Improving Open Source Face Detection by Combining an Adapted Cascade Classification Pipeline and Active Learning

Steven Puttemans[1], Can Ergun[2], Toon Goedemé[1]

steven.puttemans@kuleuven.be, can.e992@hotmail.com, toon.goedeme@kuleuven.be

Introduction

Is face detection a solved problem?

- It is as good as solved in academic context.
- However open source computer vision frameworks, like OpenCV, still provide face detection interfaces that lack high accuracies on public datasets.
- Mainly due to weakly trained detection models that are never updated.

A frequently used framework

- The cascade classification pipeline (based on the Viola&Jones algorithm) in OpenCV3.1 is still widely used to train object detector interfaces.
- However the existing face detection models have low accuracy.
- Largest issue is the existance of false positive detections, mainly if subsequent items of the pipeline require a true positive detection (e.g. face recognition).

We improve the existing framework by

- Adjusting the face annotations.
- Improving the negative training sample collection.
- Using an active learning strategy to collect hard positive and hard negative training samples to improve the detector training process.

Related work

Recent advances in computer vision

- Solve face detection using complex techniques like multi-task CNNs, CNNs combined with 3D information and R-CNNs.
- Yield very promision results, but complex to implement.
- Need for more processing power and better hardware than existing techniques.

The work of Viola & Jones

- Well established baseline for face detection used as standard in many industrial applications (e.g. digital photo cameras).
- Inside OpenCV under a BSD license → interesting for commercial use.
- Still a lot of research efficiently using this pipeline for all sorts of object detection tasks, and more specifically for industrial object detection.

Framework & dataset

OpenCV 3.1 framework using C++

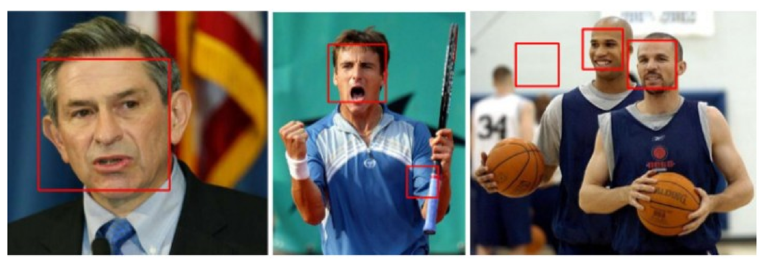
- *cv::CascadeClassifier* object detection interface.
- *opencv_traincascade* application.
- Original training data of *lbpcascade_frontalface.xml* no longer available.

Manually collected dataset for training a new model

- YouTube videos, bulk image grabber on social media, imageboards and google image search results.
- Collected data has **NO** labels of face regions available.
- Using a multi-threaded tool for efficiently using old model to look for hard negative and hard positive training samples.

Table 1: Training data overview for trained models.

Model	#pos	#neg	#stages	#stumps
OpenCVB	xxx	xxx	20	139
BoostedB	1.000	750k	26	137
IterHardPos	1.250	750k	19	146
IterHardPos+	1.500	750k	19	149



Validation on the Face Detection Data Set and Benchmark (FDDB)

- Contains 5171 face annotations in 2845 images.
- Focus of dataset = pushing limits in unconstrained face detection.
- Annotations converted to OpenCV training format.

Suggested Approach

Changing the face's region of interest

Annotations for OpenCV LBP face detector contain a lot of extra information like ears, hair, complete head and background information. We reduced the annotation to the inner face only, containing more general featres and providing robustness to both in-plane as out-of-plane rotations..

