Facial Analysis: Face Recognition, Expression Recognition, and Gender Identification

Machine Learning Group
Shraddha Mane and Gauri Shah
January 2018

PERSISTENT

Table of Contents

Executive Summary

Although humans are quite good at identifying known faces, we are not very skilled when we have to deal with a large number of unknown faces. On the other hand modern computers, with almost limitless memory and computational speed, could overcome these human limitations. One of the most successful applications of image analysis has been biometric face recognition technology; it has received significant attention, especially in the past several years due to a wide range of applications such as public surveillance and security, identity verification in the digital world, and modeling techniques in multimedia data management. Facial expression recognition is also important for targeted marketing, medical analysis, and human-robot interaction. In this white paper, we review a few techniques for facial analysis – the cloud platform AWS Rekognition, deep learning, and traditional feature extraction algorithms. We show that to get high accuracy, good quality data and processing power are a must. We present the results of our experiments which have been conducted over six different public as well as proprietary image data sets.

Introduction

Over the past few decades, face recognition and facial expression recognition have seen many commercial applications, and have become a popular area of research in the field of computer vision. Wide availability of low-cost desktop and embedded computing systems has created an enormous interest in the automated processing of digital images and videos. Recent advances in automated facial analysis have made it possible to develop automatic face recognition and facial expression recognition systems to address these requirements. Here are a few interesting recent applications:

- Alibaba has launched the 'Smile to Pay' service at KFC in China. Without a smartphone, users can make payments with just a smile.
- The iPhone X launched by Apple can be unlocked using facial recognition technology instead of tedious passwords and pins. Similarly, Facebook is testing a new feature where account owners can confirm their identity using facial recognition at the time of account recovery.
- In the UK, facial recognition is going to replace train tickets. Similarly, at the Singapore Airport, various services like bag-drop, immigration, and boarding are automated with the help of facial recognition.

Although research in automatic face recognition has been conducted since the 1960s, this problem has been largely unsolved in its entirety mainly due to the lack of processing power. Recent years have seen significant progress in this area owing to advances in data collection, face modeling and analysis techniques as well as the availability of high-quality cheap hardware that can quickly run computationally-intensive tasks. Systems have been developed for face detection and tracking, but reliable face recognition still offers interesting challenges in computer vision and pattern recognition research.

In this white paper, we overview several different techniques for a variety problems as follows:

- **Face Detection:** Given an image, locate the face in the image.
- **Face Identification:** Match a given facial image with a stored collection of identities, and return an identity.
- **Face Verification:** Check the identity alleged by the given facial image using a stored collection of images for this identity. This is a subset of the problem of Face Identification.
- **Facial Expression Recognition:** Identify if a facial image is displaying one of the seven emotions – anger, disgust, fear, neutrality, happiness, sadness, or surprise.
- **Gender Identification:** Identify if the user is a male or a female from the facial image.

Technologies

We explored three approaches used for facial analysis – feature engineering, deep learning, and cloud-based solutions. In feature engineering, we identify some unique features to classify the images. Deep learning algorithms such as Convolutional Neural Networks, require large data sets of high quality images to achieve good accuracy. Cloud-based solutions by large companies like Google, IBM, Microsoft, and Amazon use standard deep neural networks, and allow us to build customized models using specific input data.

Facial analysis has different challenges like pose, illumination, expressions, age, accessories, etc. Face detection is the first essential step for facial analysis; it locates the face in the given Image. We use the widely-used face detector proposed by Viola and Jones. After the face is located in the image, we extract the features or apply different filters to identify the user's identity, expression, and gender.

Feature Engineering

Feature Engineering is the use of domain knowledge to extract features from data used by machine learning algorithms. When the input data is too large to be processed and redundant, then it can be transformed into a reduced set of features. This process is called Feature Extraction.

> …some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used.
>
> — Pedro Domingos, A Few Useful Things to Know about Machine Learning

Facial landmarks play an important role for face recognition as well as for facial expression recognition. Landmarks locate the eyebrows, eyes, nose, lips and jawline in the face image. As these landmarks change due to different expressions or users, we can extract features from these landmarks. Other methods like HOG, SIFT, PCA, and a few more have been used for feature extraction. In some cases, these features may not be sufficient, for example, in the case of biologically identical twins. Some micro-level features on the face such as scars, moles and facial marks could be beneficial here.

Deep Learning

Getting a domain expert to identify how to extract unique features is quite a challenging and difficult task. To overcome this limitation, we use deep neural networks which are inspired by the human brain structure. We feed the input data as well as classification labels to the network, and it learns automatically how to identify objects. 2012 was the first year when neural networks grew to prominence as Alex Krizhevsky used them to win the ImageNet competition that year.

> Coming up with features is difficult, time-consuming, requires expert knowledge. "Applied machine learning" is basically feature engineering.
>
> — Andrew Ng, Machine Learning and AI via Brain simulations

Convolutional Neural Networks (CNNs) are a special type of deep neural network, specifically used for image processing. They are made up of different layers such as the convolutional layer, the pooling layer, and the fully connected layer. Typically, all the neurons in one layer perform one type of mathematical operation from which it gets its name. The machine can do image classification or object detection by looking at low level features such as edges, colors, and curves with the help of filters used in the convolution layers. Abstract concepts such as objects are then built up through a series of convolutional layers.

