Face Detection Using SURF Cascade

Jianguo Li, Tao Wang, Yimin Zhang

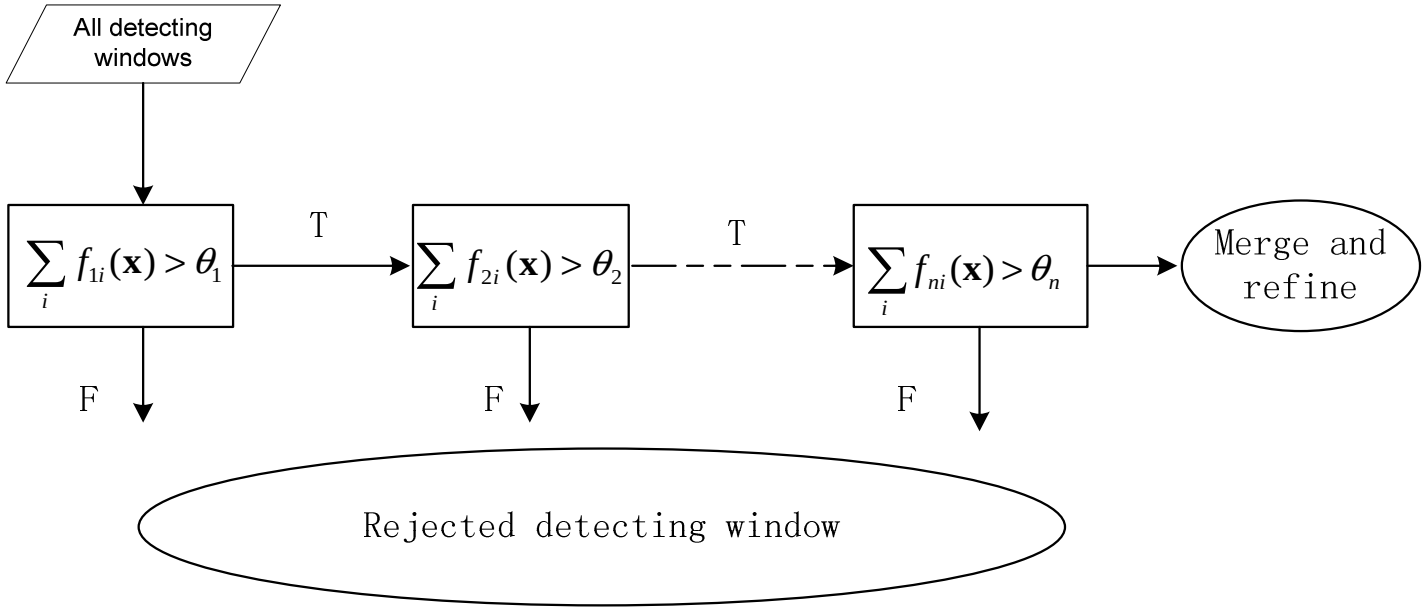
Intel Labs China

Outline

- Cascade Detection Revisited
  - Problems & motivations
- SURF-Cascade
  - SURF Feature
  - Maximizing AUC
- Benchmark
- Conclusion

Cascade Detector Revisited

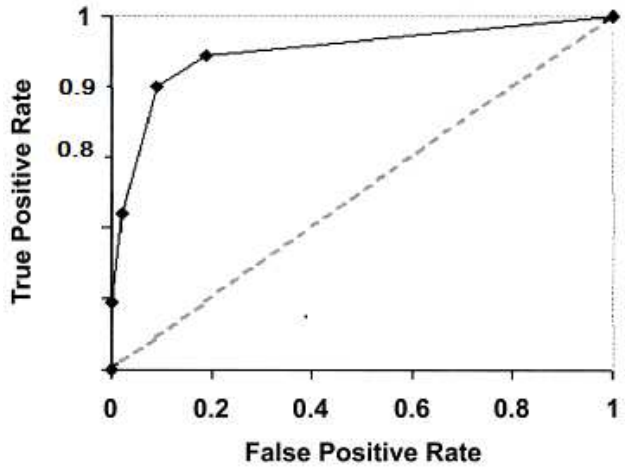
- Five ingredients
  - Feature representation: Haar, HoG, …
    - Integral image to speedup feature extraction
  - Weak classifiers: dtree, linear SVM, …
  - Training algorithm: Boosting, …
  - Cascade structure: hard, soft, chain, …
  - Scan strategy: slide-window, …



