Fast Object Detection with Occlusions

Yen-Yu Lin[1] Tyng-Luh Liu[1] Chiou-Shann Fuh[2]

[1]Inst. of Information Science, Academia Sinica, Taiwan / [2]Dept. of CSIE, National Taiwan University

1. Summary

- Our goal is to develop a robust object detection system to detect faces with or without occlusions in real time.
- Particularly, we address the following problems:
  - How to derive an appropriate boosting algorithm for face detection?
  - How to represent non-face training data for reducing false positives?
  - How to handle occlusions without compromising in efficiency?
- Our contributions
  - A *soft-boost* technique combining Bhattacharyya weak learners with soft margins.
  - A *reinforcement* training procedure to reduce false-positive rates by uniformly stage-wise adding non-face training data.
  - An *evidence cascading* approach to distinguish occluded faces from non-face patterns.

2. Boosting with Soft Margins

- Notations:
  - L training data: $D=\{(\mathbf{x}_1,y_1),\ldots,(\mathbf{x}_l,y_l)\}=D^+\cup D^-$.
  - Strong classifier: $H(\mathbf{x})=sign(f(\mathbf{x}))=sign(\sum_{t=1}^{T}\alpha_t h_t(\mathbf{x}))$.
  - Weight distribution at the tth iteration: $\mathbf{w}_t=(w_t(1),\ldots,w_t(l))$.
- An error upper bound for AdaBoost:

$$\frac{1}{l}\sum_{i=1}^{l}\frac{1}{2}|y_i-H(\mathbf{x}_i)|\le\frac{1}{l}\sum_{i=1}^{l}\exp(-y_i f(\mathbf{x}_i))=\prod_{t=1}^{T}Z_t,$$

where $Z_t=\sum_{i=1}^{l}w_t(i)\exp(-\alpha_t y_i h_t(\mathbf{x}_i))$.

2.1 Bhattacharyya (BH) Weak Learners

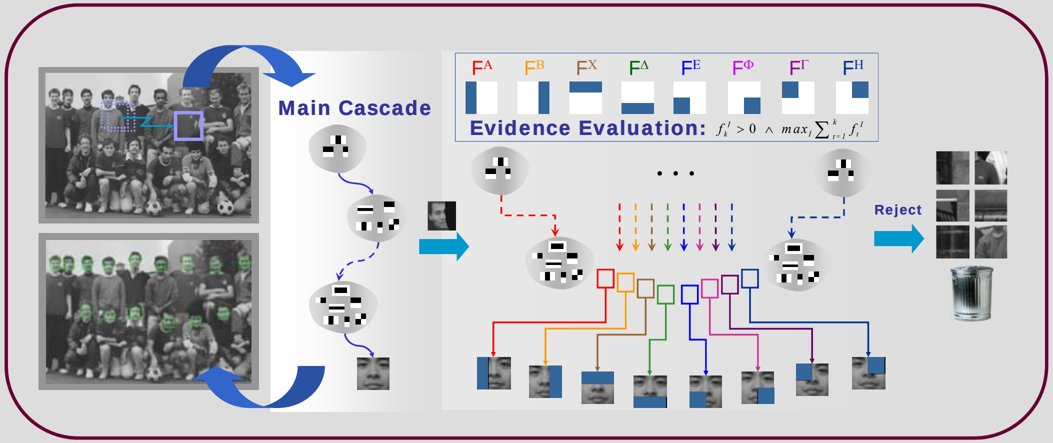
- We start with a set of projection directions, $\Phi=\{\varphi_i\}_{i=1}^{n}$.
  - φ is uniquely defined by a *Rectangle Feature* (RF).
  - φ defines a histogram of D over a partition of m equal-size bins $\{b_k\}_{k=1}^{m}$.
- Define $\mathbf{i}_k(\varphi)$ as indexes of training D data being projected by φ into bin b_k, i.e., $\mathbf{i}_k(\varphi)=\{i\,|\,\mathbf{x}_i\in D,\varphi(\mathbf{x}_i)\in b_k\}$Analogously, $\mathbf{i}_k^+(\varphi)$ and $\mathbf{i}_k^-(\varphi)$ are defined for D^+ and D^-, respectively.
- The values of the weighted positive and negative histograms in bin b_k can be evaluated by: $p_k^+(\varphi)=\sum_{\mathbf{i}_k^+(\varphi)}w(i)$ and $p_k^-(\varphi)=\sum_{\mathbf{i}_k^-(\varphi)}w(i)$.
- Suppose $h_\varphi(\mathbf{x})=s_k$ (for some bin b_k). Then, we have

$$Z=\sum_{k=1}^{m}\sum_{\mathbf{i}_k(\varphi)}w(i)\exp(-\alpha y_i s_k).$$

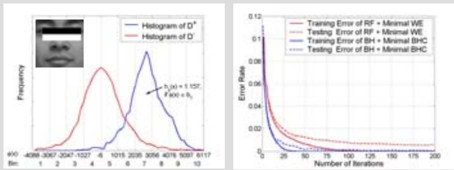
$$\frac{dZ}{ds_k}=0\Rightarrow s_k^*=\frac{1}{\alpha}\ln\sqrt{p_k^+(\varphi)/p_k^-(\varphi)}\ \text{and}\ Z^*=2\sum_{k=1}^{m}\sqrt{p_k^+(\varphi)p_k^-(\varphi)}.$$

