Fast Rotation Invariant Multi-View Face Detection Based on Real Adaboost

Bo WU[1], Haizhou AI[1], Chang HUANG[1] and Shihong LAO[2]
[1] *Department of Computer Science and Technology, Tsinghua University, Beijing, 100084, China*
[2] *Sensing Technology Laboratory, Omron Corporation*
E-mail: ahz@mail.tsinghua.edu.cn

Abstract

In this paper, we propose a rotation invariant multi-view face detection method based on Real Adaboost algorithm [1]. Human faces are divided into several categories according to the variant appearance from different view points. For each view category, weak classifiers are configured as confidence-rated look-up-table (LUT) of Haar feature [2]. Real Adaboost algorithm is used to boost these weak classifiers and construct a nesting-structured face detector. To make it rotation invariant, we divide the whole 360-degree range into 12 sub-ranges and construct their corresponding view based detectors separately. To improve performance, a pose estimation method is introduced and results in a processing speed of four frames per second on 320×240 sized image. Experiments on faces with 360-degree in-plane rotation and ±90-degree out-of-plane rotation are reported, of which the frontal face detector subsystem retrieves 94.5% of the faces with 57 false alarms on the CMU+MIT frontal face test set and the multi-view face detector subsystem retrieves 89.8% of the faces with 221 false alarms on the CMU profile face test set.

1. Introduction

Multi-view face detection (MVFD) is used to detect upright faces in images that with ±90-degree rotation-out-of-plane (ROP) pose changes. Rotation invariant means to detect faces with 360-degree rotation-in-plane (RIP) pose changes. The introduction of boosted cascade face detector by Viola and Jones [2], stimulated great interest in the development of more general systems. Their system, based on integral image and simple features, promised very high speed and performance comparable to the best of previous systems [3]; this prompted the development of more general systems such as rotation invariant frontal face detection [4] and MVFD [5]. In order to meet the needs of various applications, a real-time rotation invariant MVFD system is the ultimate goal. Although present MVFD methods can be applied to this problem by means of rotating images and repeating the procedure, the process becomes time consuming and the rate of false alarm increases. Although initially this problem may seem trivial, after careful examination it is found that it involves more complexity when compared to previous problems.

Existing work on MVFD includes Schneiderman et al.'s work [6] based on Bayesian decision rule and. Li et al.'s [5] pyramid-structured detector which was reported as the first real-time MVFD system. In this paper, we propose a novel method for rotation invariant MVFD based on Schapire and Singer's improved boosting algorithms [1] that use real-valued confidence-rated weak classifiers. We call it *Real Adaboost* in order to distinguish it from what we call *Discrete Adaboost*, that is the original Adaboost algorithm [7] adopted in [2][5] using Boolean weak classifier. The main contributions of this paper are: 1) Look-Up-Table (LUT) type weak classifier is proposed and Real Adaboost is used to boost them. 2) A novel nesting-structured detector is proposed and a corresponding pose estimation method for improving overall performance is developed. 3) A fast multi-view face detection system which can deal with profile and 360-degree rotated faces based on the above structure.

The rest of the paper is organized as follows: Section 2 introduces the Real Adaboost algorithm; Section 3, the Haar feature based LUT weak classifiers; Section 4, the view-based nesting-structured cascade; Section 5, the view-based detectors; Section 6, the pose estimation method; Section 7, the experiment results; Section 8 our conclusions and Section 9 acknowledgements.

2. Real Adaboost

Boosting algorithms can improve the performance of a weak learner *L* by iteratively calling it to find a small number of weak classifiers *h* and then combining them into a strong one *H*. Adaboost is an adaptive boosting algorithm in which the rule for combining the weak classifiers adapts on the problem. Real Adaboost algorithm deals with a confidence-rated weak classifier that is a map from a sample space $\mathcal{X}$ to a real-valued space $\mathcal{R}$ instead of Boolean prediction, it has the following form [1]:

- Given dataset $S=\{(\mathbf{x}_1,y_1),...,(\mathbf{x}_m,y_m)\}$, where $(\mathbf{x}_i, y_i) \in \mathcal{X} \times \{-1,+1\}$, the weak classifier pool $\mathcal{H}$ and the number of weak classifiers to be selected *T*.

