Enhanced Visible Light Facial Recognition Using Deep Learning

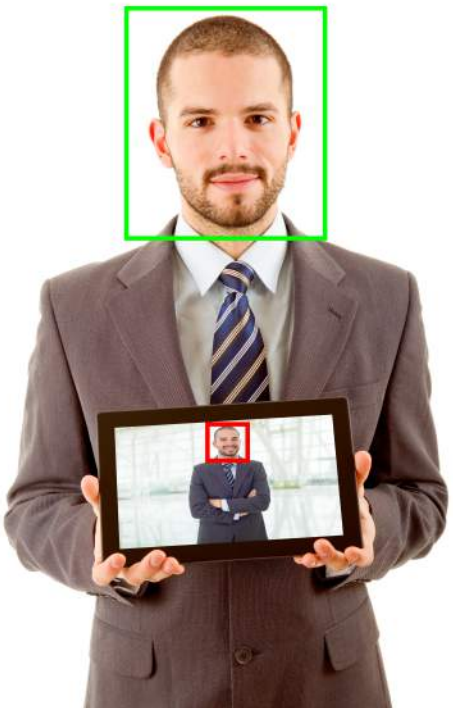
Facial recognition technology with all-rounded update

Enhanced Visible Light Facial Recognition is the latest technology of ZKTeco, aimed to overcome shortcomings of previous facial recognition and achieve an all-rounded upgrade on whatever performance and reliability by using Deep Learning. Facial Recognition has been considered as one of the rising biometric technologies and many are optimistic about its development in the future. However, the technological incapability of facial recognition has also been exposed along with the increasing usage in our daily life, mainly in the ability of anti-spoofing, threshold of processing power, environmental restriction on light source and pose of angle.

Therefore, this is precisely what the Enhanced Visible Light Facial Recognition developed for and the key to this breakthrough would be incorporating Deep Learning technology. Deep Learning is adapted to fill up those three technological incapabilities as aforesaid and has been passing through a continuous training to keep enhancing its performance.

Maximize security level by reliable anti-spoofing

Anti-spoofing is one of the most important task for all kinds of biometric technologies including facial recognition. Common spoofing attacks can be done by displaying photos or video with human faces on the camera and a liveness detection utilizing Deep Learning was introduced against it. Liveness detection will take no more than 1 second to determine whether it is a spoofing attack or not, benefited from the power of Deep Learning. Deep Learning is trained up to conduct liveness detection by learning differences such as lighting, texture and resolution between living human faces and human faces from photos or videos.



