The main existing solutions for facial recognition systems

Submitted by :

Yessine Barkia

Roaya Neffeti

Table of contents

I-Algorithms

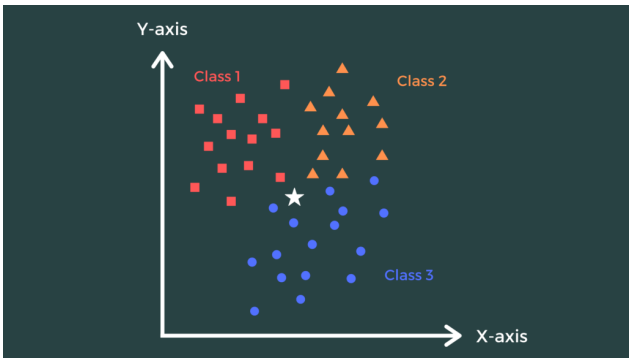
II-Tools

Face Recognition Technology has three stages:

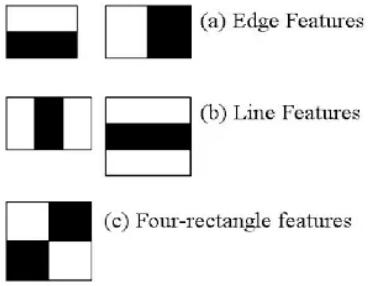
- [ ] **Detection** — the process of finding a face in an image. This can be done by training an algorithm on a vast number of photos that have faces in known positions.
- [ ] **Analysis** (attribution) is the step that maps faces often by measuring the distance between nodal points which includes the relationship between the eyes, nose, eyebrows, and the rest of the facial features. These measurements are then transcribed into a single-coded model, which can then be compared and potentially combined with known photographs within a database.
- [ ] **Recognition** is the attempt to confirm the identity of a person in a photo. It is the last step that has the goal to give the final answer to the question — Who is in this picture?

((https://medium.datadriveninvestor.com/facial-recognition-8-open-source-tools-to-detect-faces-4ec8e37bfcc6)

I-Algorithms

Algorithm	Characteristics	Advantages	Limitations
K-Nearest Neighbors (KNN) is a simple yet effective algorithm for facial recognition that uses the distance between features of two images to determine their similarity.	In KNN facial recognition, the algorithm first extracts features from the input image, such as the distance between the eyes or the angle of the nose. These features are then used to create a feature vector that represents the input image. The feature vector is compared to the feature vectors of a database of known faces to find the closest matches. The algorithm calculates the distance between the feature vectors using a distance metric, such as Euclidean distance or cosine distance, and returns the K nearest neighbors based on the distance measure. The algorithm then selects the class that appears most frequently among the K nearest neighbors as the predicted class for the input image. For example, if the K nearest neighbors of the input image are all images of the same person, then the algorithm predicts that the input image also belongs to that person.  	-simplicity and flexibility. -can work well with small datasets and can easily incorporate new data as it becomes available -can also handle non-linear relationships between features -more robust to changes in lighting, pose, and other factors	-sensitivity to the choice of K (K too small => Too specific : Overfitting/ K too large => Too general: Underfitting) -computational complexity -Require high memory as it needs to store all of the training data.

The Scale-Invariant Feature Transform (SIFT) is a computer vision algorithm used for image feature detection and description. It was proposed by David Lowe in 2004 and is considered to be one of the most robust algorithms for detecting and matching image features. (https://medium.com/data-breach/introduction-to-sift-scale-invariant-feature-transform-65d7f3a72d40)	There are mainly four steps involved in the SIFT algorithm. *) Scale-space peak selection: Potential location for finding features. *) Keypoint Localization: Accurately locating the feature keypoints. *) Orientation Assignment: Assigning orientation to keypoints. *) Keypoint descriptor: Describing the keypoints as a high dimensional vector. *) Keypoint Matching	-Locality: features are local, so robust to occlusion and clutter -Quantity: many features can be generated for even small objects -Good recall rates.	-Patent protected -computationally expensive, especially when processing large datasets -Limited to detecting specific types of features: -sensitive to changes in lighting and contrast.