Problems & motivations

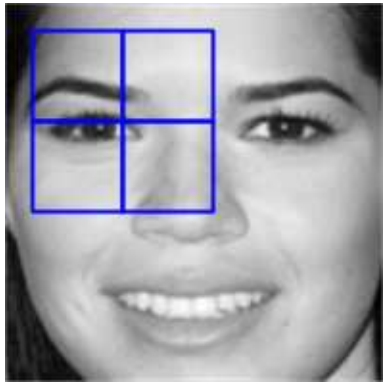
- Practical detector requires ~1e-6 FPPW
  - Huge training set required
  - need scan >10^8 negative samples
- Large feature pool
  - In Haar cascade, > 200,000 features for 20x20 template.
- Slow convergence speed
  - Training based on two conflicted objectives: TPR/FPR
    - i.e, in each stage, set 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)

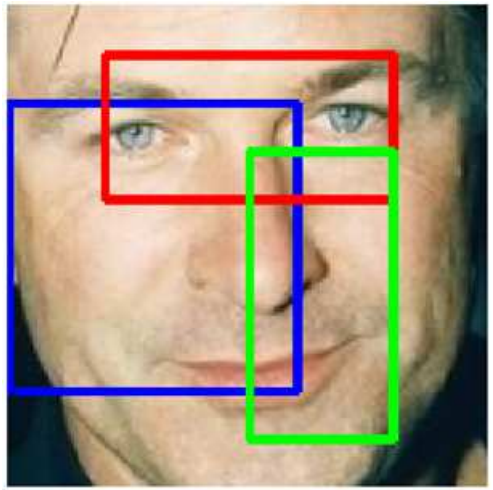
- Cascade training
  - AdaBoost in each stage

$$H^T(x) = \sum_{t=1}^{T} \alpha_t h_t(\mathbf{x}, \mathbf{w}),$$

  - Feature selection: maximize AUC score J

$$H^t(\mathbf{x}) = \arg\max_{k=1:K} J(H^{t-1}(\mathbf{x}) + \alpha_k h_k(\mathbf{x})).$$

  - Convergence test: AUC
  - Determine threshold when converged
    - Search on ROC curve with given TPR

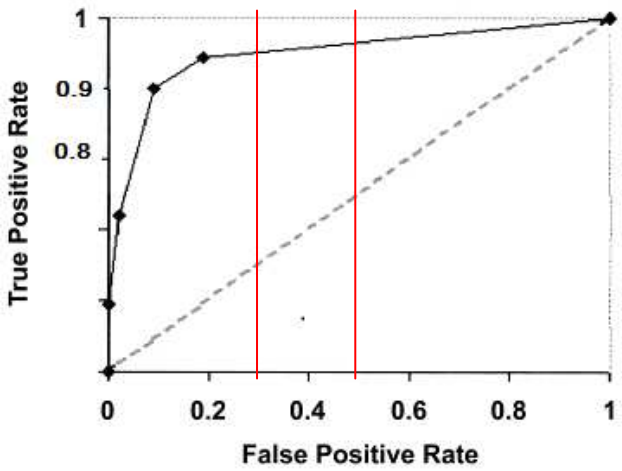


Searching on ROC curve

- In comparison, the Viola-Jones framework
  - Overall FPR 1e-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:
    - 1e-6 = 0.305x0.226x0.147x0.117x0.045x0.095x0.219x0.268
  - Overall TPR = 0.995^8 = 0.970