Streamlined algorithm brings you lightning-fast experience

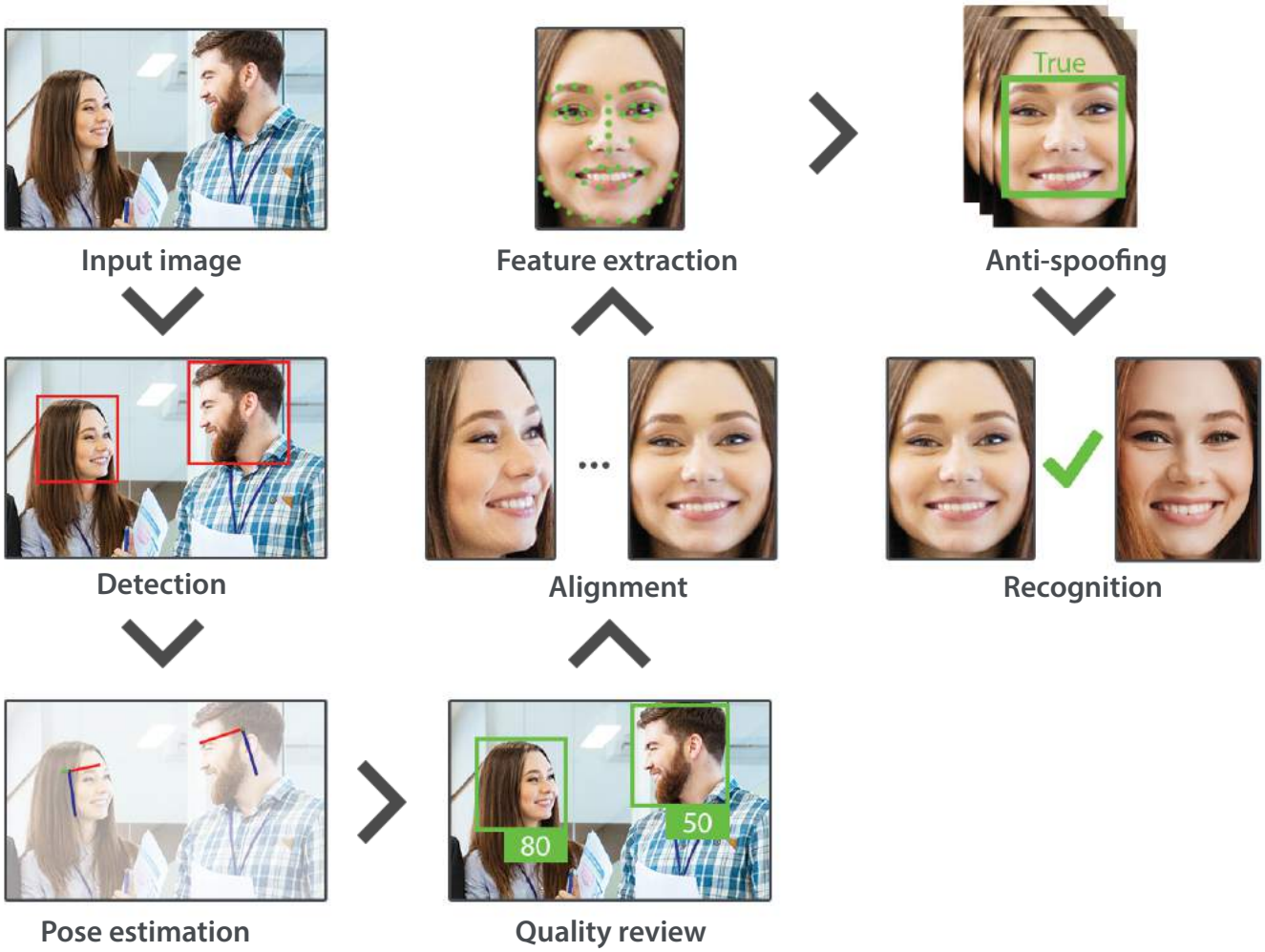
Sparse Representation-based Classification (SRC) is used to streamline to recognition process and lower threshold of processing power which facilitate the broad applications of facial recognition due to lower entry barrier on hardware requirement. SRC utilized the mathematical linear characteristics of image pixel and features extraction via Deep Learning to conducts the classification in a more efficient way by minimizing the size of processing data while not only maintaining the recognition accuracy at a high level but also improving user experience which allows user to complete recognition with less than 1 second by just a glance.

Unshaken performance under complex environment

Facial recognition is usually conducted under a dynamic environment where light source, pose of angle and camera distance can be varied. It has been a classical problem for facial recognition along these year until the introduction of Deep Learning technology. Deep learning technology allows system to extract and filter features by itself throughout the training for learning the differences of a human face under various distance, pose of angle and light source, so when it comes to an unfavorable image which distance, lighting and pose of angle are less than ideal, the recognition can refine its algorithm to moderate the adverse effect on recognition accuracy. Official figure shows that the Enhanced Visible Light Facial Recognition can adapt to a wide range of unfavorable lighting environment from 0 to 50,000 lux.

Step through the Enhanced Visible Light Facial Recognition

Traditional facial recognition will go through 4 main stages including detection, alignment, feature extraction and recognition. The Enhanced Visible Light Facial Recognition has newly incorporated multiple stages in order to complement the technology.

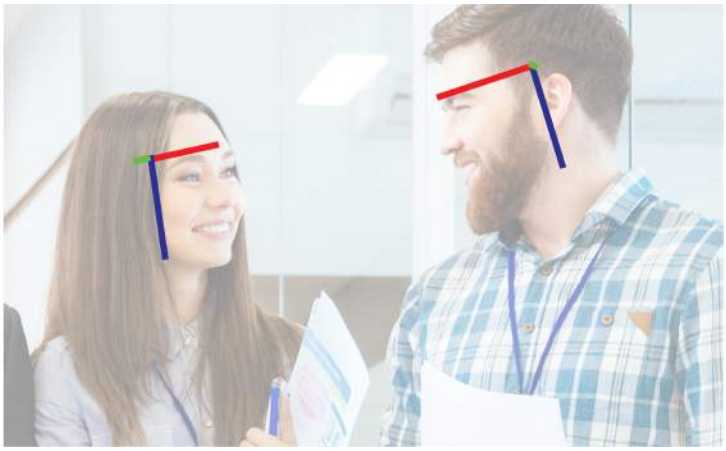


1. Detection

System will firstly identify if there is a human face in the image or video and ignore anything else. During detection, Deep Learning with high accuracy and real-time detection ability has given play to its functions of detecting human faces in different size as well as being more adaptable when dealing with backlight.

