, it is not that easy for machines, because they are expected to learn to convert the faces into machine-recognisable feature matrix and attempt to match the features with faces regardless of noise.

1.1. Challenges

While the face recognition brings opportunities into the society, the challenges remain to restrict the development of the area to a further step. Basically, there are six types of challenges that researchers have been seeking to resolve. Firstly, the low-quality resolution of an input facial image prevent itself to contain sufficient information, which makes systems hard to extract the useful feature to distinguish the input human face from other faces. Secondly, illumination variation may cause the change of face appearance on a pixel-based level. Factors, such as the light intensity, the direction of light source, and shadow, etc., can significantly affect the face recognition task. Thirdly, face recognition is sensitive to pose variation, which mainly results from camera position, camera rotation, head movement. In addition, occlusion usually occurs when some parts of the human face are covered by sunglasses, facial masks, or other face accessories. This means useful features are partially available. Moreover, human faces are dynamic because of various facial expressions caused by emotions or other cognitive activities. It is difficult for systems to find the shared features in those facial patterns from the same person, especially for one-shot learning tasks. Lastly, aging problem typically reflects on the face appearance. When a person grows up, the changes of the skin, muscle and other facial aspects will affect the appearance and thus decrease the face recognition performance. Overall, the factors mentioned above are common in face recognition tasks and they may influence the performance in an unconstraint manner. Therefore, the topic deserves further investigating in the future.

1.2. Contributions

The project is about a comparison between two approaches and a baseline method. Therefore, a traditional method and a modern method was selected to make an evaluation. All the implementation was based on MATLAB. The primary contributions of this project cover three aspects and will be described in this paperwork:

- Traditional methods have been implemented to find the one with the best performance.
- A modern face recognition approach powered by deep learning technique has been adopted to show the state-of-art performance.
- All approaches have been evaluated in comparison with the given baseline.

2. RELATED WORK

This section primarily introduces the past face recognition approaches and implementation techniques that related to the project. Apart from presenting the related work, a summary about what techniques and algorithms have been leveraged will be illustrated.

2.1. Traditional Approaches

The study of face recognition grew to a high level in the early 1990s. The introduction of Eigenface [1] has been considered as a millstone and it paved the way for the development of feature-based face recognition in the past years. Generally, it was designed to use a holistic approach to acquire and process the global features in the human face. In 1997, Fisherface [2] was derived from Eigenface [1] to linearly project the original facial image into low-level sub-spaces. It took advantages of Linear Discriminant Analysis (LDA) to alleviate the problems of illumination variation. To further solve the problems posed by uncontrolled face variation, local features attracted many researchers' attention. Local Binary Pattern (LBP) [3], a texture descriptor, was applied to the face recognition. By dividing the face image into several regions, LBP features was detected and combined to represent the facial feature. Later, the local filtering approach was extended [4] to make it robust enough to treat the faces at different scales. In 2005, HOG, short for Histogram of Oriented Gradients [5], was proposed to focus on the edge features. The extracted gradients and orientations of the edges were designed to encoded into histograms to generate a histogram-based features. In this project, Eigenface [1], Circular LBP [4], HOG [5] have been implemented as representative traditional approaches.

2.2. Deep Face Recognition

The traditional approaches were improved the face recognition slowly, because the researchers usually focused on only one facet of unconstrained noise mentioned previously while ignored other factors. However, the success of AlexNet [6] stimulated the interests in the computer vision community and many computer scientists started using deep learning to perform face recognition tasks. The empirical practices of the deep neural network have indicated that it can achieve the state-of-the-art performance. Starting from 2014, a series of deep learning architecture have been proposed to reshape the face recognition research community, such as Deepface [7], DeepID [8], [9], VGGFace [10], [11], FaceNet[12], etc. They were designed to extract high-level features by applying the input face to go through consecutive layer-based operations. Although it took a long time to train a model for face recognition, the result revealed that the performance was considerably enhanced. The deep learning models trained from those deep neural network backbones can handle the face recognition challenges perfectly. In this project, FaceNet [12] has been programed as a typical approach of deep face recognition. More details will be illustrated in the next section.

