Facial Features Extraction and Clustering with Machine Learning

Yunbo Yang
College of Arts & Sciences, University of Washington, Seattle, 98195, United States of
Americayunboy@uw.edu

ABSTRACT

Computer vision is always one of the most popular topics in machine learning, and face recognition is one of the essential parts of computer vision studies. Today, most studies focused on locating and distinguishing faces. In contrast, this paper focuses more on interpreting components of faces or facial features. Instead of using intentional high variance images to train robust models that fit multiple real-world situations, the study uses standardized photos from Chicago Face Database (CFD). Images are preprocessed to fit the transfer learning model using a slightly adjusted version of The Visual Geometry Group, 16 layers version model (VGG 16) for features extraction. Then data is again processed using Principal Component Analysis (PCA) for computational efficiency and clustered using the K-means algorithm with k election based on common knowledge. The results for flat facial features such as eyes and mouths show an effective clustering. In contrast, confusing results exist in clustering features such as noses and face shapes, which may require higher dimensional data for satisfying clustering. The resulting dataset consists of the name of each face image and its facial features labels, which can be used for further study relationships between facial features and races and genders. It can also be used for face generation programs based on selected facial features.

CCS CONCEPTS

• **Machine learning** → Neural networks; Features Extraction; VGG 16; Principal Component Analysis.

KEYWORDS

Facial features Extraction, K-means Clustering, Face Generator

ACM Reference Format:

Yunbo Yang. 2022. Facial Features Extraction and Clustering with Machine Learning. In *2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE 2022), October 21–23, 2022, Xiamen, China.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3573428.3573746

1 INTRODUCTION

In recent decades, face identification has become one of the most popular tools for classifying humans' identities based on their physical characteristics and unique features [1]. Many applications of face recognition are deployed in places under various computer vision applications like international airports, trade centers, and our phones or other devices [2]. Most current studies focus on human facial images to predict to classify an individual's demographic information such as race, age, and gender, or in other words, the classification of a specific facial image. In all the above cases, the studies focused on accuracy, trying to make a robust system capable of recognizing faces or specific results using intentionally rough data. However, the recognition accuracy will decrease with small changes in noises and variance even for the best face recognition systems [3]. The focus of increasing accuracy is on developing an overall robust algorithm or improving hardware.

Research in the area of facial recognition algorithms has developed exponentially since the late twenty century [4]. In the early time, there were several well-known identification methodologies. One of the popular algorithms is the eigenfaces model, the most popular version of which in modern days is Principal Component Analysis. For a human, facial recognition is more like an instinctive ability or a simple task learned since early childhood. However, in the field of computer vision, the image is a high-dimensional matrix with some possible patterns that are difficult for algorithms to interpret. Therefore, a reasonable solution is dimension reduction for both computational efficiency and interpretability. Even now, vector representations play a significant role in image recognition. Later, Fisher developed Linear Discriminant Analysis (LDA), which is considered as an improved version of PCA. Nevertheless, all of the algorithms focused on the classification of whole human faces.

Research in the area of facial recognition algorithms has developed exponentially since the late twenty century [4]. In the early time, there were several well-known identification methodologies. One of the popular algorithms is the eigenfaces model, the most popular version of which in modern days is Principal Component Analysis. For a human, facial recognition is more like an instinctive ability or a simple task learned since early childhood. However, in the field of computer vision, the image is a high-dimensional matrix with some possible patterns that are difficult for algorithms to interpret. Therefore, a reasonable solution is dimension reduction for both computational efficiency and interpretability. Even now, vector representations play a significant role in image recognition. Later Fisher developed Linear Discriminant Analysis (LDA), which is considered as an improved version of PCA. Nevertheless, all of the algorithms focused on the classification of whole human faces.

2 METHOD

2.1 Dataset

Most current databases involve photographs of individuals making various facial expressions and later self-validation. Many issues arise with the study using these databases, such as lack of homogeneity, standardization, and quality. This paper uses the main set of the Chicago Face Database (CFD) created by Debbie S. Ma, Joshua

EITCE 2022, October 21–23, 2022, Xiamen, China

ACM ISBN 978-1-4503-9714-8/22/10…$15.00
https://doi.org/10.1145/3573428.3573746

Figure 1: Example of one image processing

Correll, and Bernd Wittenbrink at the University of Chicago. The construction of CFD was aimed to develop an open resource for the science researcher's community to deal with the above issues as much as possible. To avoid homogeneity, the scientists included multiple white and black self-identified male and female faces. To achieve standardization, the stimuli are controlled and produced in high-resolution image files. The complete database includes both the basic image set (CFD) and several additional sets (CFD-MR, CFD-INDIA). The main CFD set consists of 597 unique images of individuals with various backgrounds; the CFD-MR extension set, where "MR" stands for multiracial ancestry, only consists of images of 88 self-reported multiracial ancestry unique individuals, and the CFD-INDIA extension set has 142 unique images of individuals from Delhi, India. All images are subjected to represent models with neutral facial expressions [5]. In a Word 2010 document, insert a picture.