- Initialize the sample distribution $D_1(i) = 1/m$.
- For $t = 1,\ldots,T$

  1. For each weak classifier h in $\mathcal{H}$ do:

  a. Partition $\mathcal{X}$ into several disjoint blocks $X_1,\ldots,X_n$.

  b. Under the distribution D_t calculate

$$W_l^j = P(\mathbf{x}_i \in X_j, y_i = l) = \sum_{i:\mathbf{x}_i \in X_j \wedge y_i = l} D_t(i) \qquad (1)$$

  where $l = \pm 1$.

  c. Set the output of h on each X_j as

$$\forall \mathbf{x} \in X_j,\ h(\mathbf{x}) = \frac{1}{2}\ln\left(\frac{W_{+1}^j + \varepsilon}{W_{-1}^j + \varepsilon}\right) \qquad (2)$$

  where ε is a small positive constant.

  d. Calculate the normalization factor

$$Z = 2\sum_j \sqrt{W_{+1}^j W_{-1}^j} \qquad (3)$$

  2. Select the h_t minimizing Z, i.e.

$$Z_t = \min_{h\in\mathcal{H}} Z$$
$$h_t = \arg\min_{h\in\mathcal{H}} Z \qquad (4)$$

  3. Update the sample distribution

$$D_{t+1}(i) = D_t(i)\exp\left[-y_i h_t(\mathbf{x}_i)\right] \qquad (5)$$

  and normalize D_{t+1} to a p.d.f.
- The final strong classifier H is

$$H(\mathbf{x}) = \text{sign}\left[\sum_{t=1}^{T} h_t(\mathbf{x}) - b\right] \qquad (6)$$

  where b is a threshold whose default is zero. The confidence of H is defined as

$$Conf_H(\mathbf{x}) = \left|\sum_t h_t(\mathbf{x}) - b\right| \qquad (7)$$

It can be seen that Eq.1 and Eq.2 define the output of the weak classifier, so all that is left to the weak learner is to partition the domain $\mathcal{X}$.

3. Haar Feature based LUT Weak Classifier

Haar features are simple rectangle features proposed by Viola et al. [2]. For each Haar feature, one weak classifier is trained. In [2], threshold-type weak classifiers are used which output Boolean values, i.e. $h(\mathbf{x})$ = sign[$f_{Haar}(\mathbf{x})$-b], where f_{Haar} is the Haar feature and b is a threshold, see Fig.1.a. The main disadvantage of the threshold model is that it is too simple to fit complex distributions. Furthermore, in order to use Real Adaboost, we need a weak learner L which can give a partition of $\mathcal{X}$. Therefore we propose a real-valued LUT weak classifier illustrated in Fig.1.b. Assuming f_{Haar} has been normalized to [0, 1], the range is divided evenly into n sub-ranges:

$$bin_j = [(j\text{-}1)/n, j/n), j = 1,\ldots,n \qquad (8)$$

A partition on the range corresponds to a partition on $\mathcal{X}$. Thus, the weak classifier can be defined as

$$\text{If } f_{Haar}(\mathbf{x}) \in bin_j \text{ then } h(\mathbf{x}) = \frac{1}{2}\ln\left(\frac{\overline{W}_{+1}^j + \varepsilon}{\overline{W}_{-1}^j + \varepsilon}\right) \qquad (9)$$

where

$$\overline{W}_l^j = P\left(f_{Haar}(\mathbf{x}) \in bin_j, y = l\right), l = \pm 1, j = 1, \quad ,n \quad .$$

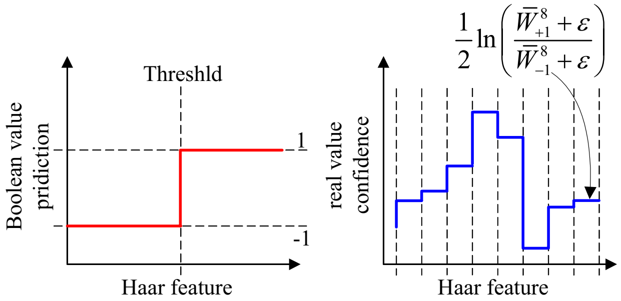
Given the characteristic function

$$B_n^j(u) = \begin{cases} 1 & u \in \left[j-1/n, j/n\right) \\ 0 & u \notin \left[j-1/n, j/n\right) \end{cases}, j = 1, \quad ,n \qquad (10)$$

the LUT weak classifier can be formally expressed as:

$$h_{LUT}(\mathbf{x}) = \frac{1}{2}\sum_{j=1}^{n}\ln\left(\frac{\overline{W}_{+1}^j + \varepsilon}{\overline{W}_{-1}^j + \varepsilon}\right)B_n^j\left(f_{Haar}(\mathbf{x})\right) \qquad (11)$$

In our experiments, the LUT size n = 64.



