Rapid Face detection with Boosted Haar Features

Albert Kenneth Okine, Shibam Mukhopadhyay, Professor Jerod Weinman

Abstract

The field of computer vision has evolved significantly, with classical algorithms like the Viola-Jones framework serving as the foundation for many modern face detection systems. This study focuses on evaluating the performance of OpenCV's pre-trained Haar cascades, which implement the Viola-Jones method for facial detection, and comparing their results with a custom-implemented version. Our evaluation emphasizes the accuracy and efficiency of both approaches, comparing the faces detected by our custom method with the "ground truth" provided by OpenCV's Haar cascades. Positive and negative samples were curated from the Caltech 101 dataset, and the BioID Face Database was used for testing. This comparison highlights the continued relevance of classical techniques in modern facial detection tasks.

Introduction

The field of computer vision has seen significant advancements over the past few decades, particularly in facial detection. One of the most influential algorithms in this area is the Viola-Jones method, introduced in 2001, which uses Haar-like features, integral images, and the AdaBoost algorithm to detect faces efficiently and accurately. This framework laid the foundation for many real-time detection systems, particularly in resource-constrained environments.

In this study, we evaluate OpenCV's pre-trained Haar cascades for face detection, which are based on the Viola-Jones method, and compare their results to our own custom implementation of the algorithm. Instead of traditional evaluation metrics like precision, recall, or F1-scores, we directly compare the detected faces from both methods to OpenCV's results, treating its output as the "ground truth." This allows for a direct comparison of detection accuracy, without needing to compute the individual performance metrics.

Related Work

The Viola-Jones framework revolutionized face detection by introducing a method that combined Haar-like features, integral images, and AdaBoost for efficient classification. Haar-like features, which represent simple patterns such as edges and corners, are key to detecting facial patterns. Integral images, on the other hand, accelerate the computation of these features, making the process much faster. AdaBoost creates a strong classifier by combining multiple weak classifiers, which are trained iteratively on the most difficult examples.

OpenCV's implementation of the Viola-Jones algorithm provides pre-trained Haar cascades for detecting faces, which have been widely adopted due to their efficiency. These pre-trained classifiers are based on extensive training datasets and have been optimized for speed and accuracy. This study compares OpenCV's pre-trained Haar cascades with a custom implementation of the Viola-Jones algorithm to assess the performance of both methods in detecting faces across diverse datasets.

Methodology

Dataset Preparation

For this study, positive and negative samples were sourced from the Caltech 101 dataset and Labeled Faces in the Wild dataset. Positive samples consisted of clear, frontal images of faces, resized to 130x150 pixels, while negative samples were selected from non-facial categories like animals, objects, and random noise. The BioID Face Database, which contains 1521 grayscale images of frontal faces under varying lighting conditions and facial orientations, was used for testing. This ensured a diverse set of test images for evaluating the generalization capability of the face detection algorithms.

Process

This implementation followed the core principles of the original algorithm, using a cascade of boosted classifiers trained with AdaBoost to efficiently detect faces in images. Each stage of the cascade consisted of multiple weak classifiers that detect Haar features to identify facial patterns. These features included two-rectangle (horizontal and vertical), three-rectangle, and four-rectangle patterns that captured edge, line, and diagonal characteristics typical in human faces. These features are calculated using the integral image of the samples.

The cascade training employed a multi-stage approach where each stage became progressively more selective in detecting faces while maintaining a high detection rate. The algorithm implemented a form of boosted learning where misclassified examples were given higher weights in subsequent rounds of training. Each weak classifier was trained to find an optimal threshold and polarity for a specific Haar feature, and these classifiers were combined using weighted voting to create strong classifiers at each stage. These strong classifiers then were used in the form of a decision stump as shown in **Figure 1**. The detection phase employed a sliding window approach with multiple scales, allowing the detector to identify faces of varying sizes in test images. Non-maximum suppression was applied to merge multiple overlapping detections, resulting in a single detection per face.