For this paper, the study used only the main CFD set, consisting of 597 standard white-background 1718 by 2444, red, green, and blue (RGB) facial images of unique individuals who self-identified as Black, White, Asian, and Latino males and females. Images are divided and cropped into different sets of left eyes, mouths, noses, and processed faces. To avoid the influence of other facial features on face types clustering, eyes, noses, and mouses are covered to maintain generality. All sets are resized to 224 by 224 for later usage. Figure 1 shows the processing of one image.

2.2 Features Extraction

Recently, Convolutional networks (CNNs) have shown many excellent outcomes in large-scale image recognition. Due to its lack of interpretability and expensive computational cost, the pretrained CN model has become a commodity in deep learning, especially for computer visions. The Visual Geometry Group, 16 layers version model (VGG 16) achieves top 5 with a test accuracy of 92.7% test accuracy in Large Scale Visual Recognition Challenge 2014 (ILSVRC2014), using a subset from dataset ImageNet, a dataset of over 14 million web-collected and human-labeled images belonging to 1000 classes. The input layer is fixed to 224 by 224 RGB matrix, and in the previous step, the images used in this paper are processed to fit the layer. The input is then passed through multiple convolutional layers along with ReLU and max pooling layers, using kernel size three by three and one stride. The model utilizes a one-by-one convolution filter as a linear transformation of the input values. In the next three dense layers after the sets of convolutional layers, the first two layers each have 4096 channels, and the third one implements a 1000-way ILSVRC classification, thus containing 1000 target labels [6].

Figure 2 shows the structure of the original VGG 16 model. For the feature extraction task, this paper used a slightly modified version of the model, removing the output layers for clustering purposes, since the original model is for classification so that the 4096 channels are kept as the extracted features.

2.3 Dimension Reduction

Although each image's features are expected to be extracted through the previous step, the data representation is still considerably large. Therefore, further data processing is required for computational efficiency. A simple approach can be merging pixels and binarizing the images for dimension reduction purposes, but it tends to produce biased results. As the dataset consists of people with different skin colors, binarization with any threshold leads to a binary clustering based on skin colors, where the cut line is a specific threshold value. This paper used Principal Component Analysis (PCA) for dimension reduction and computational efficiency. PCA is a multivariate statistical algorithm for principal feature extraction [7]. For its simpleness and effectiveness, PCA is considered as a popular and standard tool in different image processing systems for exacting relevant information [8]. The desired result is the representation of data points as a set of orthogonal vectors, or principal components. Probably PCA is the most popular tool in various computer vision image processing applications for its simpleness and efficiency in exacting relevant information and patterns [9]. In this study, n-components for each subset are selected at 95% explained variance

2.4 Clustering

The paper uses Kmeans for data clustering. Kmeans algorithm is a recursive algorithm that tries to allocate points in the dataset into k unique non-overlapping clusters, where each point belongs to only one cluster. It recursively attempts to make the data points in one cluster as similar as possible while at the same time keeping points across the clusters as different as possible as well. It assigns data points to clusters wit ccentroids $\mu 1$, ..., μk , such that the sum of the squared distance, calculated based on a pre-defined distance function, between the data points and the cluster's centroid is at the minimum. For each set of points in cluster j, the function below is applied to find the center of the cluster as the new centroid, μj , the sum of data points belongs to cluster j divided by the number of data points in cluster j. Kmeans clustering is a very computationally cheap and interpretable algorithm. Possible issues may arise since the algorithm is sensitive to noises, which can be avoided through the previous data standardization and projection steps.

The only problem left is the selection of k. Considering the comprehensiveness of the dataset, k for each set can be selected based on the common knowledge that there are 6 eye types, 14 nose types, 9 mouse types, and 7 face types, so k selection of each set of eyes, noses, mouths, and faces can be 6, 14, 9, and 7, respectively. Other approaches to k value selection may be the "elbow method"

Figure 2: VGG -16 Structure

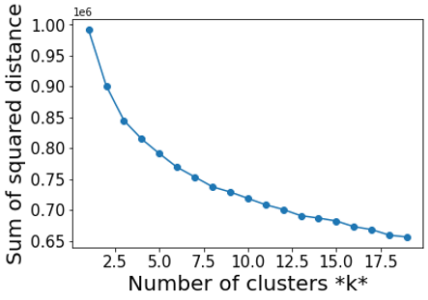

Figure 3: Sum of squared distance

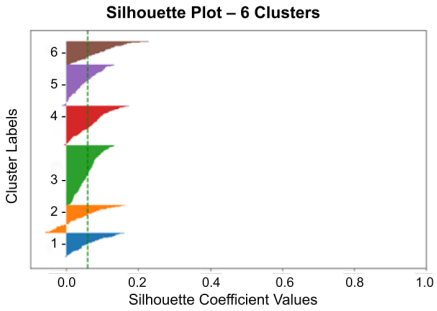

Figure 4: Silhouette scores

and silhouette score evaluation, but neither of the two seems to be plausible in this case.

