Efficiency and Accuracy Improvements to the Viola-Jones Object Detection Algorithm Using ML Methods

James Skelly, ECG703 Final Presentation

Presentation Overview

- Introduce the face detection problem and its importance
- Overview of the original (2001) Viola-Jones algorithm
  - Outline the operation
  - Discuss the original results
- Provide examples of Viola-Jones in applications outside of face detection
- Highlight improvements made to the original algorithm
  - Redundancy reduction
  - Evolutionary pre-selection of features
  - Application of composite features
- Compare Viola-Jones performance to that of other modern algorithms
- Conclusion and wrap up with Q/A

Face Detection

- The first step of all other facial analysis algorithms
  - Face recognition, facial feature recognition, face scanning
- Recognition has several applications
  - Preventing retail crime, smart advertising, finding missing persons, aiding forensic investigation, protecting schools from potential threats, diagnosis of various diseases
- Easy for humans, not so easy for computers

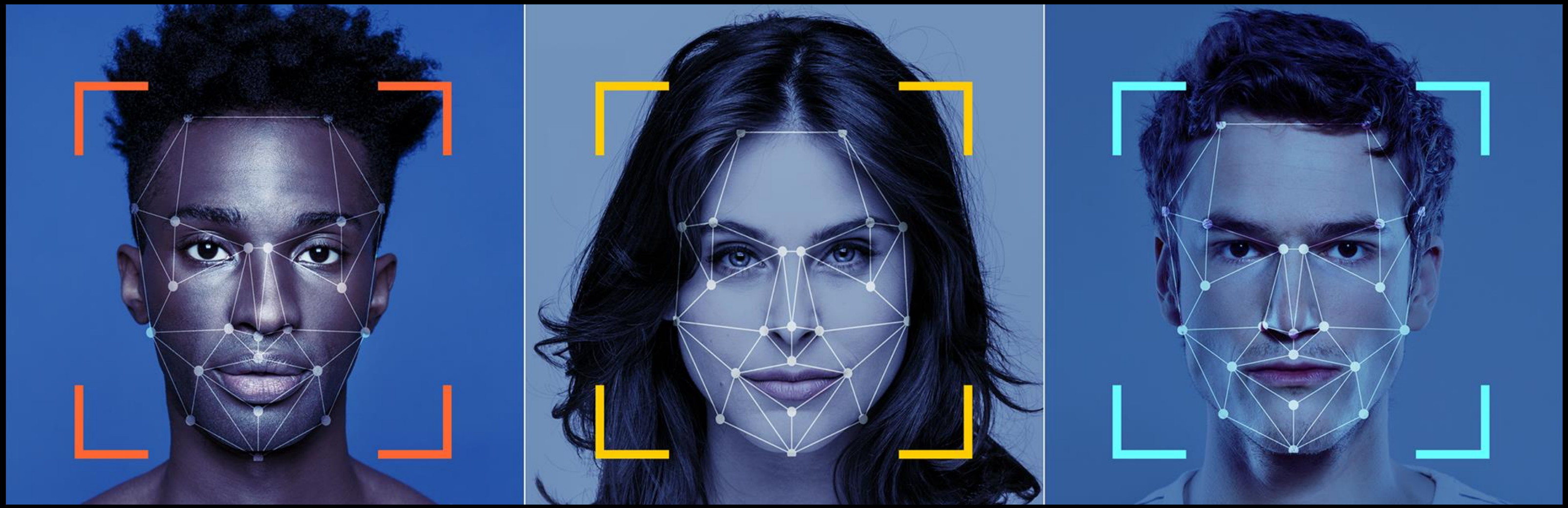

Figure 1: Facial Recognition Graphic [2]

The Viola-Jones Algorithm

The Viola-Jones Algorithm

- Proposed in 2001 as an object detection algorithm, mainly proposed for use as a face detection algorithm
- Looks at regions of pixels and compares the sum of their values to find edges and other defining features
- 15 times greater frame rate than competing algorithms in 2001
- Four main components of the algorithm
  - Haar-like rectangular features
  - Concept of the integral image
  - Boosting with AdaBoost
  - Attentional cascade of classifiers