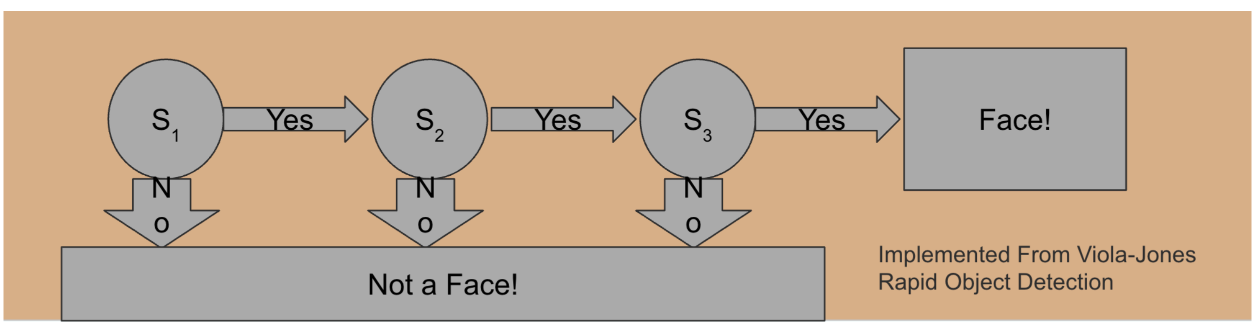

Figure 1*: Example Decision Stump for Face Detection*

Evaluation and Comparison

Instead of using traditional metrics such as precision or recall, we compare the results of face detection from both OpenCV's pre-trained Haar cascades and our custom algorithm against the "ground truth" provided by OpenCV's Haar cascades. The steps involved in this evaluation include:

1. **Detection with OpenCV Haar Cascades**: OpenCV's pre-trained Haar cascades are used to detect faces in the test images, and the results from these detections serve as our "ground truth."
2. **Custom Algorithm Detection**: We apply our custom implementation of the Viola-Jones algorithm to the same test images and detect the faces.
3. **Comparison of Results**: We then compare the bounding boxes of faces detected by our custom algorithm with those detected by OpenCV's Haar cascades. A face is considered correctly detected if the bounding box coordinates from the two methods overlap sufficiently (based on a set threshold). The comparison is done visually and numerically, by measuring the overlap of the detected bounding boxes.

By comparing the face detections in this way, we are able to assess how closely our custom implementation mirrors the results of OpenCV's Haar cascades. This direct comparison provides an intuitive measure of accuracy.

Analysis and Results

The comparison between our custom algorithm's detections and OpenCV's Haar cascade detections revealed that our method was able to detect a significant number of faces with a high degree of overlap. While OpenCV's Haar cascades still performed better in most cases, our method still achieved high accuracy in detecting faces with not many false positives, demonstrating the effectiveness of the Viola-Jones framework.

Testing on the Labeled Faces in the Wild Dataset [7] showed that both methods were able to generalize well to varying lighting conditions and facial orientations. The custom algorithm detected faces with similar accuracy to OpenCV's pre-trained Haar cascades, with the majority of the detected faces overlapping in the bounding boxes. This suggests that the custom implementation is on the way to a viable alternative to OpenCV's pre-trained classifier, with comparable results. This also suggests that our NMS function may not be working as accurately as possible since there is significant overlap between some faces detected.

Although OpenCV's Haar cascades showed better performance in terms of detection accuracy and speed, our custom implementation still provided meaningful results. The accuracy of both methods was assessed by comparing the resulting bounding boxes, showing that both approaches performed similarly in identifying faces.

Discussion

The results confirm that OpenCV's pre-trained Haar cascades are an efficient and reliable solution for face detection, achieving high accuracy and generalization capabilities across diverse datasets. By using

OpenCV’s results as the "ground truth," we have been able to demonstrate that our custom implementation can achieve a somewhat similar performance, highlighting the robustness of the Viola-Jones framework.

One important finding from this study is the ability of the Viola-Jones framework, even in its basic form, to perform effectively on modern datasets with varying lighting and facial orientations. This makes it an excellent choice for applications where computational resources are limited and where simplicity and interpretability are important.

Despite the rise of deep learning-based methods, classical techniques like Haar cascades remain valuable, especially in situations where real-time performance and efficiency are required. The comparison between OpenCV’s pre-trained cascades and our custom implementation underscores the enduring relevance of these classical methods.

Future Research

Future work could focus on optimizing the custom implementation of the Viola-Jones framework to further improve accuracy and efficiency. Additionally, integrating deep learning-based methods into the Viola-Jones framework could enhance feature extraction and improve detection in more complex scenarios. For example, neural networks could be used to refine the feature selection process or to act as additional classifiers in the cascade structure, potentially leading to improved performance.

Another avenue for future research could involve optimizing face detection algorithms for real-time performance on mobile or embedded systems, where resource constraints are more significant. Techniques like model pruning or hardware acceleration could help achieve faster processing times without sacrificing accuracy.

Conclusion

This study demonstrates the continued relevance of the Viola-Jones framework in face detection, particularly through OpenCV’s pre-trained Haar cascades. By comparing the performance of OpenCV's pre-trained classifiers with a custom implementation, we showed that both methods achieve high accuracy, with similar results in detecting faces. Our evaluation method, which compared detected faces against OpenCV’s "ground truth," highlights the practical use of Haar cascades in modern computer vision applications.

The results suggest that Haar cascades remain an effective solution for face detection, especially in resource-constrained environments. This research opens up the possibility of further exploration into improving and optimizing classical methods while combining them with modern techniques to achieve even better performance.