Figure 3 shows that the sum squared distance decreases smoothly for the k value from 2 to 19, so an "elbow" point can hardly be selected in this case. Likewise, Figure 4 shows that a k value of 6, which means 6 clusters of eyes, results in a shallow silhouette score that is not representative of the clusters. Other subsets produced similar plots, proving that k selection methods may not be effective for this case. Therefore, k values based on existing knowledge will be kept.

2.5 Experimental Results and Analysis

Table 1 is the facial feature type of each image, represented by the image name. Eye type range: 0 – 5; Nose type range: 0 – 13; Mouth type range: 0 – 8; Face type range: 0 – 6.

Figure 5 includes three samples from each cluster. Based on human vision, the algorithm considered eye shapes, eyelids, and eyelashes and performed a neat clustering. In each cluster, a diversity of gender and race can be observed, showing that the algorithm is inclusive. Mouth clustering also has a good outcome, but the clustering for other facial features may not be as good as for eyes. For example, Figure 6 shows similar clusters and one sample from each cluster.

Figure 5 includes three samples from each cluster. Based on human vision, the algorithm considered eye shapes, eyelids, and eyelashes and performed a neat clustering. In each cluster, a diversity of gender and race can be observed, showing that the algorithm is inclusive. Mouth clustering also has a good outcome, but the clustering for other facial features may not be as good as for eyes. For example, Figure 6 shows similar clusters and one sample from each cluster.

3 DISCUSSION

In modern days, a human's face is the most preferred biometric characteristic, compared to fingerprints or DNA, which is probably most frequently used in different situations for ensuring identity security, such as customs, and governmental identity card management. Among all the applications of facial recognition for computer algorithms or human visions, individuals' facial features are the most significant for distinguishing different people. For any individual, it is easy for anyone that has seen the person to describe the person's physical appearance in plain text. However, the inverse progress, or visualizing the person's image based on the description in plain text, may not be as simple as the original task. One possible scenario can be the visualization of a criminal from eyewitness descriptions. At the same time, the dataset produced in this paper can be used for source picture selection since it has each picture with corresponding facial feature labels. The overall steps are: 1) Read text description input. 2) Find the suitable images in the dataset that have labels that match the description the best. 3)Morph selected images and output the result for verification. The first two steps are straightforward and to be achieved with simple coding. At the same time, issues arise with step three, even when using mature face morphing algorithms.

The study has attempted to use the face morphing algorithm, InterFace [10], to produce designated visualization using selected pictures. The InterFace is a software package written with MatLab

Table 1: Facial Feature Types

Picture Name	Eye Type	Nose Type	Mouth Type	Face Type
CFD-BF-200-080-N.jpg	2	2	3	4
CFD-BF-201-080-N.jpg	1	5	1	6
. . .. . .	. . .. . .	. . .. ..	. . .. . .	. . .. ..
CFD-WM-258-125-N.jpg	3	0	2	0
CFD-WM-255-219-N.jpg	3	6	5	0

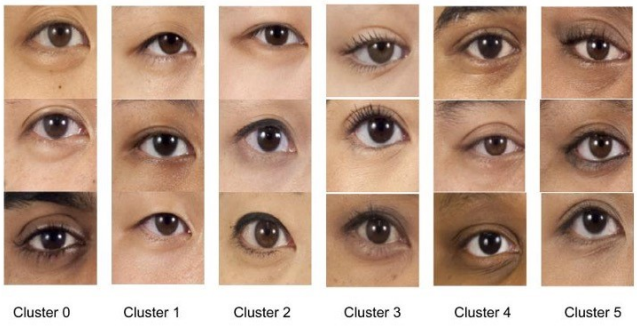

Figure 5: Sample Images for Eye clusters

Figure 6: Similar Noses from Different Clusters

for various studies in face recognition, which is capable of image reshaping and wrapping, averaging multiple face images, and morphing any two different faces. Another software figure includes tools for exploring the face features and structures based on PCA. The critical concept of InterFace is the distinction between facial textures and shapes. In recent years the software has become an essential way of segmenting images in many fields, such as graphical techniques and psychological hypotheses. Facial shapes are constructed according to the structure of the assigned points according to facial feature positions such as eye corners, nose tips, mouth sizes, and cheek outlines. The interface has 82 generated key points, which are annotated in the figure on the left. Using the above techniques, relying on the generated annotation of both images, the program stored the standardized texture and shape data for each image and then combined them in the later morphing part.