(a) (b)

Figure 1. Discrete threshold and real LUT weak classifiers

4. Nesting-Structured Cascade

The cascade-structured classifier of Viola et al. [2] has been proved very efficient for many object detection problems. Nevertheless there is still room to enhance its efficiency. We noticed that although each layer of the cascade is trained by Adaboost to consist of closely coupled features, successive layers are rather loosely correlated. More specifically, it is normal for the predecessor to be considered a binary function by its successor because it only tells whether or not the current sub-window is a face (Eq.6). It is a great waste that the confidence value in Eq.7 is omitted since this value can be used to efficiently classify all preceding training samples. This implies that the confidence given by the predecessor can be a strong and efficient feature value to help its successor classify newly collected samples. Therefore, we propose a novel nesting-structured cascade, in which each layer is considered as not only an independent node of the cascade classifier but also as a component of its successor. During training, after each layer is learned, first the layer bootstraps the sample set, and then it is taken as the first weak classifier of its successor in a new boosting round. In this way, the classification ability of each layer is inherited from its predecessor.

The nesting-structured cascade, shown in Fig.2, is made of both common Haar-feature-based weak classifier and the newly developed nested weak classifier; within these two components the only difference lies in the type of feature value. In the Haar feature based weak classifier, the feature value is the Haar feature itself, while in nested weak classifier it takes the confidence value (defined in Eq.7) output by its predecessor as its feature value. Compared with the neural network structure, each layer can be regarded as a linear network of Haar feature based weak classifiers except for the first node in layer two and higher. In this network, the input value of each node is the Haar feature value or the output of the nodes in the previous layer, the LUT weak classifier becomes a more precise approximation of the distributions of training samples than the threshold function.

As expected, the Real Adaboost algorithm performs better than the Discrete Adaboost algorithm in our nesting-structured cascade since the Real Adaboost algorithm can output a more continuous confidence value than Discrete Adaboost, especially in the first few layers. The reason for this is that they consist of only a few weak classifiers, and accordingly our LUT weak classifier functions especially well with the Real Adaboost algorithm.

The training algorithm is as follows:

```
● Given the maximum false positive rate per layer $f$, the minimum detection rate per layer $d$, the overall false positive rate $F_{target}$, the positive training set $P$ and the negative training set $N$.
● Denote the $i$-th layer with $L_i$, and the nesting detector with $ND$
● Initialize: $F_{all} = 1.0$, $D_{all} = 1.0$, $ND = 0$, $i = 0$
● While $F_{all} > F_{target}$
    1. $i = i + 1$
    2. If $i > 1$
        i) take $L_{i-1}$ as the first weak classifier of $L_i$, and update the sample distribution.
        ii) Use Real AdaBoost algorithm to learn the remaining weak classifiers of $L_i$, until $L_i$ satisfies the condition $f$ and $d$.
       else
           Use Real AdaBoost to learn $L_i$ directly, until $L_i$ satisfies the condition $f$ and $d$.
    3. Evaluate the actual false positive rate $F_i$ and detection rate $D_i$ of $L_i$
    4. $F_{all} = F_{all} \times F_i$, $D_{all} = D_{all} \times D_i$
    5. $ND = ND + L_i$
    6. Resample the negative training set $N$ with $ND$
● $ND$ is the final nesting detector with false positive rate $F_{all}$ and detection rate $D_{all}$
```

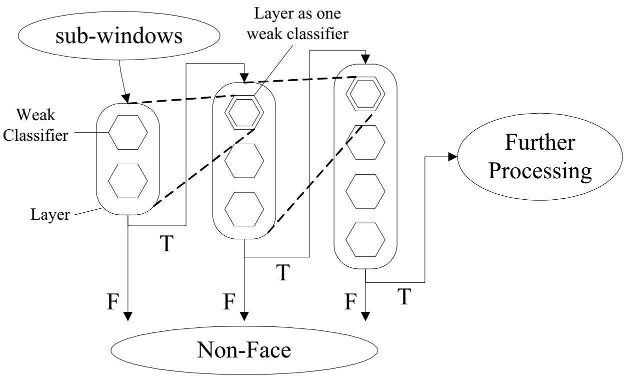


Figure 2. Nesting-structured cascade

5. View Based Detector

According to the left-right ROP angle, multi-view faces are divided into five categories, full left profile, half left profile, frontal, half right profile and full right profile, covering [-90°, -50°], [-50°, -20°], [-20°, +20°], [+20°, +50°], [+50°, +90°] respectively. According to the RIP angle, faces are divided into twelve categories, each covering 30 degrees. So there are in total 5×12 view categories corresponding to 60 detectors. Since the Haar features can be flipped horizontally and rotated 90°, only eight detectors have to be trained and all the others can be generated from the original ones, see Fig.3.