$$F = \prod_{i=1}^{K} f_i,$$

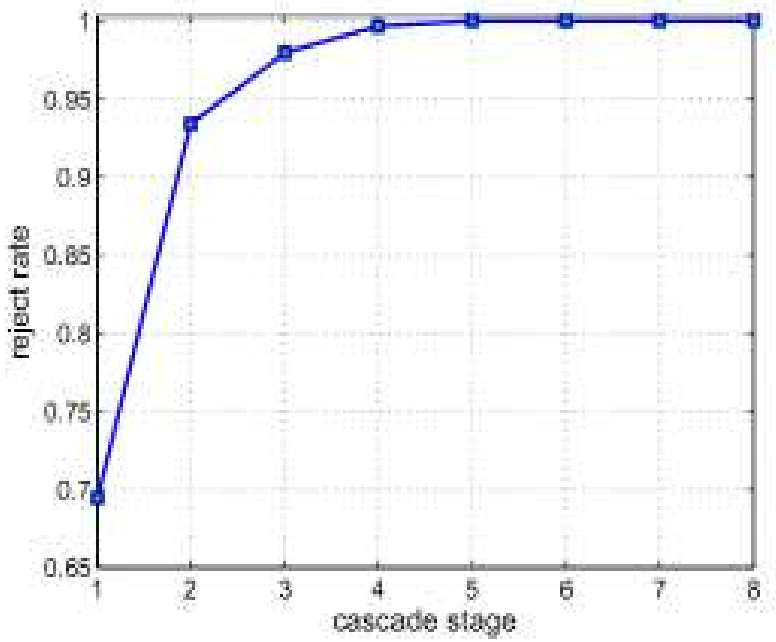
$$D = \prod_{i=1}^{K} d_i,$$

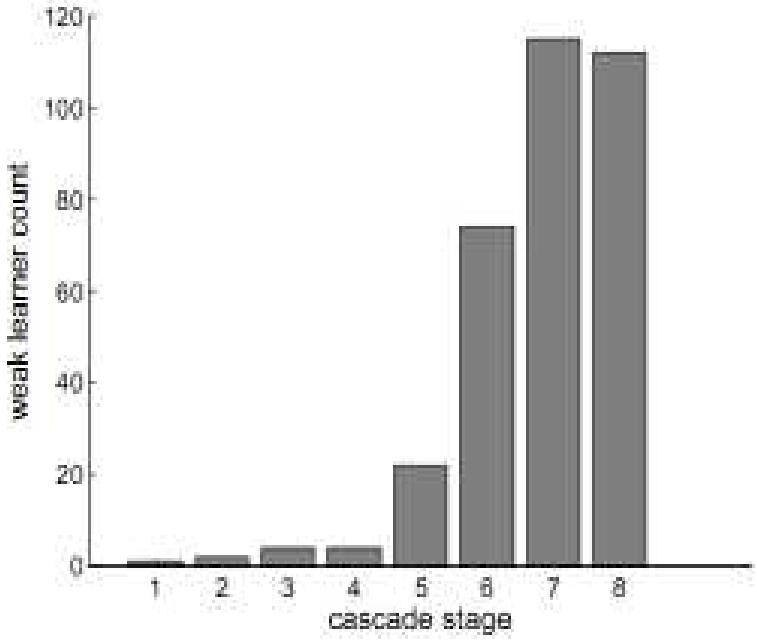


Training performance

- 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