However, despite the software's excellent performance in morphing two faces, the algorithm can not generate satisfying results that fit the text description input, especially when the quantity of selected images is large. The algorithm tends to find the "average" face of all inputted images as the number of images used in the morphing process increase, consequently averaging the facial features of the images. The issue is not limited to any one morphing algorithm only. Still, it is likely to appear in most of the existing morphing techniques. A different approach is to treat parts of facial features separately continuously. Instead of morphing whole images of faces, the desired components are the average of each facial feature in this case. The overall steps are: 1) Read text description input. 2)Find the suitable images in the dataset that have labels that match the description the best. 3)Find the average image of each facial feature generated from a small subset of random images from the selected images. 4) Combine the selected images into a face for verification. The new approach ignores the concerns of averaging the features since the merging process is separate for each facial feature. Nevertheless, the result images do not look as natural as the previous algorithm since it combines parts from different faces.

4 CONCLUSION

The paper used standardized multi-background face data from CFD for facial feature extraction and clustering. The images used for facial features analysis include subsets of processed images of eyes, mouths, noses, and face shapes. For preprocessing and facial feature extraction, the paper referenced VGG – 16 and used a slightly adjusted model version. The following clustering steps involve classic algorithms of PCA and K-means clustering. A table is generated with the first column being the name of the original image and the other columns being the types of each facial feature. Through an analysis of the results, the paper demonstrated a robust clustering of flat facial features such as eyes and mouths.

In contrast, the clustering for tridimensional facial features such as noses and face types is not as good as the clustering of the other

two since the input data is entirely flat images. Then the paper discussed some future work and possible applications of the study. Effective clustering of more facial features, such as ears, should be achieved with higher dimensional data. For current data, models for the prediction of race and gender or the study of the distribution of facial features across various groups can be constructed. One further application is the face generation system, which uses morphing algorithms to generate faces with specific chosen facial features or text descriptions. However, most existing face morphing techniques average the input faces, especially when the input quantity is large. This results in the vanishment of initially outstanding facial features. Thus, the development of a face morphing system capable of keeping specific facial features in the future is desired for this application.

REFERENCES

[1] S. M. M. Roomi, S. L. Virasundarii, S. Selvamegala, S. Jeevanandham and D. Hariharasudhan, "Race Classification Based on Facial Features," 2011 Third National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics, 2011, pp. 54-57, doi: 10.1109/NCVPRIPG.2011.19.

[2] R. Raghavendra, K. Raja, S. Venkatesh and C. Busch, "Face morphing versus face averaging: Vulnerability and detection," 2017 IEEE International Joint Conference on Biometrics (IJCB), 2017, pp. 555-563, doi: 10.1109/BTAS.2017.8272742.

[3] Wei, P., Zhou, Z., Li, L. *et al.* Research on face feature extraction based on K-mean algorithm. J Image Video Proc. 2018, 83 (2018). https://doi.org/10.1186/s13640-018-0313-7

[4] K. Marwa and O. Kais, "Perspectives Of Classical Methods Of Face Recognition," 2022 International Conference on Intelligent Systems and Computer Vision (ISCV), 2022, pp. 1-8, doi: 10.1109/ISCV54655.2022.9806068.

[5] Ma, D.S., Correll, J. & Wittenbrink, B. The Chicago face database: A free stimulus set of faces and norming data. Behav Res 47, 1122–1135, 2015. https://doi.org/10.3758/s13428-014-0532-5

[6] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[7] Sheng Huang, Dan Yang, Ge Yongxin, Xiaohong Zhang, Combined supervised information with PCA via discriminative component selection, Information Processing Letters, Volume 115, Issue 11, 2015, Pages 812-816, ISSN 0020-0190, https://doi.org/10.1016/j.ipl.2015.06.010.

[8] X. Kang, X. Xiang, S. Li and J. A. Benediktsson, "PCA-Based Edge-Preserving Features for Hyperspectral Image Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 55, no. 12, pp. 7140-7151, Dec. 2017, doi: 10.1109/TGRS.2017.2743102.

[9] Abdi, Hervé and Williams, Lynne J., Principal component analysis, WIREs Computational Statistics 2, 2010, 433 – 459, https://doi.org/10.1002/wics.101

[10] Kramer, R.S.S., Jenkins, R. & Burton, A.M. InterFace: A software package for face image warping, averaging, and principal components analysis. Behav Res 49, 2002–2011, 2017. https://doi.org/10.3758/s13428-016-0837-7