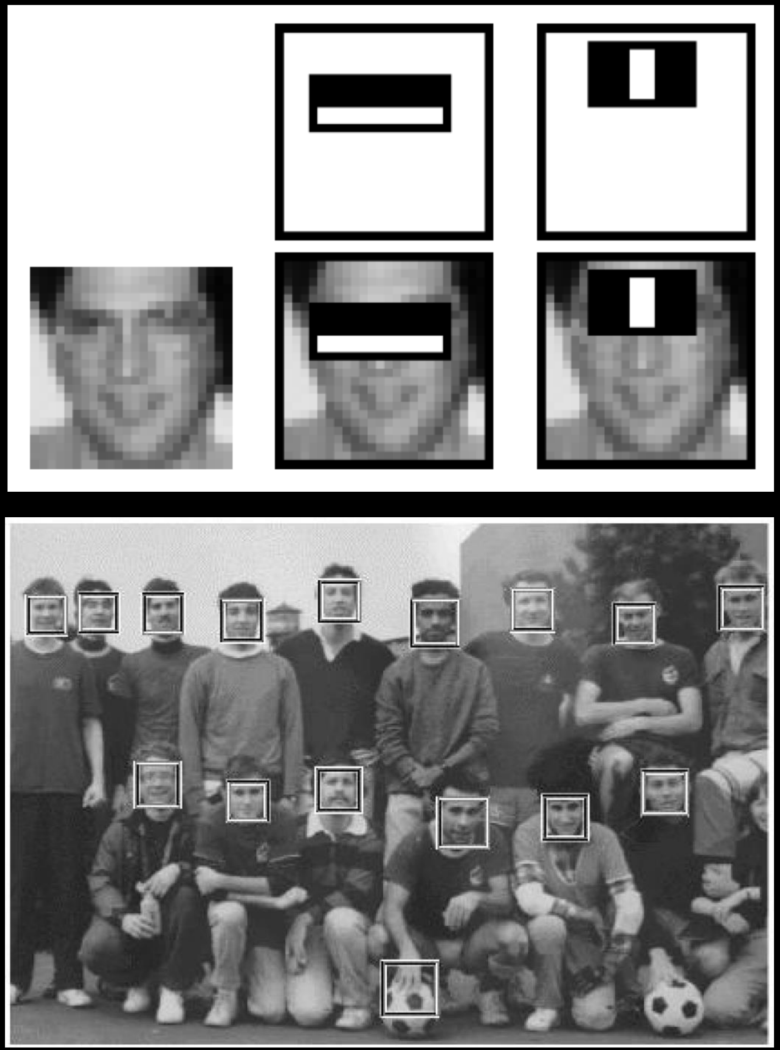

Figure 2: Images from Viola-Jones Paper, 2001 [3]