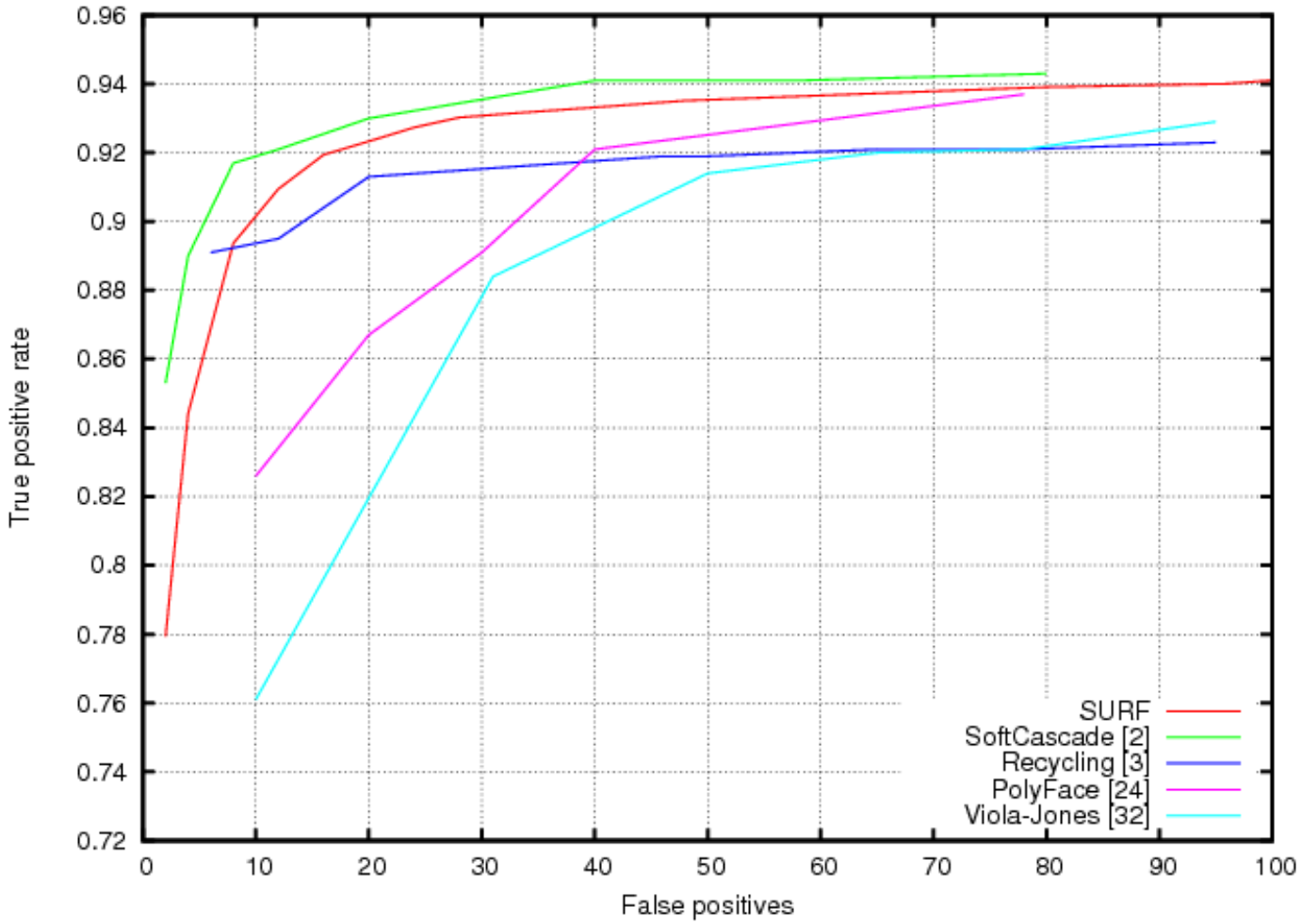


	#stages	#weak	Model-size	Hit-rate (CMU Frontal)	training-time on Core i7
VJ (OpenCV)	24	2912	>1MB	76.1%	~3 days
SURF	8	334	58KB	90.8%	47min