Deep neural networks usually need large quantities of data for training. Transfer learning is a relatively new approach where pre-trained CNNs are used for classification using smaller data sets. Standard CNNs like AlexNet, ZFNet, VGGNet, GoogleNet, and ResNet are trained on the ImageNet data set. We can use the pre-trained CNNs in two ways:

1. To extract the features from an image: We use these extracted features with any classical machine learning algorithm such as Naïve Bayes, SVMs, Random Decision Forests, etc. for pattern matching and classification.
2. To fine-tune the pre-trained CNNs with the smaller data set and use it directly for classification.

Cloud Based Services -

Companies like Google, Microsoft, IBM, Amazon, and other smaller start-ups provide cloud-based services for image analysis. These companies train deep neural nets using billion of images; they also keep updating the model by adding new images. We can then use their pre-trained model for classification of our own images. Due to transfer learning, even with a small data set we can achieve a reasonably good accuracy using these cloud services.

Experiments

Face Recognition

Face recognition is defined as the process of identifying a user from the provided facial image or video. Face detection is the first step for face recognition. After locating the face from an image, to identify the user we need to represent the detected face with features.

We used several different methods for face recognition which include the Amazon Rekognition cloud service, transfer learning with Google's Inception V3 model, and facial landmarks. We created our own dataset by considering real world challenges such as pose, illumination, expressions, presence of spectacles, etc. We experimented with two sets of training data using photographs of Persistent Systems employees: (1) One nearly frontal image per user, and (2) a black-and-white identity card with a facial photograph per user. The test dataset consisted of multiple images per user taking into consideration the above challenges. We obtained accuracy more than 90% using Amazon Rekognition. However, we got poor results with transfer learning and facial landmarks as the dataset is very small in size.

Facial Expression Recognition

Identifying the expression of a user from an image or a video is facial expression recognition. The expression can be neutrality, anger, disgust, fear, happiness, sadness, or surprise. Facial components like eyebrows, eyes, nose, lips etc. play an important role in expression recognition. Raised eyebrows, raised eyelids, flared nostrils, nose wrinkles, etc. decide the expression of the user. For example, when a person smiles the lip edges get stretched and muscles around the eye sockets contract. Using combinations like this one, we can identify the facial expression.

We used various methods like deep learning, transfer learning with the Inception V3 model, Amazon Rekognition, and facial landmarks for facial expression recognition on publicly available datasets such as JAFFE, CK, FER2013 and our proprietary employee dataset. We found that the accuracy of expression recognition decreases as we increase the number of expressions to classify. The highest accuracy achieved is around 97% with only two expressions (happy and neutral) using facial landmarks. Using all the 7 expressions for classification, accuracy goes down to nearly 65%.

Gender Identification

Gender identification is a more difficult problem compared to facial recognition and facial expression recognition. However, the difference in the face shape, eyebrows, chin, jaw, haircut and facial hairs are certain features which help to differentiate gender.

The IMDb data set has around 500K images with gender labels. As thousands of images are available, deep learning using CNNs is best choice to predict the gender. We used TensorFlow to train a CNN, and we got 96% accuracy for gender recognition.

Amazon Rekognition API also provides fairly accurate results for gender identification. We can also use these APIs to predict age but judging age is subjective, even for humans. Many times, even humans fail to predict the age of the person. The results given by these cloud services for age prediction are not up to the mark and there is room for many improvements.

PERSISTENT

Conclusions

We have evaluated several techniques in the area of facial analysis primarily focusing on face recognition and expression detection. In summary, Amazon Rekognition has the best performance for both face recognition and facial expression detection. It can handle challenges like pose, illumination, expression, age, accessories etc. According to the security requirements, we can adjust the threshold in the user identification system – the higher the confidence with which a face must be identified, the higher should be the threshold.

Convolutional Neural Networks also perform well when the data set is large enough for adequate training. For example, in case of gender identification, where we have ~500K images, the accuracy is 96%; however, for facial expression recognition where we have only 1200 images, the accuracy is significantly lower at 66%. In the case of facial expression recognition, we see that we get higher accuracy when we use fewer emotions for classification. Feature extraction using facial landmarks performs better than CNNs as well as transfer learning for facial expression recognition.

In conclusion, state-of-the-art techniques do a fairly good job at facial analysis with good input data, but there is room for significant improvement to push the accuracy to even higher limits in the absence of suitable training data.

PERSISTENT

About Persistent Systems

Persistent Systems (BSE & NSE: PERSISTENT) builds software that drives our customers' business; enterprises and software product companies with software at the core of their digital transformation. For more information, please visit: www.persistent.com

India

Persistent Systems Limited

Bhageerath, 402,
Senapati Bapat Road, Pune 411 016
Tel: +91 (20) 6703 0000 • Fax: +91 (20) 6703 0009
Email: info@persistent.com

USA

Persistent Systems Inc.

2055 Laurelwood Road, Suite 210,
Santa Clara, CA 95054
Tel: +1 (408) 216 7010 • Fax: +1 (408) 451 9177
Email: info@persistent.com