The Viola-Jones Algorithm

Haar-like Rectangular Features

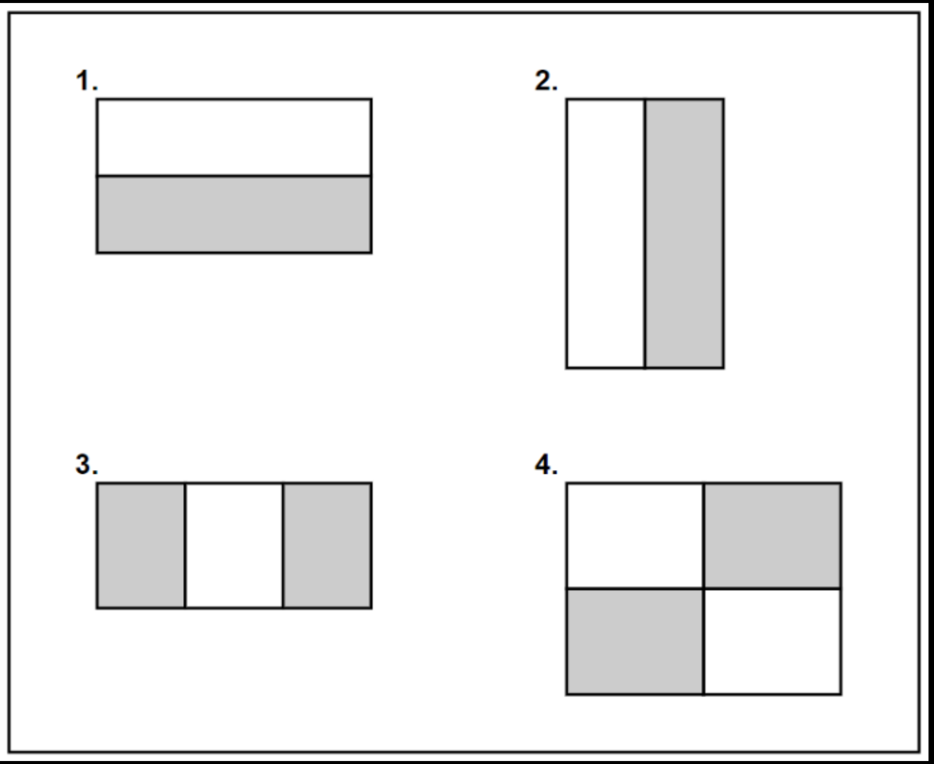

Figure 3: Examples of Haar-like Rectangular Features

The Concept of the Integral Image

Original Image

	0	1	2	3
0	30	50	70	80
1	70	90	50	60
2	90	10	30	60
3	60	40	40	20

Integral Image

	0	1	2	3
0	30	80	150	230
1	100	240	360	500
2	190	340	490	690
3	250	440	630	850

Figure 4: Example of Computing the Integral Image

The Viola-Jones Algorithm

Boosting with AdaBoost

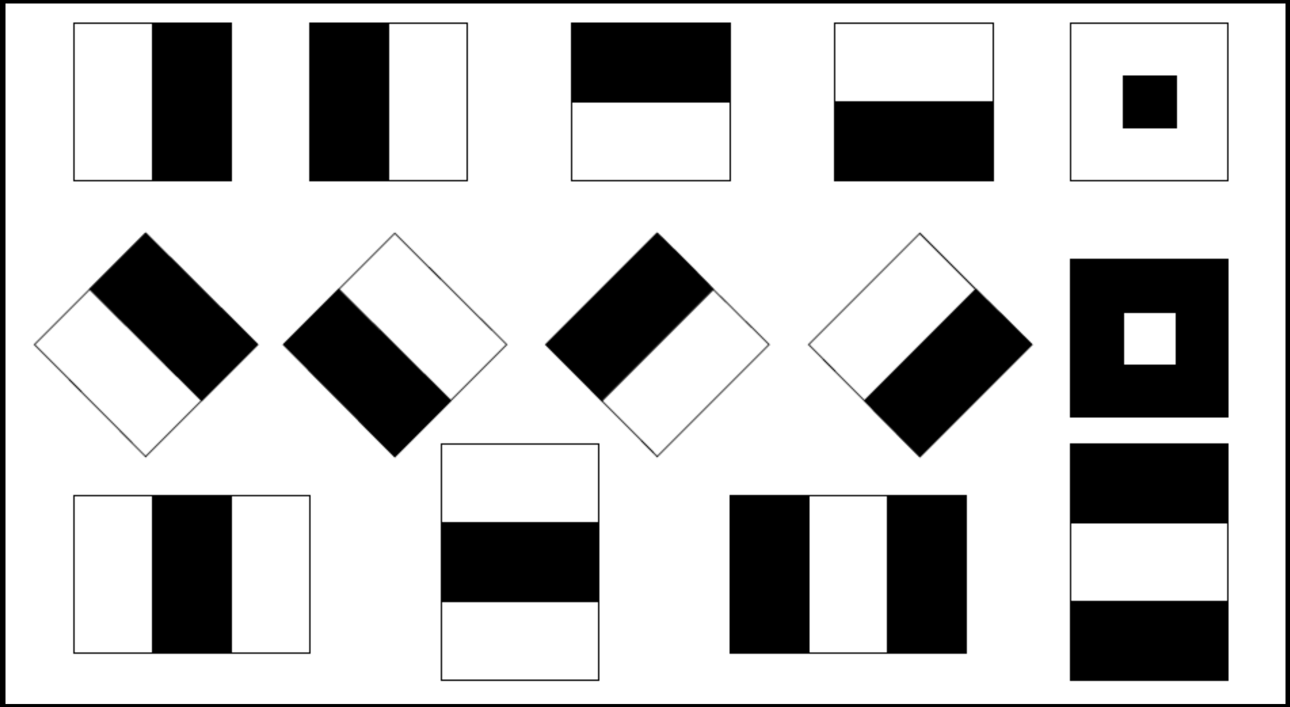

Figure 5: Various Rectangular Features [5]