2.3. Data Augmentation

The performance of face recognition solutions significantly depends upon the quality and volume of the training dataset, especially for deep face recognition methods. However, sufficient face dataset that is both high-quality and balanced is not accessible and available, because the labeling is a laborious and high-cost process. Besides, the capturing human faces from multiple facets and under different environmental conditions is impossible. To resolve the data shortage problem, data augmentation approaches are commonly utilized to enlarge the dataset. According to a survey for face data augmentation [13], the techniques can be classified as generic transformation, component transformation and attribute transformation. In the process of implementation, geometric transformation, a type of generic transformation technique, was applied to expand the dataset for traditional approaches.

3. METHODOLOGY

Face recognition aims to find the identity of the input face from a face database. Different from face verification, it is a one-to-many matching task. According to the given source code of the baseline method, the implementation is based on intensity-based template matching. This can be regarded as "one-to-many" face feature verification. For each facial image in the dataset, it traverses the whole testing dataset to perform the cross correlation and finds the target with the maximum correlation value, where all the pixel values are served as features without further processing. In general, the core idea of the two alternatives was modified from the baseline method. The implementation mainly contains three phases: 1) Face Localization, 2) Feature Extraction, 3) Classification.

3.1. Face Localization

Face localization is the process to detect the human faces in an image and use a bounding box to indicate its location and size. This idea was leveraged in the project in that it was able to prevent unnecessary information in the image from discouraging the recognition performance. In MATLAB, the built-in function, "*vision.CascadeObjectDetector*", for face localization is provided to help cropping the face. The face detector adopts the Viola-Jones algorithm [14], which receives an input image and outputs a square human face. In the process of the implementation, sometimes the human faces were failed to be detected. Therefore, the unexpected cases have been handled: 1) If no face is found, then nothing needs to be done. 2) If the bounding box is smaller than a predefined threshold, then nothing needs to be done. 3) If multiple bounding boxes are detected and the largest one does not match the previous case, then only the largest one will be cropped out. After cropping all the images, the human faces were resized to the same dimension as a pre-processing step for feature extraction.

3.2. Feature Extraction - FaceNet

FaceNet [12] was adopted to extract the facial features in the project. Different from other deep learning approach, it learns the features from a facial image with the size of 224*224*3, and then encoded them into a 128-dimensional face embedding as an output. This was implemented based on the Tensorflow-Keras Deep Learning Package in MATLAB. A pre-trained model with weights was downloaded from the web, and then was imported into a format supported by MATLAB. Since the model was pre-trained on a large face dataset, it already had the ability to generate a valid facial embedding and there was no need to fine-tune the model. The process of importing was complicated, because some layers were not supported by MATLAB. Therefore, those layers were re-defined in MATLAB and the whole network architecture was assembled again. In addition, there was no common output layer in the original architecture and MATALB did not accept this manner. To output the 128-dimensional face embeddings, an output layer was defined as well.

3.3. Feature Extraction - HOG

HOG [5] was the second feature extraction approach to be used in the project. To obtain the HOG-based facial features, a built-in function, "*extractHOGFeatures*", was leveraged to perform feature extraction. The parameters "CellSize" and "BlockSize" were iteratively tried to find an optimal combination for better performance.

3.4. Classification - SVM

Support vector machine (SVM), a supervised machine learning approach, is often used in the classification problems. According to the labeled data, it can learn the data properties and separate the classes by decision boundaries. SVM was utilized to classify the face features and map the features with the target human face. The built-in function, "fitcecoc", was called to perform the "one-to-many" feature classification.

3.5. Data Augmentation

In the original dataset, there is only one facial image for each category. However, the testing images within the same class are multiple. For traditional approaches, inferring the face identity from one image may not be able to handle the facial variation in the testing dataset. Therefore, a set of geometric transformation techniques were applied to the dataset, such as flipping, rotating, etc.

4. EVALUATION

This section presents the evaluation of the face recognition approaches. Initially, a traditional approach and a modern approach were design to be implemented for this project. However, there was no idea about what existing algorithms outperform the baseline. Finally, seven combinations of feature extraction and classification were implemented. The feature extraction approaches include FaceNet [12], Eigenface [1], Circular LBP [4], HOG[5], while the classification methods are based on SVM and Cosine Similarity. Therefore, all the approaches will be compared with the baseline.