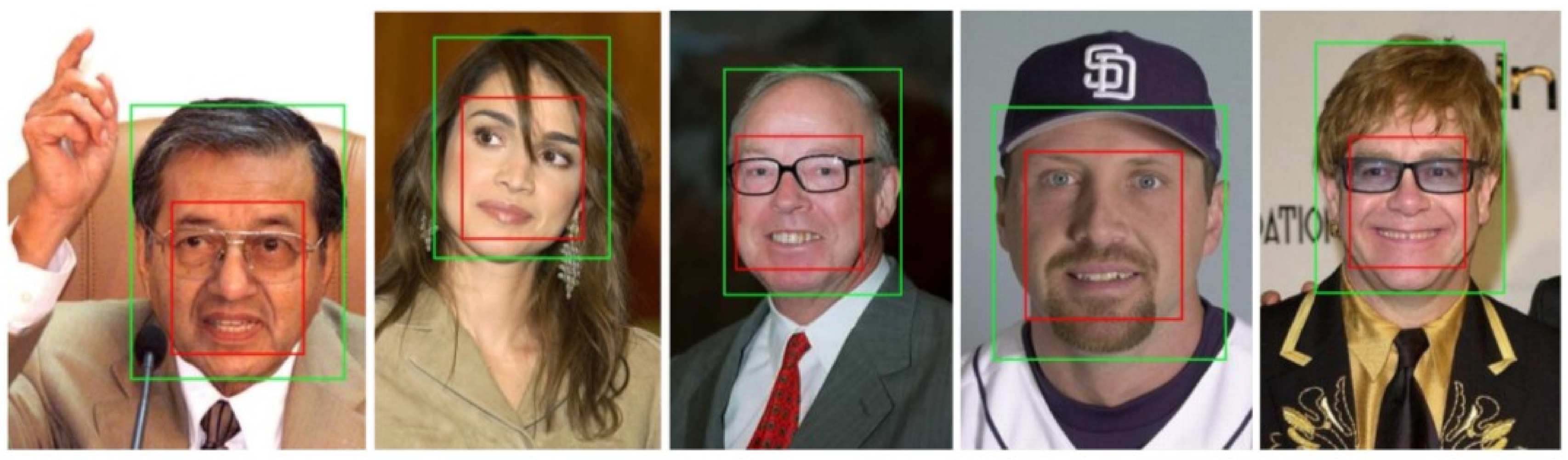
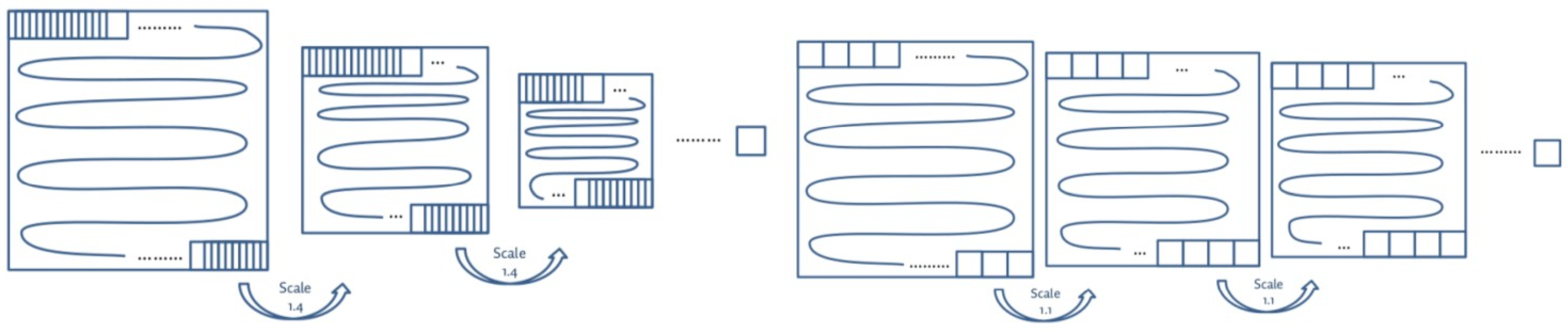


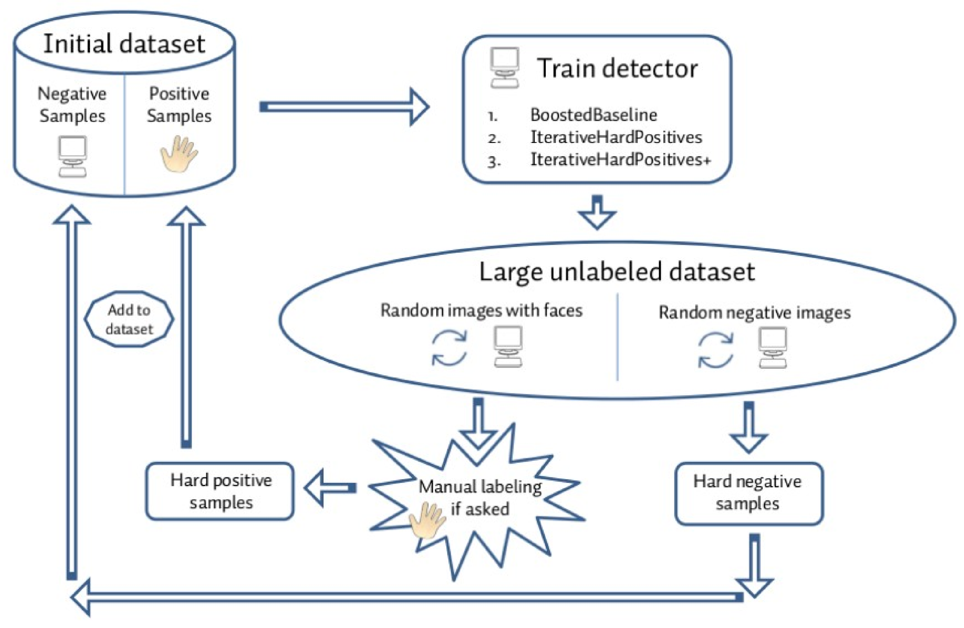
Figure 2: Changing the annotations from full-face to inner-face: (green) OpenCV (red) ours.