The Attentional Cascade

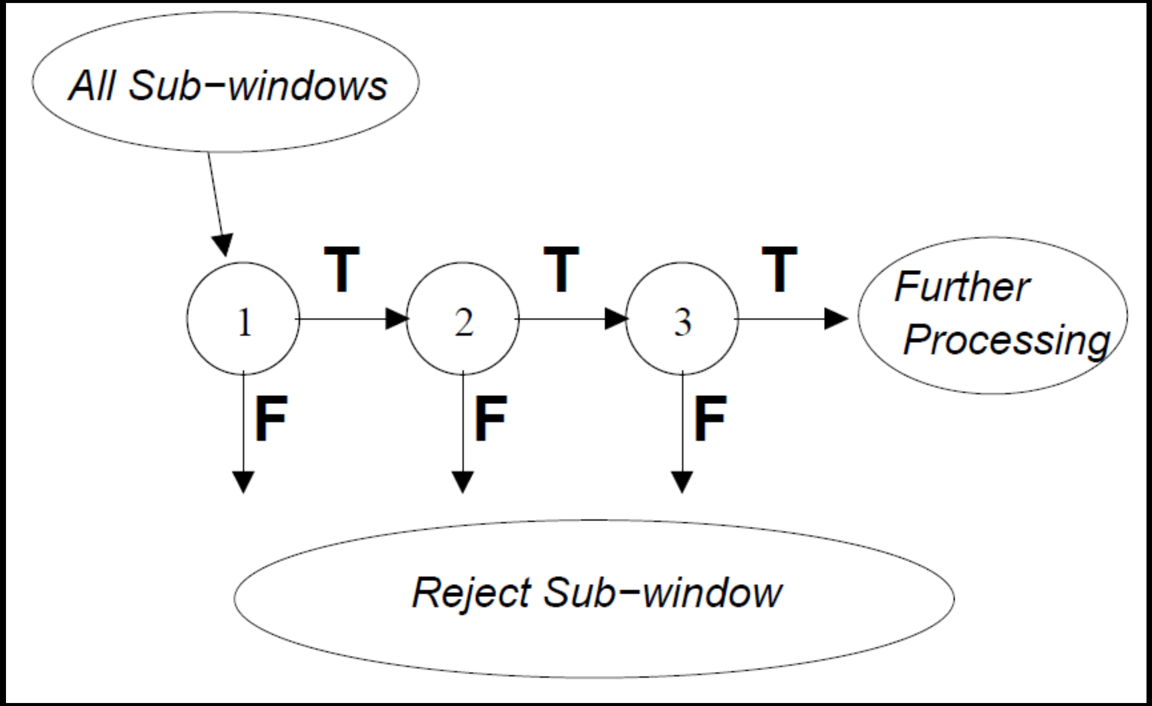

Figure 6: Degenerate Decision Tree (Cascaded Classifiers) [3]

Results of Viola-Jones, 2001

Detector \ False detections	10	31	50	65	78	95	167
Viola-Jones	76.1%	88.4%	91.4%	92.0%	92.1%	92.9%	93.9%
Viola-Jones (voting)	81.1%	89.7%	92.1%	93.1%	93.1%	93.2 %	93.7%
Rowley-Baluja-Kanade	83.2%	86.0%	-	-	-	89.2%	90.1%
Schneiderman-Kanade	-	-	-	94.4%	-	-	-
Roth-Yang-Ahuja	-	-	-	-	(94.8%)	-	-

Figure 7: Table Showing Detection Rates of Original Viola-Jones Algorithm
Compared Against Other Algorithms of the Time [3]

Applications Outside of Face Detection

- Detection and tracking of locomotive activity of animals in wildlife video
- Emotion recognition to determine success in the learning environment
- Drowsiness detection to improve brain-computer interfaces, prevent auto accidents
- Vehicle counting system for traffic monitoring and surveillance
- Hand gesture recognition

