Face Detection Using SURF Cascade

Jianguo Li, Tao Wang, Yimin Zhang

Intel Labs China
Outline

• Cascade Detection Revisited
  – Problems & motivations

• SURF-Cascade
  – SURF Feature
  – Maximizing AUC

• Benchmark

• Conclusion
Cascade Detector Revisited

- Five ingredients
  - Feature representation: Haar, HoG, ...
    • Integral image to speedup feature extraction
  - Weak classifiers: dtree, linear SVM, ...
  - Training algorithm: Boosting, ...
  - Cascade structure: hard, soft, chain, ...
  - Scan strategy: slide-window, ...

\[
\sum_{i} f_{1i}(x) > \theta_1 \quad \text{T} \quad \sum_{i} f_{2i}(x) > \theta_2 \quad \text{T} \quad \sum_{i} f_{ni}(x) > \theta_n 
\]

- All detecting windows
- Merge and refine
- Rejected detecting window
Problems & motivations

- Practical detector requires ~1e-6 FPPW
  - Huge training set required
  - need scan >10^8 negative samples

- Large feature pool
  - In Haar cascade, > 200,000 features for 20x20 template.

- Slow convergence speed
  - Training based on two conflicted objectives: TPR/FPR
    - i.e, in each stage, set minTPR (0.995) and maxFPR (0.5)
    - Reach FPR=0.5 is easy in early stages
    - But TPR is not converged simultaneously
    - 1e-6=0.5^20, while 0.995^20=0.905

Weeks => Days => Hours?
SURF Cascade (1)

- **Features: SURF**
  - 2x2 cell of patch
  - Each cell is 8-dim vector
    - Sum of dx, |dx| when dy >=0
    - Sum of dx, |dx| when dy <0
    - Sum of dy, |dy| when dx >=0
    - Sum of dy, |dy| when dx <0
  - Total is 2x2x8 = 32 dim feature vector
  - 8-channel integral images

- **Feature Pool**
  - In a 40x40 face detection template
  - Slide the patch (x, y, w, h) with fixed step = 4 pixels
  - Each cell at least 8x8 pixels, w or h at least 16 pixels
  - with 1:1, 1:2, 2:3... aspect-ratio (w/h)
  - Totally 396 local SURF patches

- **Weak classifier: logistic regression on 32dim SURF**
  - h(x) = P(y|x, w) = 1/(1+exp(-ywx)).
SURF Cascade (2)

- Cascade training
  - AdaBoost in each stage
    \[ H^T(x) = \sum_{t=1}^{T} \alpha_t h_t(x, w), \]
  - Feature selection: maximize AUC score \( J \)
    \[ H^t(x) = \arg \max_{k=1:K} J(H^{t-1}(x) + \alpha_k h_k(x)). \]
  - Convergence test: AUC
  - Determine threshold when converged
    - Search on ROC curve with given TPR
Searching on ROC curve

• In comparison, the Viola-Jones framework
  – Overall FPR $10^{-6} = 0.5^{20}$
  – One stage TPR=0.995, overall $0.995^{20} = 0.905$

• Given TPR while FPR is adaptive
  – The FPR on 8-stage may like:
    • $10^{-6} = 0.305 	imes 0.226 	imes 0.147 	imes 0.117 	imes 0.045 	imes 0.095 	imes 0.219 	imes 0.268$
    – Overall TPR $= 0.995^{8} = 0.970$
Training performance

- Implement in C/C++ on X86
  - Parallelize the feature search step using OpenMP
  - SIMD for classifier (wx) and feature extraction

- Training dataset
  - 13000 faces from GENKI/FaceTracer database
    - With mirrors and resampling to obtain 39000 faces in total
  - 18000 non-face images from caltech101, image-net, etc.

- Training status
  - Platform: Intel Core-i7, 3.2GHz, 4-core, 8-thread.
  - On demand search of negative from non-face images
    - Totally scanned 13.6 billions of negative samples
  - Reach 1e-6 FPPW in 8-stages
Cascade statistics

<table>
<thead>
<tr>
<th></th>
<th>#stages</th>
<th>#weak</th>
<th>Model-size</th>
<th>Hit-rate (CMU Frontal)</th>
<th>training-time on Core i7</th>
</tr>
</thead>
<tbody>
<tr>
<td>VJ (OpenCV)</td>
<td>24</td>
<td>2912</td>
<td>&gt;1MB</td>
<td>76.1%</td>
<td>~3 days</td>
</tr>
<tr>
<td>SURF</td>
<td>8</td>
<td>334</td>
<td>58KB</td>
<td>90.8%</td>
<td>47min</td>
</tr>
</tbody>
</table>
What if?

• OpenCV Haar-training on the same dataset
  – Need 3 days (OpenMP tuned on)

• VJ’s criteria (TPR + FPR) for SURF?
  – Need 5 hours to reach 1e-6 at the 19-th stage
Evaluation on CMU+MIT frontal-set

The graph shows the performance of different face detection algorithms on the CMU+MIT frontal-set. The x-axis represents false positives, while the y-axis represents true positive rate. The algorithms compared include SURF, SoftCascade [2], Recycling [3], PolyFace [24], and Viola-Jones [32].
Evaluation on UMass FDDB (frontal)
Multi-view SURF cascade on UMass
Some detection results
Detection speed

- Test on three videos
  - A,B,C --- B has more faces than C in average

<table>
<thead>
<tr>
<th>Videos</th>
<th>resolution</th>
<th>Ours (fps)</th>
<th>OpenCV (fps)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.avi</td>
<td>352×288</td>
<td>30.1</td>
<td>25.0</td>
</tr>
<tr>
<td>B.mpg</td>
<td>640×480</td>
<td>5.8</td>
<td>4.6</td>
</tr>
<tr>
<td>C.avi</td>
<td>640×480</td>
<td>7.3</td>
<td>5.4</td>
</tr>
<tr>
<td>A.avi</td>
<td>352×288</td>
<td>312</td>
<td>190</td>
</tr>
<tr>
<td>B.mpg</td>
<td>640×480</td>
<td>71.3</td>
<td>49.7</td>
</tr>
<tr>
<td>C.avi</td>
<td>640×480</td>
<td>90.5</td>
<td>58.2</td>
</tr>
</tbody>
</table>

- Why SURF cascade is faster than Haar-cascade
  - Average number of weak classifiers evaluated
    - SURF-cascade: 1.5
    - Haar: 28
  - Easy SIMD for SURF-cascade
    - 32-dim float => 128bit SIMD, 4-data in parallel
    - $1.5 \times \frac{32}{4} = 12$
Conclusion

• Contributions
  – Introduce SURF feature for fast face detection
  – Propose AUC as single criterion for cascade training
  – Build a cascade face detector from billions of samples on PC within one hour.

• Advantages of SURF cascade
  – Very short cascade and small size (8 stages, ~58KB)
  – Accuracy is comparable to stage-of-the-art detectors.
  – Even faster than OpenCV optimized Haar-cascade
Thanks!