Haar cascade Classifier : Object Detection using Haar feature-based cascade classifiers is an effective object detection method proposed by Paul Viola and Michael Jones in their paper, "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001. (https://medium.com/analytics-vidhya/haar-cascades-explained-38210e57970d)	Haar cascades is a machine learning algorithm that detects objects in images using a trained function from positive and negative images. *) Calculating Haar Features: summing pixel intensities in regions. Here are some examples of Haar features:  *) Creating Integral Images : essentially speed up the calculation of these Haar features. *) Adaboost Training: Adaboost essentially chooses the best features and trains the classifiers to use them. => Based on this prediction, the classifier either decides to indicate an object was found (positive) or move on to the next region (negative).	-Fast and efficient -High accuracy if trained with a large dataset -robust to variations in lighting and orientation	-Sensitive to the quality of training data -Limited to detecting specific types of objects (distinctives features such as faces, eyes and cars) --High computational requirements
Eigenfaces is a computer vision algorithm used for face recognition. It was developed by Sirovich and Kirby in 1991 and is based on the concept of Principal Component Analysis (PCA).	1. **Data collection**: The algorithm requires a dataset of face images to be used for training. 2. **Preprocessing**: The algorithm preprocesses the images by normalizing their size and converting them to grayscale. 3. **Feature extraction**: The algorithm extracts features from the preprocessed images by applying PCA to the image matrix. The resulting eigenvectors are called eigenfaces, which represent the principal components of the face images. 4. **Classification**: The algorithm uses the eigenfaces to classify new face images by projecting them onto the eigenspace and comparing the resulting coefficients to those of the training images.	-High accuracy -Robustness to variations in facial expression -Efficient computation	-Limited to recognizing faces in controlled environments -Sensitivity to variations in pose -Vulnerability to spoofing attacks (fake face image)

Fisherfaces is a computer vision algorithm used for face recognition. It was proposed by Belhumeur, Hespanha, and Kriegman in 1997 and is based on the concept of **Linear Discriminant Analysis** (LDA).	Fisherface is a technique similar to Eigenfaces but it is geared to improve clustering of classes. While Eigenfaces relies on PCA, Fischer faces relies on LDA (aka Fischer's LDA) for dimensionality reduction. The FLDA maximizes the ratio of between-class scatter to that of within-class scatter. It reduces the noise and captures the most discriminative features of the input data, and represents them in a low-dimensional space that maximizes the separability between classes. The resulting features can then be used for classification tasks	-High accuracy -Robustness to variations in lighting and facial expressions -Ability to handle multiple classes	-Sensitivity to variations in pose -Computationally expensive -Vulnerability to overfitting (the model may fit too closely to the training data)

CNN :

The field of facial recognition and identification didn't really take off until developers stopped trying to design the perfect matching algorithm and instead embraced deep learning—the use of a mathematical model that mimics the brain's neural networks to build understanding from existing data. Deep learning-based approaches such as **Convolutional Neural Networks** (**CNNs**) have achieved state-of-the-art performance in face recognition.

Characteristics:

- a type of artificial neural network commonly used in image recognition and processing tasks.
- CNNs are designed to automatically learn and extract relevant features from images, making them highly effective at tasks such as image classification, object detection, and segmentation.
- A CNN consists of multiple layers, each of which performs a specific operation on the input data.

The basic layers of a CNN are:

- **Input Layers**

- **Hidden Layers:**

    - Convolutional layers: apply a convolution operation to the input. This passes the information on to the next layer.

    - Pooling combines the outputs of clusters of neurons into a single neuron in the next layer.

    - Activation Layers: These layers apply an activation function to the output of the convolutional and pooling layers. (Most commonly used function: Rectified Linear unit, ReLU: **f(z)=Max(0,z)**)

    - Fully connected layers connect every neuron in one layer to every neuron in the next layer.

- **Output Layers**

Advantages:

- High Accuracy: CNNs can achieve high accuracy in image recognition and classification tasks.
- Robustness to Variations: CNNs are able to recognize objects in images even if they are rotated, scaled, or partially occluded.
- Automated Feature Extraction: CNNs can automatically learn and extract relevant features from images, without requiring manual feature engineering.
- Transfer Learning: CNNs can be pre-trained on large datasets and then fine-tuned for specific tasks, making them highly efficient in real-world applications.

Limitations:

- High Computational Requirements: CNNs require significant computational resources to train and run, making them unsuitable for some low-resource applications.
- Large Datasets: CNNs require large datasets for effective training, which can be difficult and expensive to obtain in some cases.
- Limited Interpretability: The complex nature of CNNs makes it difficult to interpret how they make decisions or which features are being used for classification.
    -

It would be ideal, obviously, to use CNN as our algorithm of choice in our facial recognition system.

⟶However, and fortunately, it is not necessary to develop a model from scratch using these convoluted algorithms. That is because several tools that provide pre-built models are available, which can save time and effort in the development process. These pre-built models have already been trained on large datasets and are ready to be used for specific tasks, allowing developers to focus on the application of these models rather than the model construction itself.