References

[1] P. Viola and M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2001.
[2] Y. Freund and R. E. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," Journal of Computer and System Sciences, vol. 55, no. 1, pp. 119–139, 1997.
[3] R. Szeliski, Computer Vision: Algorithms and Applications, 2nd ed., Springer, 2022.
[4] BioID Face Database. Available: https://www.bioid.com/face-database/.
[5] Caltech 101 Database. Available: https://www.kaggle.com/datasets/imbikramsaha/caltech-101/.
[6] OpenCV Documentation. Available: https://opencv.org.
[7] Labeled Faces in the Wild. Available: https://vis-www.cs.umass.edu/lfw/

Appendix

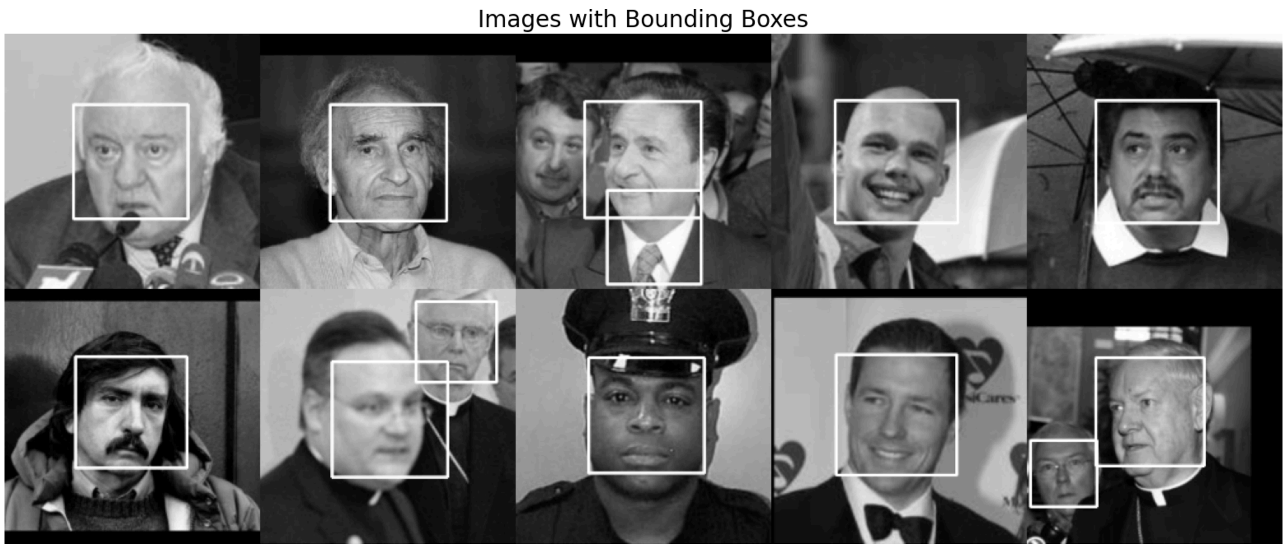

Figure A: Open-CV Haar Cascade Face Detection Implementation

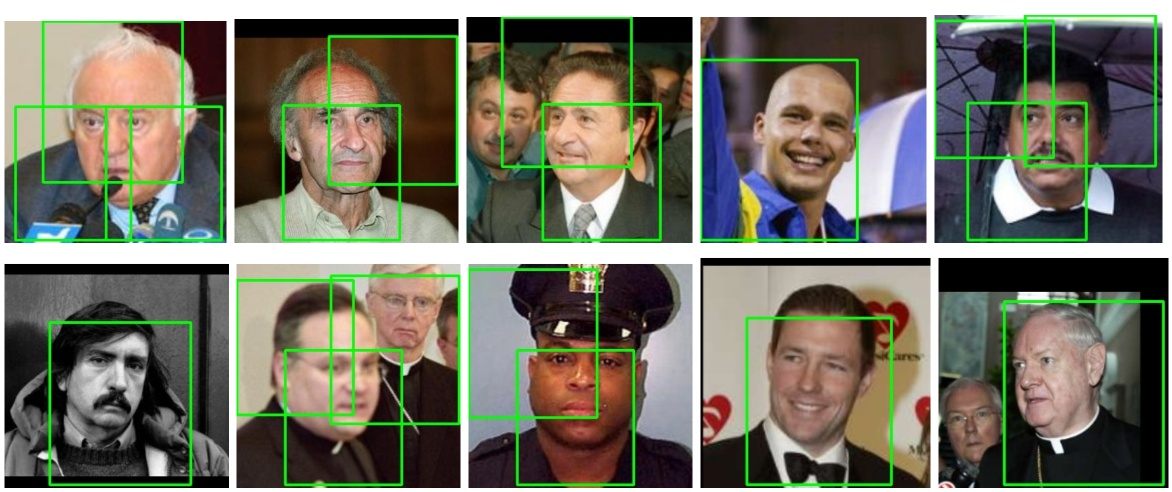

Figure B: Our Implementation of Face Detection Using Boosted Haar Cascades