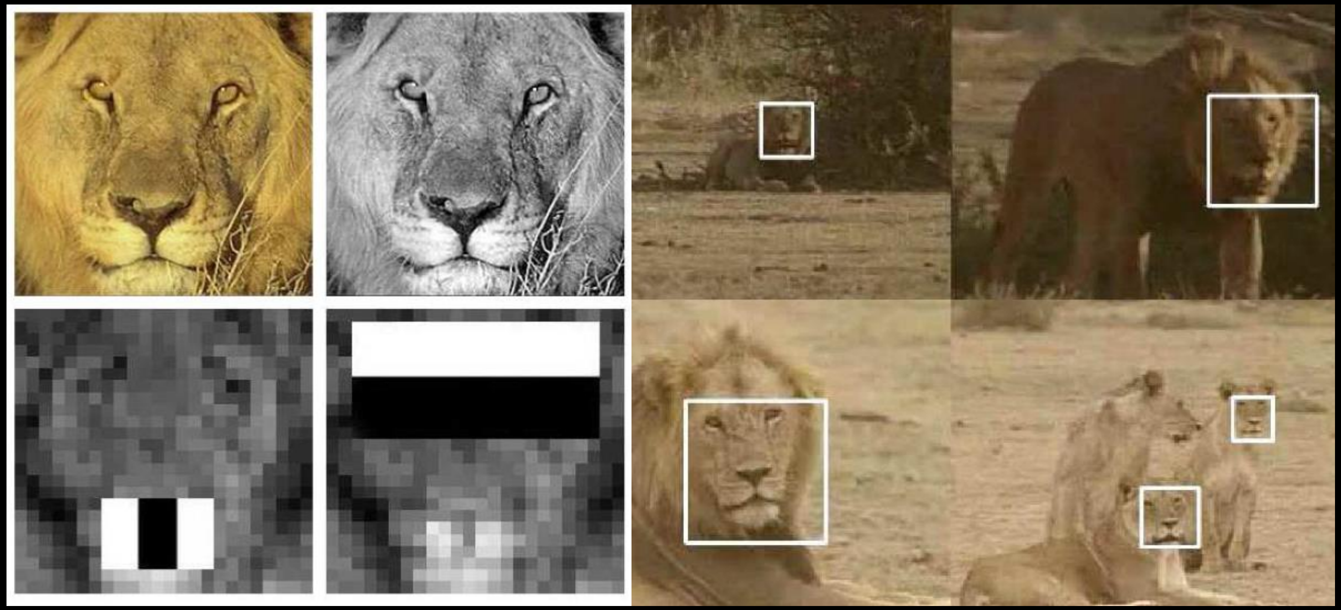

Figure 8: Face Detection on Lions Using Viola-Jones [8]

Improvements to the Algorithm

Redundancy Reduction

- Multiple similarly sized, similarly located windows detect the same face
- Accuracy, precision, and recall of the algorithm are improved
- True positive, true negative, false positive, false negative rates improved

No	Comparison		
	Test	*Proposed Method*	*Traditional Viola-Jones* [2]
1	True Positive	740	520
2	True Negative	25	0
3	False Negative	100	155
4	False Positive	35	225
5	Accuracy	85%	77%
6	Precision	95%	71%
7	Recall	88%	81%

Figure 9: Table of Reduced Redundancy Results [6]

Figure 10: Example Image Showing Redundancy (left) and Removal of Redundancy (right) [6]

Feature Pre-Selection

- Artificial evolutionary process to pre-select features for the classifiers, reduce training time and overfitting
- Incremental size changes for features, duplicates are skipped so that only unique classifiers are formed
- Only aiming for cascades with fewer stages than Viola-Jones
- Obtain 5 cascades with better performance, less stages, less features