II-Tools

Description: A commercial-grade emotion analysis, face detection and recognition engine provided as a public API. Kairos takes the complexity out of facial recognition and emotion analysis so the focus is left on building a great product.

Features :

- Detect faces and face features within images
- Match faces against galleries of previously enrolled faces
- Analyze emotion of faces in still images and video
- Can estimate gender , age and emotions
- Can be used for a wide range of applications, including identity verification, access control, and personalized marketing.

Advantages :

secure **speedy** **scalable**

- Easy to use and integrate into applications
- There’s no need to build your own face database or understand complicated statistical algorithms
- Can process images and videos in real time
- Offer high-level security and encryption to protect user data

Limits :

- Sensitive to occlusions and variations in facial appearance, such as changes in lighting or facial expressions.
- May have difficulty recognizing faces with certain skin tones or facial features.
- May require a stable internet connection to function optimally.
- Potential privacy concerns regarding the use of facial recognition technology

Description: Amazon Rekognition is a cloud-based facial recognition system that uses deep learning algorithms to analyze images and videos.

It becomes easier to identify objects, texts, activities, and scenes in images and videos. All we have to do is provide the AWS Rekognition API with an image or video, and it will identify people, texts, objects, etc. It also helps in detecting inappropriate content.

How it works :

Amazon Rekognition works with two KPI* sets – Amazon Rekognition Image and Amazon Rekognition Video. These KPIs help in image and video analysis

→ KPI : KPIs are metrics that are used to evaluate the performance and effectiveness of a facial recognition system. These metrics can include measures such as accuracy, precision, recall, and false positive rates

⟶ It detects labels — objects (i.e., people, cars, furniture, clothes, pets), scenes (i.e., woods, beach, a city street) or concepts (outdoors), activities (i.e., playing soccer, skating). You can detect a person in a photo or video, detect facial landmarks, expressed emotion, and save facial metadata. Besides, you can also compare a face in an image with faces detected in another image

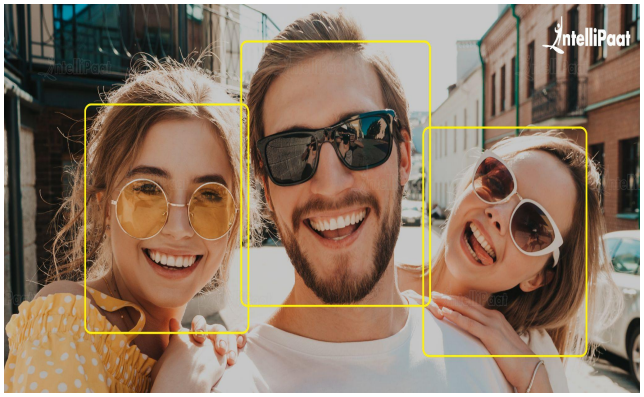
Features :

- Cloud-based facial recognition system.
- Can detect, recognize, and compare faces in images and videos.
- Can estimate age range, gender, and emotions.
- Can recognize celebrities and detect text in images and videos.
- Uses deep learning algorithms to analyze facial features.
- Can be used for a variety of applications, including security, marketing, and content moderation.

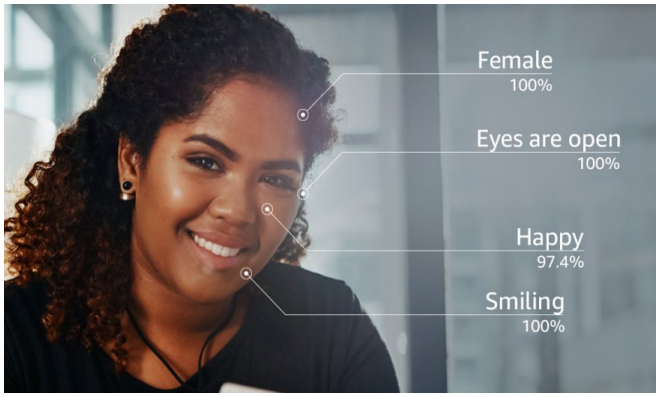
Use cases :