Adapting the negative training sample collection



(a) Original proces inside OpenCV framework. (b) Suggested adapted pipeline.

Iterative active learning strategy



- Iterative processing with input of domain expert identifying wrong classifications.
- *Hand symbol* = manual intervention/annotation
- *Computer symbol* = fully automated processing

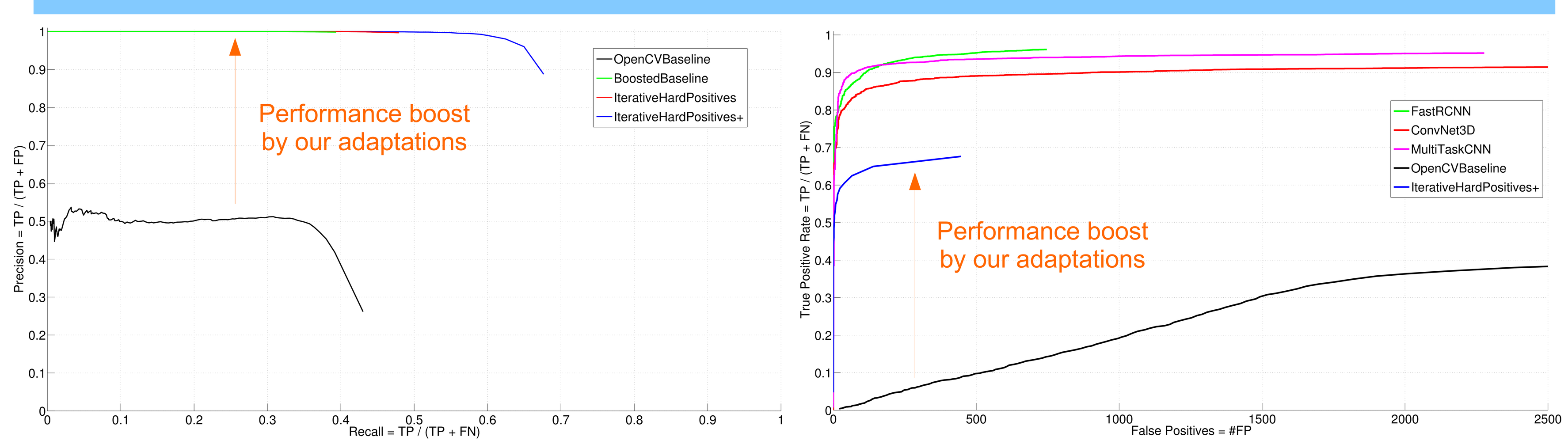
- Focus is to collect hard negative and hard positive training samples for improving the existing detection model.

Halting training if dataset is consumed

OpenCV uses pixel wise shifts in the negative training data once the negative dataset is fully processed. This generates a lot of redundant data. We allow to halt the training and finalize the model once the negative set is depleted.