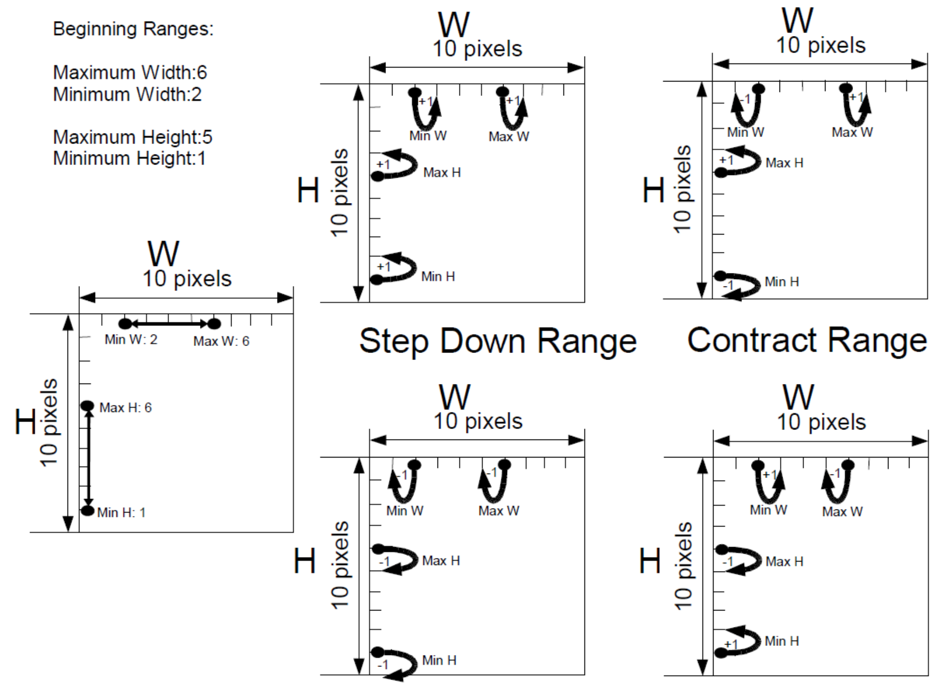

Figure 11: Incremental Feature Size in Evolutionary Process [5]

Cascade	Evaluated	Recall	Precision	False Pos. Rate	Stages	Features
Control	1	0.9398	0.0925	0.3931	12	68
Evolved	120	0.9297±0.0022	0.1044±0.0033	0.3867±0.0141	11.33±0.07	76.46±1.05
+ Perf.	26	0.9487±0.0011	0.1054±0.0027	0.3493±0.0090	11.35±0.15	75.46±1.79
+ Perf., - Stages	13	0.9501±0.0015	0.1019±0.0030	0.3605±0.0097	10.85±0.19	71.54±2.32
+ Perf., - Stages, - Feat.	5	0.9493±0.0023	0.0988±0.0003	0.3692±0.0005	10.2±0.2	64±0

Figure 12: Table of Results When Features are Pre-Selected [5]

Application of Composite Features

- Alternative weak learner to the simple rectangular features
- Drastically decreases the number of false positives detected
- Slight decrease in frame rate results from more complex features

Table 3 Comprehensive comparison of the two algorithms

Algorithm name	Total face number	Total missing count	Total error count	Total missing rate	Gross error rate
Viola-Jones	1372	189	21	0.138	0.015
Article method		75	1	0.055	0.0007

Figure 13: Table Comparing Results with and without Composite Features [7]