2. Pose Estimation

Estimation model is developed via Deep Learning for precisely identify the pose angle such as yaw, pitch and roll. Learning the precise pose angle can benefit subsequent moderation on angular distance.



3. Quality Review

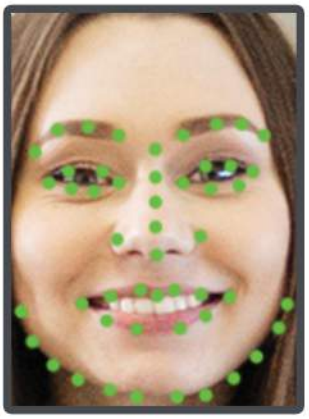
After estimation of the pose angle of input faces, the image will undergo a quality review in 3 dimensions of brightness, clarity and pose angle to ensure quality of images falls into an acceptance range and screen out images which classified as unrecognizable for preventing any waste of processing power.

4. Alignment

Face alignment refers to the positional adjustment of a person's eyes, nose and mouth to the indicated positions. This process requires 2D affine transformation including movement, zooming, opposition and rotation. The aligned face images are more efficient to be verified by verification algorithms.

5. Feature Extraction

Face alignment refers to the positional adjustment of a person's eyes, nose and mouth to the indicated positions. This process requires 2D affine transformation including movement, zooming, opposition and rotation. The aligned face images are more efficient to be verified by verification algorithms.



6. Anti-spoofing

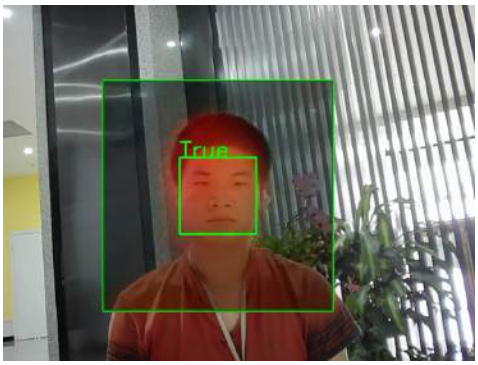
A liveness detection has been incorporated in order to further enhance the ability of Anti-spoofing. Liveness detection can be separated in whether individual's cooperation is needed.

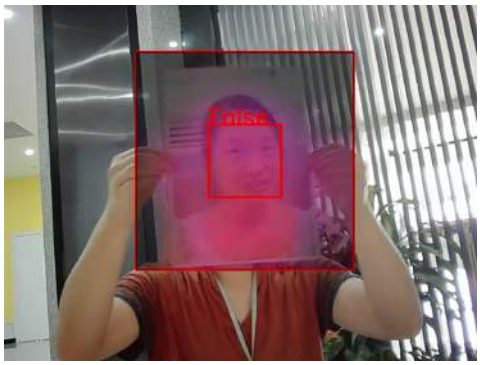
Passive liveness detection, refers to detection requires no individual's cooperation, will send out instructions such as blink through devices and the system will capture the movement to identify whether user has deliver as instructed or not.

Active liveness detection, refers to detection requires individual's cooperation, is developed against common spoofing attacks by displaying photos or video with human faces.

- **Spoofing attack through photos**

  Anti-spoofing on photo display focuses on detecting movement of human faces. Face within a photo usually move along with the entire photo including the photo background. However, movement of a real human faces and the background behind are usually inconsistence.
  With this assumption, identification via Deep Learning can be done through judging the optical flow and semantic segmentation. Facial recognition camera will first capture a series of video sequence and sample multiple frames, then both real living human face and face with photo will be separated from background by semantic segmentation and analyze the association among pixels movement over time, namely optical flow, so liveness can be detected eventually by comparing results.





- **Spoofing attack through videos**

  Anti-spoofing on photo display is done by capturing partial facial features then analyze via Deep Learning in details. Deep Learning was trained to identify the differences between a living human face and human face played in a video in terms of texture and resolution. And it is also capable to recognize the association between a human face and objects surrounding, for example, video and its player (smart phone, tablet) will be captured together, this association of human faces showed within a player will be recognized by system as a spoofing.