Table 1 Result without Data Augmentation

Methods	Time (s)	Accuracy (%)
Baseline	*29.352*	*25.372*
FaceNet	263.1094	97.247
HOG+SVM	138.7571	48.3631
LBP+SVM	73.2132	35.7143
EIG+SVM	128.3148	26.8601
HOG+COS	100.8934	43.1548
LBP+COS	41.4549	35.3423
EIG+COS	96.3114	22.9911

Table 2 Result with Data Augmentation

Methods	Time (s)	Accuracy (%)
Baseline	29.352	25.372
FaceNet	263.1094	97.247
HOG+SVM	160.4739	52.753
LBP+SVM	83.0495	37.128
EIG+SVM	166.6573	28.4226
HOG+COS	131.2572	47.6935
LBP+COS	61.4945	34.9702
EIG+COS	116.7089	28.4226

From the tables above, the result reveals that the combination of HOG [5] and SVM outperforms other traditional approaches. However, the time spent on the training process is the longest among traditional methods. There is no doubt that FaceNet [12] is the best approach, because it reaches above 97% recognition score, although the training time is much longer. Besides, this deep recognition method does not rely on the additional data augmentation. All the learning process is based on one facial image. This indicates that the approach is promising and can make up for the influence caused by the data shortage. The reason why the approach takes longer time is that it takes some time to load the model before the feature extraction.

5. CONCLUSION AND SUMMARY

In this project, traditional approaches and a deep learning approach for face recognition were evaluated. The result shows that the deep learning approach is robust to handle the facial recognition challenges, such as lighting condition, occlusion, and face expression.

6. REFERENCES

[1] M. Turk and A. Pentland, “Eigenfaces for Recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Jan. 1991, doi: 10.1162/jocn.1991.3.1.71.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997, doi: 10.1109/34.598228.

[3] T. Ahonen, A. Hadid, and M. Pietikäinen, “Face Recognition with Local Binary Patterns,” in *Computer Vision - ECCV 2004*, Berlin, Heidelberg, 2004, pp. 469–481.

[4] T. Ahonen, A. Hadid, and M. Pietikainen, “Face Description with Local Binary Patterns: Application to Face Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006, doi: 10.1109/TPAMI.2006.244.

[5] N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, San Diego, CA, USA, 2005, vol. 1, pp. 886–893. doi: 10.1109/CVPR.2005.177.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks,” *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.

[7] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1701–1708. doi: 10.1109/CVPR.2014.220.

[8] Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation from Predicting 10,000 Classes,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, Jun. 2014, pp. 1891–1898. doi: 10.1109/CVPR.2014.244.

[9] Y. Sun, X. Wang, and X. Tang, “Deep Learning Face Representation by Joint Identification-Verification,” *arXiv:1406.4773 [cs]*, Jun. 2014, Accessed: May 12, 2021. [Online]. Available: http://arxiv.org/abs/1406.4773

[10] O. M. Parkhi, A. Vedaldi, and A. Zisserman, “Deep Face Recognition,” in *Procedings of the British Machine Vision Conference 2015*, Swansea, 2015, p. 41.1-41.12. doi: 10.5244/C.29.41.

[11] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, “VGGFace2: A Dataset for Recognising Faces across Pose and Age,” in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi’an, May 2018, pp. 67–74. doi: 10.1109/FG.2018.00020.

[12] F. Schroff, D. Kalenichenko, and J. Philbin, “FaceNet: A unified embedding for face recognition and clustering,” in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 815–823. doi: 10.1109/CVPR.2015.7298682.

[13] X. Wang, K. Wang, and S. Lian, “A Survey on Face Data Augmentation,” *Neural Comput & Applic*, vol. 32, no. 19, pp. 15503–15531, Oct. 2020, doi: 10.1007/s00521-020-04748-3.

[14] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, Kauai, HI, USA, 2001, vol. 1, p. I-511-I–518. doi: 10.1109/CVPR.2001.990517.