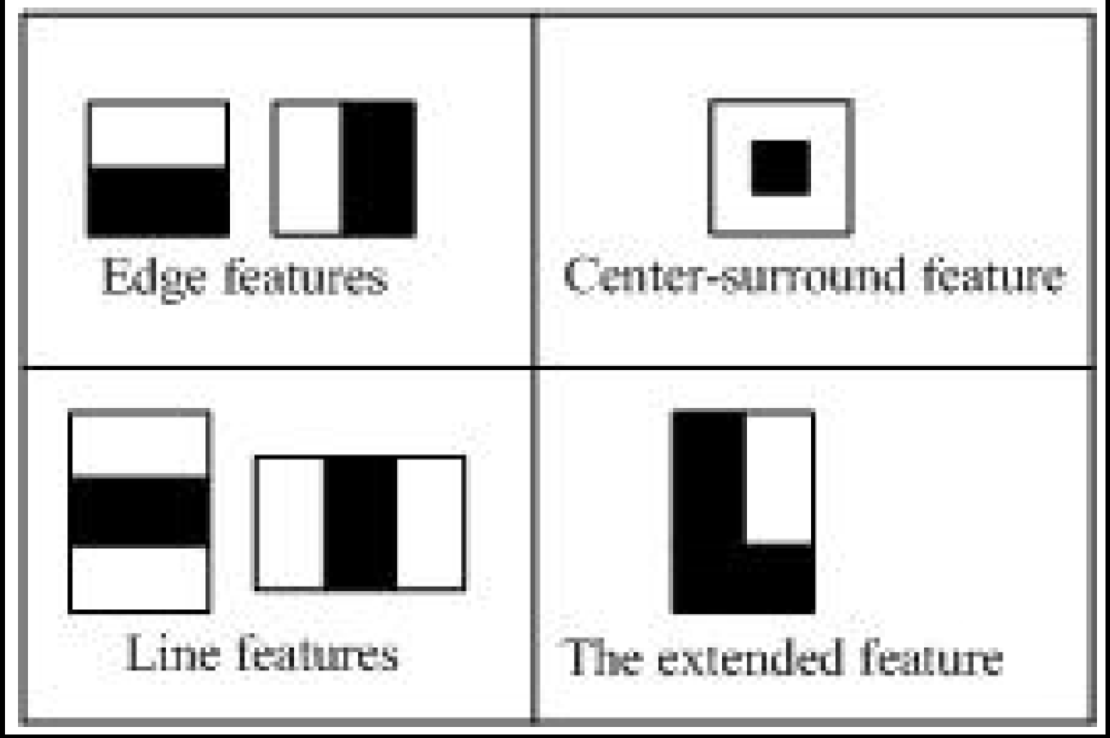

Figure 14: Simple Features vs. Composite Features [9]

Comparison to other Algorithms

Viola-Jones Compared with Other Algorithms

- Other algorithms have been written in recent years using:
  - Neural networks
  - SMQT features
  - Support Vector Machines
- Neural network-based face detectors outperform Viola-Jones in terms of detection rates
- Viola-Jones outperforms NN-based detectors in terms of time complexity and memory consumption