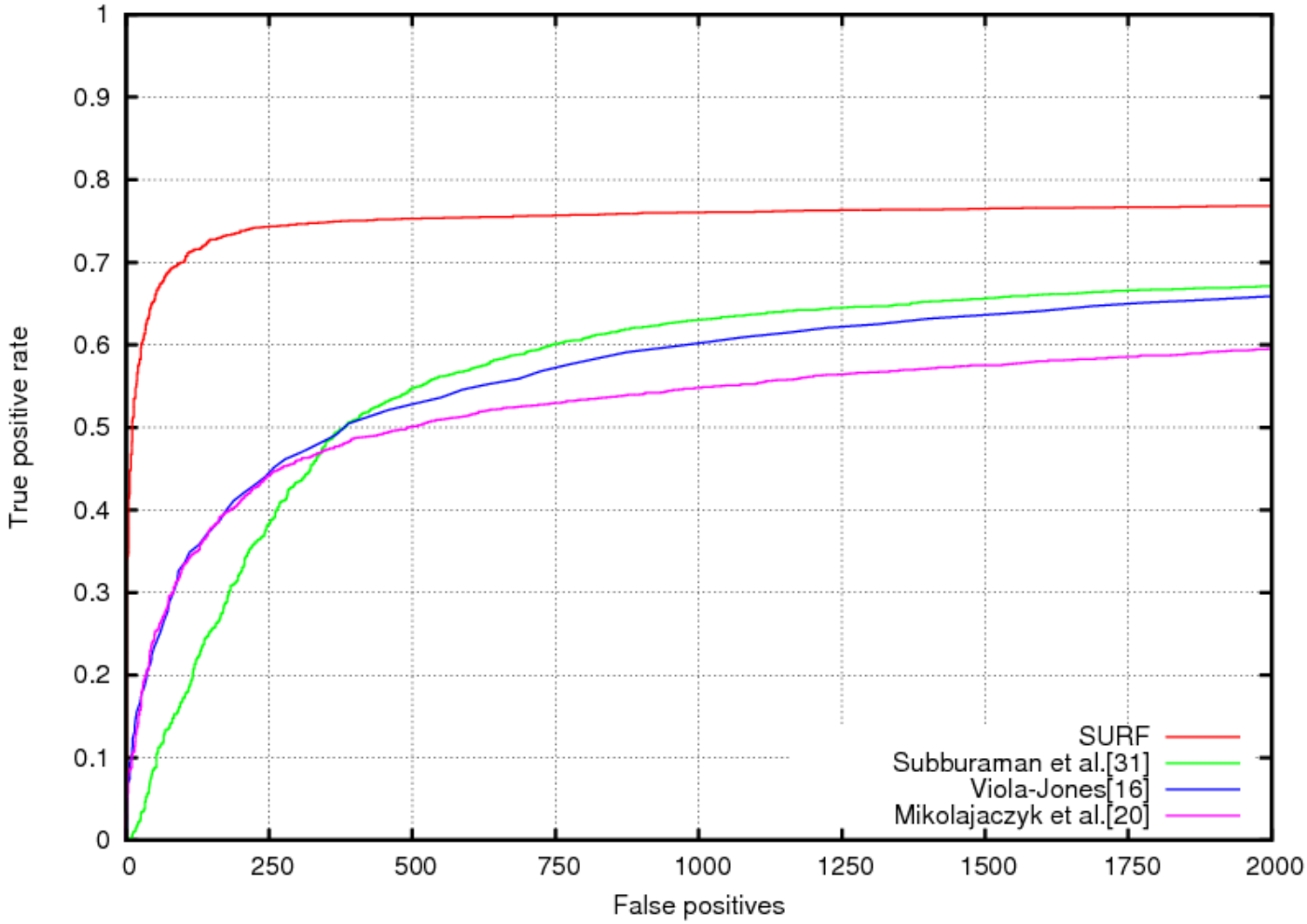
What if?