Results & conclusions

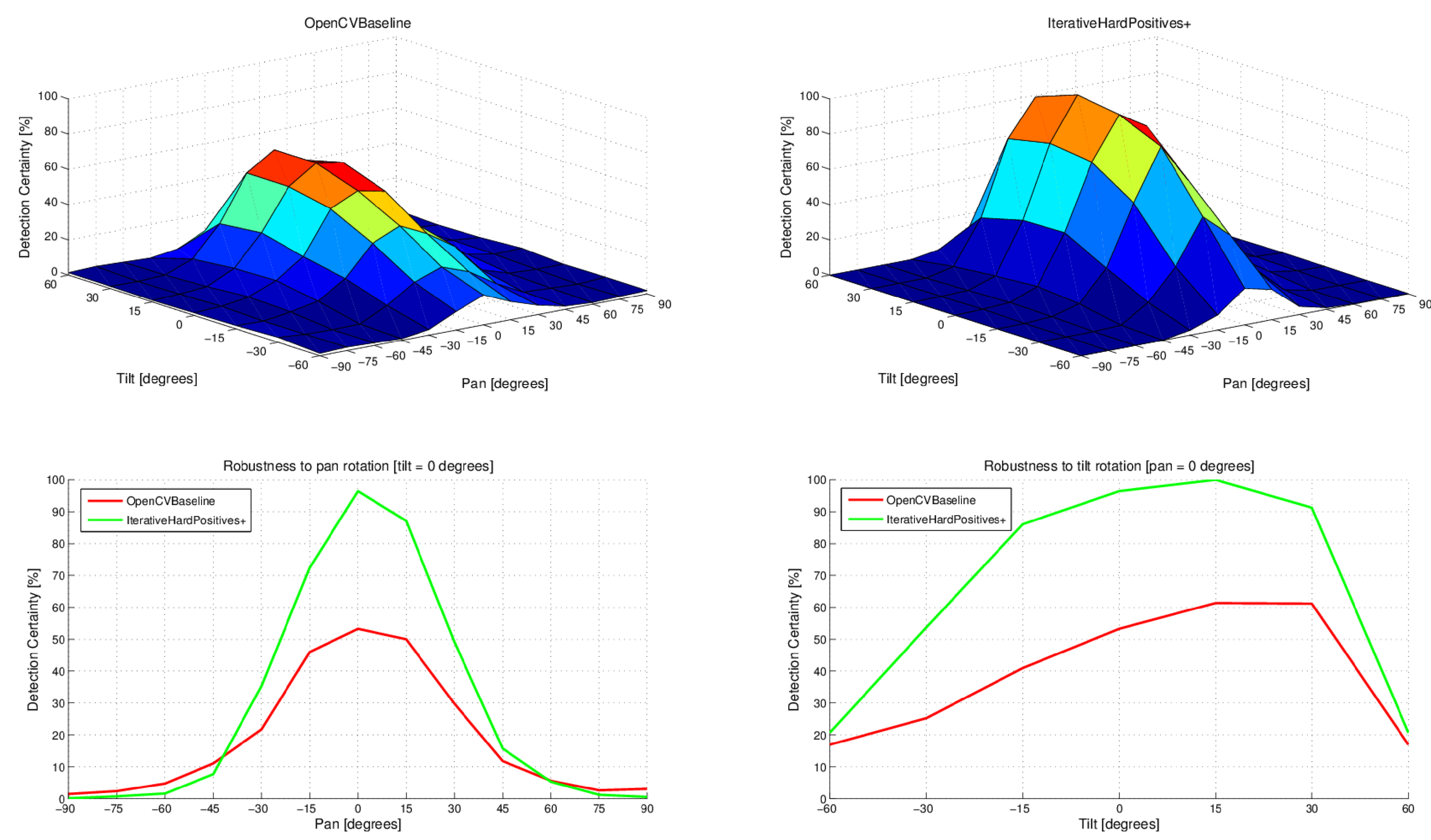
Performance of trained models



- *OpenCVBaseline* achieves a **40% recall** at a **precision of 40%**.
- Our *BoostedBaseline* detector already increases **precision** to **99.5%** at **40% recall**.
- Our *IterativeHardPositives+* detector achieves **90% precision** at **68% recall**.

- FDDB uses true positive rate versus the number of false positive detections. Our adaptations clearly pushes the accuracy of open source face detection using cascade classification closer to the state-of-the-art CNN based algorithms.

Out-of-plane rotation robustness



Conclusions

- Using our IterativeHardPositives+ detector we achieve an increase to 68% of recall, while maintaining a high precision of 90%.
- Combining multiple views (frontal and sideview) could increase overall performance on FDDB dataset.
- New functionality officially integrated in **OpenCV3.2 master** branch!

[1] EAVISE (Embedded Artificially Intelligent Vision Engineering), KU Leuven, Campus De Nayer, Sint-Katelijne-Waver, Belgium

[2] Istanbul University, Faculty of Economics, Beyazt, 34452 Fatih/Istanbul, Turkey.

EAVISE
Embedded & Artificially intelligent VISion Engineering

KATHOLIEKE UNIVERSITEIT LEUVEN