SMQT – Successive Mean Quantization Transform

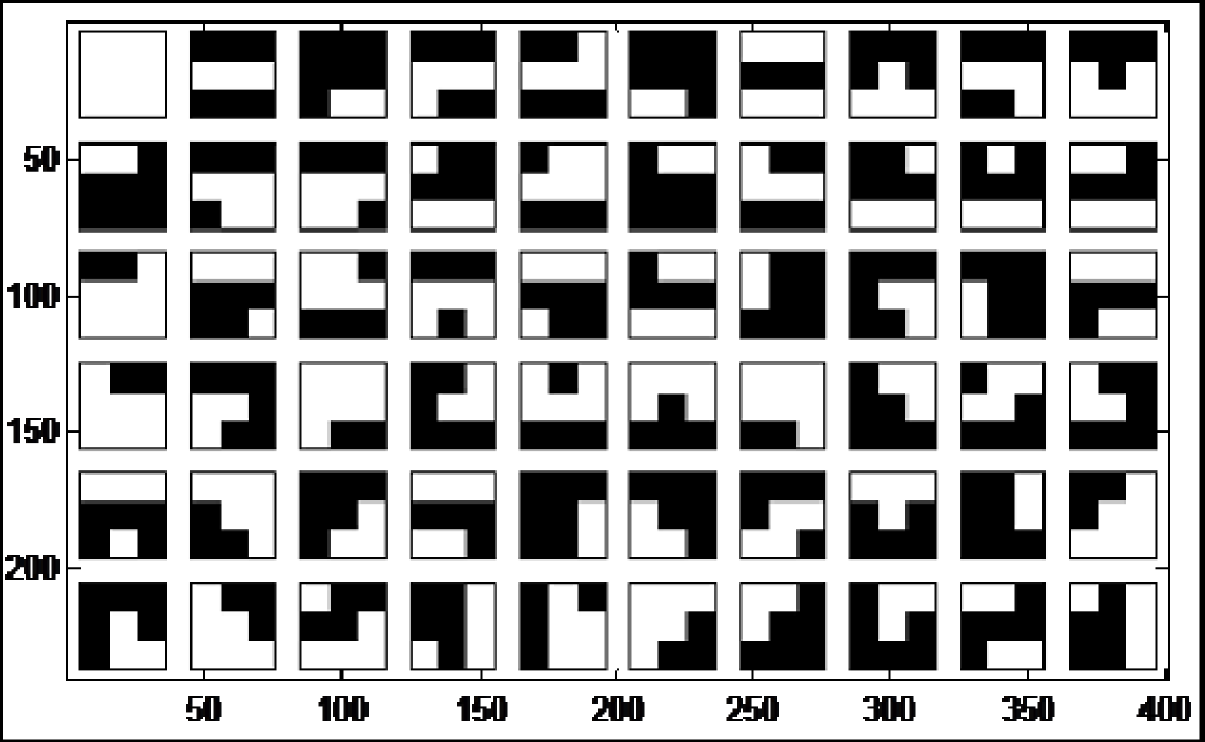

Figure 15: SMQT Feature Examples [1]

Viola-Jones Compared with Other Algorithms

Detector	Time		Memory
	GFLOPS	FPS	consumption (GB)
Viola-Jones [18]	0.6	60.0	0.1
HeadHunter DPM [20]	5.0	1	2.0
SSD[6]	45.8	13.3	0.7
Faster R-CNN [5]	223.9	5.8	2.1
R-FCN 50 [3]	132.1	6.0	2.4
R-FCN 101 [3]	186.6	4.7	3.1
PVANET [13]	40.1	9.0	2.6
Local RCNN [19]	1206.8	0.5	2.1
Yolo 9000 [16]	34.90	19.2	2.1

Figure 16: Table Showing Superior Time Complexity, Memory Consumption of Viola-Jones Compared to Competing Algorithms [4]

Thank you! Questions?

References

[1] Artan, Y., Burry, A., Kozitsky, V., & Paul, P., “Efficient SMQT features for snow-based classification on face detection and character recognition tasks,” 2012.

[2] B. Smith, “Finally, progress on regulating facial recognition,” March, 2020. [Online]. Available: https://blogs.microsoft.com/on-the-issues/2020/03/31/washington-facial-recognition-legislation/. [Accessed May 4, 2021].

[3] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, Kauai, HI, USA, 2001, pp. I-I, doi: 10.1109/CVPR.2001.990517.

[4] L. T. Nguyen-Meidine, E. Granger, M. Kiran and L. Blais-Morin, “A Comparison of CNN-based Face and Head Detectors for Real-Time Video Surveillance Applications,” 2017 The 7th International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, Canada, Nov. 28 to Dec. 1, 2017.

[5] S. R. Lang, M. H. Luerssen and D. M. W. Powers, "Evolutionary Feature Preselection for Viola-Jones Classifier Training," 2012 Spring Congress on Engineering and Technology, Xi'an, China, 2012, pp. 1-4, doi: 10.1109/SCET.2012.6342142.

[6] K. C. Kirana, S. Wibawanto and H. W. Herwanto, "Redundancy Reduction in Face Detection of Viola-Jones using the Hill Climbing Algorithm," 2020 4th International Conference on Vocational Education and Training (ICOVET), Malang, Indonesia, 2020, pp. 139-143, doi: 10.1109/ICOVET50258.2020.9230349.

[7] W. Lu and M. Yang, "Face Detection Based on Viola-Jones Algorithm Applying Composite Features," 2019 International Conference on Robots & Intelligent System (ICRIS), Haikou, China, 2019, pp. 82-85, doi: 10.1109/ICRIS.2019.00029.

[8] T. Burghardt and J. Calic, "Real-time Face Detection and Tracking of Animals," 2006 8th Seminar on Neural Network Applications in Electrical Engineering, Belgrade, Serbia, 2006, pp. 27-32, doi: 10.1109/NEUREL.2006.341167.

[9] L. Yun and Z. Peng, "An Automatic Hand Gesture Recognition System Based on Viola-Jones Method and SVMs," 2009 Second International Workshop on Computer Science and Engineering, Qingdao, China, 2009, pp. 72-76, doi: 10.1109/WCSE.2009.769.