CS 6384 Computer Vision 18S

Hao WAN (hxw161730)

Project2: Wink detection and Shush detection

Approaches Explanation

In this project we follow the object detection method proposed by Paul Viola and Michael Jones, using several Haar feature-based cascade classifiers to detect faces, eyes and mouths. The classifiers are already trained and given, our input is as well determined – live video on the face of ourselves and two folders of static images. Our goal is to obtain a best detection result on the given input. To achieve this goal, many approaches are used. Some approaches are explained as the following:

1. Detect the face(s) firstly, then try to find the eyes and mouth within the range of face. This will make the programs works more accurately and faster.

2. Choose an appropriate classifier (or appropriate classifiers) to get a maximum ratio of successful detection.

There is no perfect single classifier for all circumstances. But our input source is determined, so we can try different classifiers and choose a better one.

After several tests, for part I of this project, I use "haarcascade_frontalface_alt.xml" and "haarcascade_eye.xml" to do the first stage of detection; and for part II, I use "haarcascade_frontalface_default.xml" and "Mouth.xml".

These classifiers can make sure a relatively high recall on the corresponding input source - most of the images do get detections (but some of the detections may be false positives, so we need further optimizations).

3. Fine-tune the parameters of a classifier to reduce the number of false positive detections.

The reference goes to:
https://stackoverflow.com/questions/20801015/recommended-values-for-opencv-detectmultiscale-parameters

scaleFactor: I use 1.1 and 1.15 as the scaleFactor. The lower the value is, the slower the detections would be, but we also could have a higher sensitivity of the detection (at the same time a higher rate of false positive).

minNeighbors: Higher value results in less detection but higher accuracy.

minSize: This is the smallest size we want to detect. For some images, the mouths and eyes appear too small, in this case we should diminish this value or we may miss the small features.

4. Use a combination of two classifiers to detect objects in some delicate images.

For example, for detecting eyes with glasses, if we use the classifier "haarcascade_eye_tree_eyeglasses.xml" directly, we could detect the eyes with glasses, but the rate of false positive would be very high at the same time. So firstly, I put the normal eye classifier "haarcascade_eye.xml", using a relatively high value of minNeighbors to accurately detect normal eyes without many false positives. Some eyes with glasses may not be detected in this situation, then I apply the eyeglasses classifier again to those images so that the eyes with glasses could be detected. This approach works better for a single face, or for multiple faces entirely with or entirely without glasses, because I use an if statement (if there is no eye detected) to decide if we should go into the eyeglasses classifier.

5. Pre-process the input images.

This includes the color processing and the resizing.

For the shush part I simply did a histogram equalization to enhance the contrast of the images, for the wink part, because in the input folder there exist several images with bad brightness, I did some transformations to enhance the brightness of the images.

The size sometimes could have a bad influence on the performance of the detection. If the image is extremely huge, the detection would be slow and give a higher rate of false positive; if the image is too tiny, then sometimes the size of certain mouth or eyes would be below the minSize, and could not be detected.

6. Apply some common senses to filter out some false detections and make the detections more accurate.

By observing the input images, for most of the cases, we have normal, ortho faces without flip or huge rotation. So to set the ROI for eye detection, we could use the upper part of the face, and for mouth detection we could use the lower part of the face. This will largely increase the accuracy of detection.