Assignment 2: Ear Detector

Image Based Biometry, Assignment 2
Faculty of Computer and Information Science
University of Ljubljana

Jurij Nastran

I. Description

In order to complete this assignment, I decided to use the Viola-Jones classifier. The theoretical information about the classifier was obtained from [1]. The classifier was trained using the instructions given in [2] using the pictures from the folder "train" that was given to us at the start of this assignment. The trained classifier is stored in the folder "classifier" on my GitHub repository here: https://github.com/NastranJurij/ear_detect. There one can also find the instructions regarding the environment as well as the datasets.

II. Results

Sadly the resulting classifier is not optimal, even for Viola-Jones classifiers. I estimate that more training data would be needed to greatly improve the classifier. In order to replicate my results, please follow instructions on the GitHub repository linked above.

A. Metrics

When measuring the performance of my classifier, I tested it on the 250 480 × 360 pixel colour images from the already given dataset in the folder "test". The classifier returned a list of rectangles which were immediately converted to a mask-like image. This image was compared with the ground truth images provided in the folder "testannot_rect" pixel by pixel. This allowed me to store the values TruePositive(TP), FalsePositive(FP), TrueNegative(TN) and FalseNegative(FN) for each of the 250 classified images and calculate the Precision, Recall and F1 scores after.

When optimal parameters for classification using the chosen classifier were found, the following results were obtained:

Table I: The final results of the classification

TP-mean:	1216.916
FP-mean:	3650.532
TN-mean:	166317.876
FN-mean:	1614.676
F1-mean:	0.3207
precision-mean:	0.2984
recall-mean:	0.4429

Please note, that the values like TP-mean are unexpectedly high because of the fact that for each image I obtained the final measures using pixel by pixel comparisons between my classified image and correctly classified image.

B. Analysis of results

When choosing the optimal parameters of classification such as "scaleFactor" and "minNeighbours", these two parameters were chosen in such a way as to maximize the obtained F1 score. In spite of this, the best obtained F1 score was 0.3207 or 32.07% which is not good enough to be useful over modern Viola-Jones classifiers trained for this task and especially not good enough to compete with the state of the art deep convolutional networks that are often used for such and similar tasks. The best F1 score was obtained when classifying with the parameters set to: scaleFactor= 1.01611 and minNeighbours= 12, which is a higher value than generally recommended. A better understanding of the failing of the classifier can be achieved by looking at the test results in the form of the images. A sample of test results is shown below:

C. Examples

The classifier performs especially well when the background behind the person is monotonous as shown on **Figure 1**

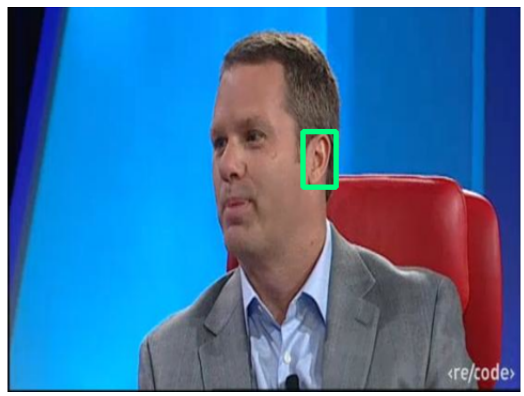

Figure 1: Picture 0049.png - Good performance in the case of monotonous backgrounds.

This is often not the case with the test images. Many have backgrounds with sections which are falsely classified by the classifier as an ear leading to a larger number of false positives. See **Figure 2** for a demonstration of the background contributing to the false positive statistic.

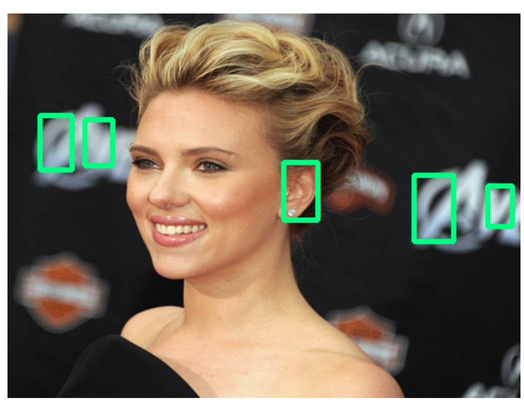

Figure 2: Picture 0209.png - The effect of non-monotonous backgrounds.

The background is shown to be a major contributor of false positives, but it cannot contribute to the false negatives. False negatives are often a result of relatively poor contrast of the ear relative to the surrounding image. This effect seems to be more prevalent among pictures of people of colour as seen on **Figure 3**. It is a major contributor to the false negative statistic.

Figure 3: Picture 0211.png - The effect of poor contrast around the ear.

Lastly many mistakes indicate that the classifier is simply undertrained meaning it would need to be trained using many more images to obtain better results in the case when the ears are seen at a somewhat different angle as is the case in **Figure 4**.

Figure 4: Picture 0017.png - Many mistakes could be repared with more training data.

In conclusion the Viola-Jones classifier can reach much better results as was shown here, but it would need to be trained on more training examples showing many more noisy backgrounds as well as many ears in different positions and lightning conditions in order for the trained classifier to be more robust.

References

[1] A. K. Jain, A. A. Ross, and K. Nandakumar, *Introduction to biometrics*. Springer Science & Business Media, 2011.

[2] "Cascade classifier training," https://docs.opencv.org/3.4/dc/d88/tutorial_traincascade.html, last accessed online: 06.12.2020.