- 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

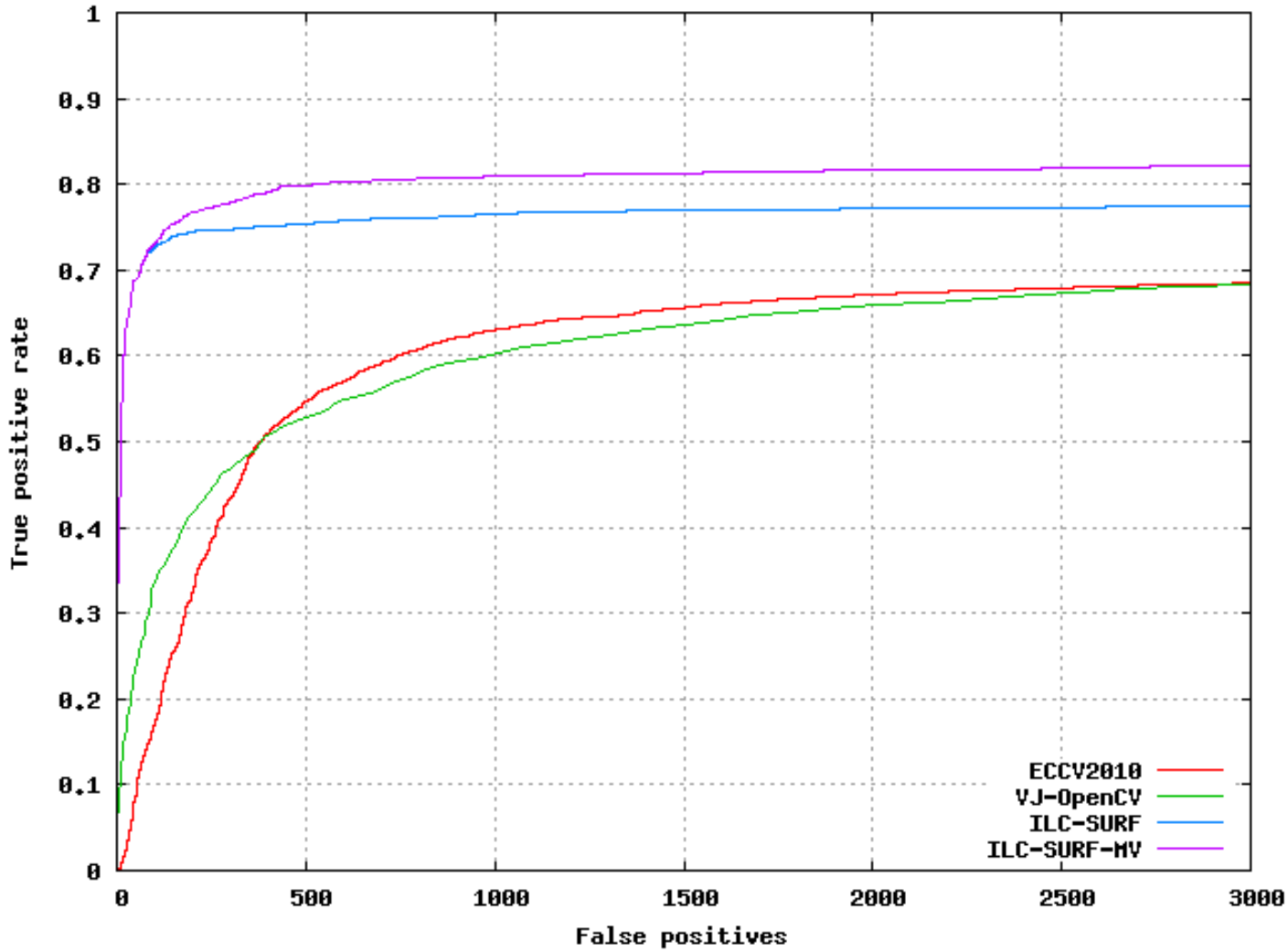
Evaluation on CMU+MIT frontal-set



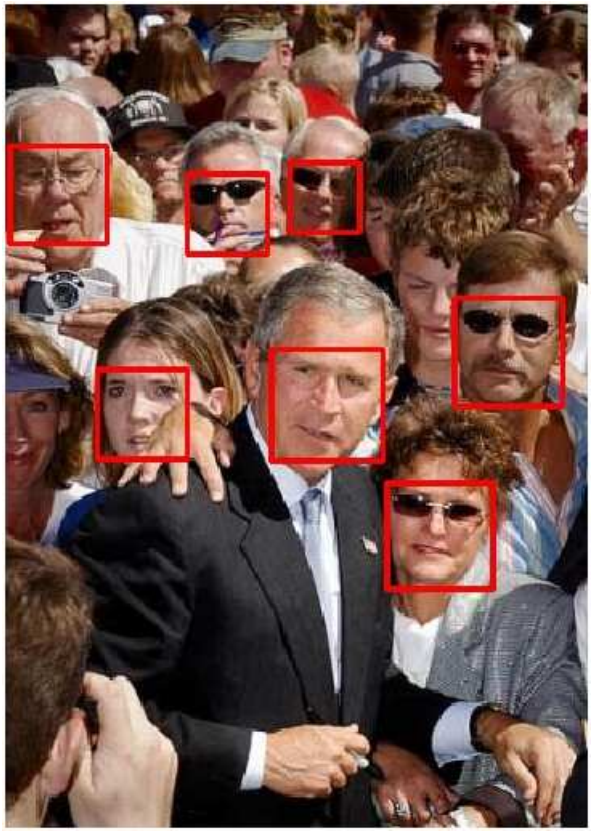
Evaluation on UMass FDDB (frontal)



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

Videos	resolution	Ours (fps)	OpenCV (fps)
A.avi	352×288	30.1	25.0
B.mpg	640×480	5.8	4.6
C.avi	640×480	7.3	5.4
A.avi	352×288	312	190
B.mpg	640×480	71.3	49.7
C.avi	640×480	90.5	58.2

Intel Atom 1.6GHz

Intel Core i7 3.2GHz

- 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*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!