- Summarize the observations as follows.
  - The *goodness* of a weak learner h_φ can be measured by its corresponding Bhattacharyya Coefficient.
  - We can *re-scale* h_φ by multiplying with α such that the selection of weak learner and its coefficient can be decided *simultaneously*.
  - A weak learner h_φ of coefficient 1 can be defined by projection φ:

$$h_\varphi(\mathbf{x})=\ln\sqrt{p_k^+(\varphi)/p_k^-(\varphi)},\ \text{for all}\ \mathbf{x}\in D\ \text{and}\ \varphi(\mathbf{x})\in b_k.$$

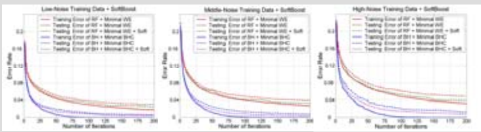


2.2 BH Learners vs. Rectangle Features



2.3 SoftBoost

- Some observations:
  - While the data become noisier, the gaps (overfitting) between training and testing error become larger.
  - The noisier the data, the larger performance difference between the two resulting classifiers (RF and BH).
  - For the classifier using BH weak learners, the problem of overfitting is more serious.
- We apply LP_{Reg}-AdaBoost (Rätsch et al., 2001) to achieve soft margins via linear programming.
  - While the data are noisier or the number of iterations are increasing, the improvements are appreciable.
  - Furthermore, the improvements for RF are even more noticeable!



3. Detecting Faces with Occlusions

- To detect occluded faces, adopting a more lenient policy solves this problem to some degree, but the false positive rates increase.
- A naïve approach is to train a view-based system, i.e., train one detector for each type of occluded faces. ➔ time-consuming.

3.1 Reinforcement Training

- Reinforcement training is used for a boosted cascade.
- A large number of false positives are often generated in the earlier stages of a cascade. At each stage, we use *k-means clustering* to divide false positives from the latest sub-cascade into 6 clusters, and uniformly select non-face training data from each cluster.

3.2 Evidence Cascading

- Distinguishing a rejected occluded face from non-face patterns is vital in our approach---*The clues lie in the evidence left behind when a testing sample is rejected.*
- Let O_I denote the occluded region of the type-I occluded faces. We define the largest subset of Φdisjoint from O_I by:

$$\Phi_I=\{\varphi\,|\,\varphi\in\Phi\ \text{and}\ \varphi\cap O_I=\varnothing\},\ \text{for}\ I=A,B,\ldots,H.$$

- The *evidence* of a testing pattern $\mathbf{x}$ at the kth stage is defined by

$$\varepsilon_k(\mathbf{x})=(f_k^A(\mathbf{x}),f_k^B(\mathbf{x}),\ldots,f_k^H(\mathbf{x}))\ \text{and}\ f_k^I(\mathbf{x})=\sum_I\beta_t h_t(\mathbf{x}).$$

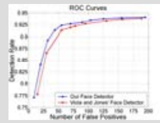
where $\sum_I$ involves only those weak learners that the involving projections do not intersect with O_I .

- The main ideas of *cascading-with-evidence*:
  - The main cascade M is designed to replace the original cascade detector. Besides the frontal faces detection, M also stage-wise accumulates evidence, which is useful when a pattern is rejected.
  - Since each $\beta_t h_t(\mathbf{x})$ has been evaluated in the computation of $H_k(\mathbf{x})$, the evidence vector $\varepsilon_k(\mathbf{x})$ is *easier to derive*.
  - Assume a type-A occluded face $\mathbf{x}$ is rejected at stage k. The positive response of $f_k^A(\mathbf{x})$ in evidence indicates $\mathbf{x}$ as a type-A occluded face. Such a property is *not shared* by most true non-face samples.
  - Unlike a view-based system, only patterns with positive responses are dispatched into the corresponding detectors.

4. Experimental Results

- The regular cascade to detect frontal faces contains 21 stages with 872 BH weak learners.
- 10,000 face samples and 10,000 non-face samples are used in training.
- Averagely, a pattern is classified by 5.36 BH weak learners.
- The main cascade contains 18 stages. Each occluded cascade contains from 23 to 28 stages.
- Since a BH weak learner may be associated with more than one type, the number of weak learners in each stage of main cascade is still *manageable*.
- The first three stages of M consist of 7, 9, 12 weak learners respectively.
- Our system is more robust to detect faces with exaggerated expressions, since only partial information is extracted to detect faces in occluded cascades.

Stage	1	2	3	4
Weak learner #	2	4	5	8
Passing rate (PR)	35.8	39.2	40.5	41.7
Cumulative PR	35.8	14.0	5.7	2.4



4.1 Detection Results