Feature	Use cases
Face verification and search	Applications can use AWS Rekognition to verify a user's identity. This can be done easily by comparing the live image with the reference image.AWS Rekognition has an index of faces called a face collection. We can search videos and images for a particular face from this index.
Face analysis and detection	Applications can use AWS Rekognition to detect the sentiment in an image or video. It can detect if the subject is happy, sad, crying, etc. The gender recognition feature also helps in gathering demographic details.
Content moderation	AWS Rekognition eases the detection of unwanted content. If there is some violent or adult content in a video or image, it can easily be detected by the API. It can flag the unsafe content and send a hierarchical label list with a set score to indicate the particular category the content belongs to.
Labels	Rekognition can be used to search for various objects and scenes in the images and videos stored on a device or cloud
Workplace safety	you can figure out images from your on-premises system devices (IoT sensors, cameras) at scale to automatically detect if persons in images are wearing Personal Protective Equipment (PPE) such as hand covers (gloves), face covers (face masks), and headcovers (helmets) and whether the protective equipment covers the corresponding body part (nose for face covers, head for head covers, and hands for hand covers)

Face verification and search



Face analysis and detection



Labels

Workplace safety

Advantages :

- Scalability and flexibility in processing large amounts of data, it can analyze a large number of images and in turn create a huge database of visual data
- High accuracy in facial recognition
- Image and video analysis based on Deep-learning
  Amazon Rekognition uses Deep-learning to accurately interpret images, compare and find faces in an image, and recognize scenes and objects in images and videos
- Wide range of features and functionalities

Limits :

- Potential biases and inaccuracies in the recognition of faces with certain skin tones or facial features
- Limited accuracy when faces are partially occluded or in low-light conditions

---

Description : TensorFlow is an open source software library created by Google that is used to implement machine learning and deep learning systems,These two names contain a series of powerful algorithms that share a common challenge , to allow a computer to learn how to automatically spot complex patterns and/or to make best possible decisions

How it works:

TensorFlow can be used for facial recognition by training a neural network model on a large dataset of images of faces. During training, the model learns to identify the unique features and characteristics of individual faces, which it can then use to classify and recognize new faces.

Training : collect a dataset of labeled images of faces,TensorFlow is used to build and train a neural network model, using techniques such as convolutional neural networks (CNNs) and transfer learning to improve the accuracy and efficiency of the model.

→ Once the model has been trained, it can be used to recognize new faces by inputting an image and comparing it to the patterns and features that the model has learned during training

Features :

- Open source
- It provides a range of tools and APIs for building and training machine learning models, including neural networks
- High scalability of computation across machines and huge data sets
- Efficiently works with mathematical expressions involving multi-dimensional arrays

Advantages :

- Flexible and customizable
- Very high accuracy
- Efficiency and scalability
- Chart generation

→ chart generation : refers to the process of creating visual representations of data in the form of charts or graphs

Limits :

- Model training requires a lot of time
- Not very well optimized
- Requires many resources
- Hard to interpret

Description : FaceNet provides a unique architecture for performing tasks like face recognition, verification and clustering. It uses deep convolutional networks along with triplet loss to achieve state of the art accuracy.

Features:

- FaceNet is a deep learning neural network model that is specifically designed for face recognition tasks
- The model is trained on a large dataset of face images and can generalize well to new faces and environments

Advantages:

- Highly accurate and efficient for facial recognition tasks , outperforming many traditional face recognition methods
- It can handle large datasets and can identify faces even in low-resolution images or under challenging lighting conditions
- Flexible and adaptable to different use cases
- The feature vectors generated by FaceNet are highly informative and can be used for other computer vision tasks, such as face clustering or image retrieval

→ Feature vector : refers to a high-dimensional vector representation of a face image that has been processed by the neural network

Limits:

- requires large amounts of labeled data for training, which can be time-consuming and expensive to collect
- The model can be sensitive to variations in facial expression, pose, and occlusion, which can affect its accuracy
- Unable to find good matches if the image is edited with words in front of it.

Description: OpenCV is a cross-platform library with which we can develop real-time computer vision applications. It mainly focuses on image processing, video capture and analysis including features like face detection and object detection

Features :

- Read and write images
- It includes over 2500 optimized algorithms that can be used to perform various image and video processing tasks.
- Process images (filter, transform)
- Perform feature detection
- Detect specific objects such as faces, eyes, cars, in the videos or images.
- Analyze the video, i.e., estimate the motion in it, subtract the background, and track objects in it

Advantages :

- Highly optimized for performance with real time image processing
- Multitude of implemented algorithms
- 30+ frame per second
- High accuracy

Limits :

- It may not always provide the most accurate results, especially in challenging environments or with low-quality images.
- OpenCV is not a complete end-to-end solution and may require integration with other tools and frameworks to build a complete application
- May have memory leaks