7. Recognition

Recognition is the final process which classifies the detected, aligned and normalized faces into known identities by crosschecking the extracted features on templates. Recognition is usually done for training purpose or identification task, different approaches or techniques will be applied either way:

- **Training purpose**

  By inputting training dataset and corresponding face anchor template, Deep Learning will started to learn how human face is by restructuring human faces from the smallest pixels of the photos to form curves, edges, shapes and finally define faces. Moreover, comparing training dataset which may vary in pose angle, distance and lighting allows the system to recognize the variation between so that it can explore the way to widen the acceptance range on angular distance and make augmentation be more precise and effective.

- **Recognition task**

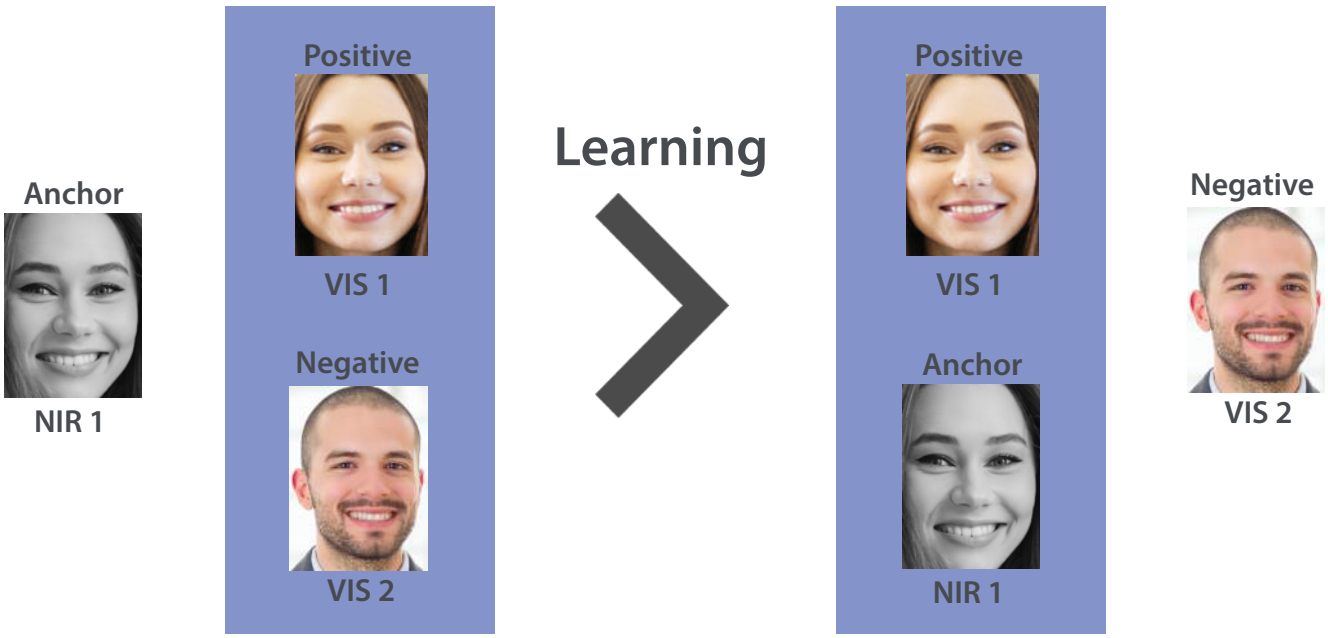
  Recognition task can be generally divided into two purposes, namely, 1:1 verification and 1:N comparison. The actual recognition on one's identity refers to the process of pushing the received data through classifiers. Classifiers or the way of classification will significantly affect the recognition quality and the processing time. Sparse Representation-based Classification (SRC) is applied in order to increase the efficiency by minimizing the size of processing data which may results in shortening processing time and lowering threshold of processing power.

Performance on practical applications

The Enhanced Visible Light Facial Recognition is rather an all-rounded upgraded version on previous facial recognition technology, not only qualified for general applications, it is also competent for advanced recognition task under a dynamic environment. Especially for its cutting edge anti-spoofing function, any business with high level of security needs will only find it a more reliable technology over the rest.

Future development on Visible Light Facial Recognition

Multi-model facial recognition based on Near-infrared (NIR) and Visible light (VIS) is currently under development. There are two type of image input utilized for having stronger compatibility on both NIR and VIS templates. The Deep Learning training model picks the NIR image as anchor and will be given a set of 2 VIS images including one positive sample and one negative sample which refers to the same person's image and false person's image respectively. The way that picking the NIR image as anchor and triplet loss to train Deep Learning model, will have the model to be further closing up the gap between anchor and positive sample and enlarger distance between anchor and negative positive by not only telling the model what is correct but also simply showing what is wrong.



V1.2 2018.07.18