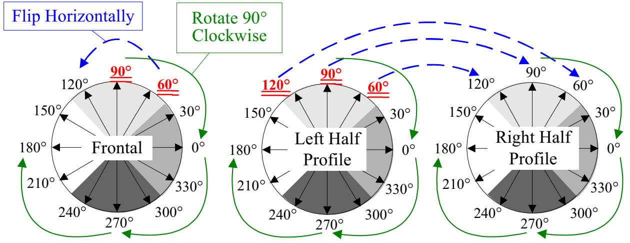


Figure 3. Flip and rotate detectors. (The original detectors of frontal view are the 60-degree and 90-degree ones. The original detectors of half profile view are the left half profile, that is, the 60-degree, 90-degree and 120-degree ones. Full profile situation is the same as half profile.)

6. Pose Estimation

During face detection, the input image is scanned by all view-based detectors and the outputs are merged. The scanning procedure is an exhaustive search, as millions of sub-windows will be involved. This is very time-consuming. To improve performance, we propose a pose estimation method based on the afore mentioned nesting structure. Suppose there are d view-based detectors and each of them has n layers. To denote the confidence of the j-th layer of the i-th detector $Conf_j^{(i)}$, i=1,…,d,

0-7695-2122-3/04 $ 20.00 © 2004 IEEE

j=1,…,*n*, the confidence of the first *k* layers can be defined as:

$$Conf^{(i)}_{[1,k]}(\mathbf{x}) = \begin{cases} \prod_{j=1}^{k} Conf^{(i)}_{j}(\mathbf{x}) & \mathbf{x} \text{ passes the first } k \text{ layers} \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

If the view category is represented by the detector's index, then the pose estimation method with first *m* layers is formally expressed as:

$$pose(\mathbf{x}) = \arg\max_{1\le i\le d}\left[conf^{(i)}_{[1,m]}(\mathbf{x})\right] \quad (13)$$

In our method, pose estimating is not separate from face detecting therefore it will not introduce extra computation. This is superior to Rowley's method [2]. Fig.4 illustrates the pose estimation procedure.

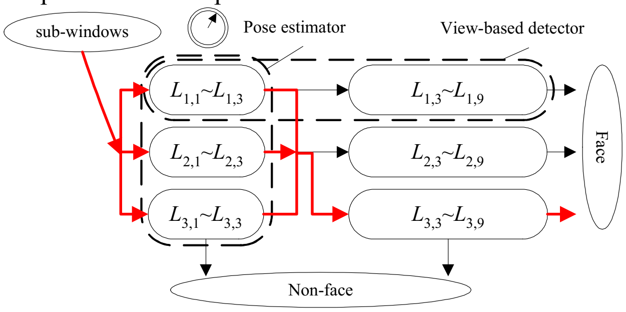


Figure 4. Pose estimation. (There are three view-based detectors, each has nine layers, and the first three layers are used to estimate the pose of face. The bold line is a possible path of an input window.)

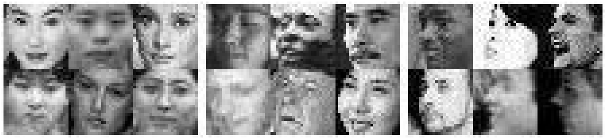


Figure 5. Standard samples. (the left six are frontal faces, the center six are left half profile faces and the right six are full profile faces.)

7. Experiments

We have collected approximately 44,000 faces which include all ethnicities, a variety of ages and both genders, of which the samples are partitioned into five ROP view-based sub-categories as illustrated in Section 5. Each category also covers [-15°, +15°] RIP changes and [-30°, +30°] up-down ROP changes. All the samples are normalized to the standard 24×24 pixel patch, see Fig.5.

7.1. Frontal Face Detection

In our face dataset there are about 20,000 frontal faces. With these samples, we have trained a nesting detector that has 16 layers and 756 Haar features. Compared to the Viola's cascade detector [2] that has 32 layers and 4297 features, our method is much more efficient. We have tested the frontal detector on CMU+MIT frontal face test set. Table 1 shows the results, Fig.6 is the ROC curve and Fig.7 shows some detection results. The running time of frontal face detection is about 18 ms for a 320×240 image on a Pentium 4 2.4 GHz PC.

Table 1. Frontal face detection results on CMU+MIT frontal face test set (130 images, 507 faces)

False Alarm / Method	3	10	13	31	50	57	95	213	422
Ours	89.0%	90.1%	90.7%	-	-	94.5%	-	96.5%	-
Viola-Jones	-	78.3%	-	85.2%	88.8%	-	90.8%	-	93.7%
Rowley	-	83.2%	-	86.0%	-	-	89.2%	-	89.9%

Table 2. Multi-view face detection results on CMU profile face test set (208 images, 441 faces, 347 of them are non-frontal)

False Alarm / Method		8	12	34	89	91	109	221	415	700
Ours	With PE	79.4%	-	84.8%	-	-	87.8%	89.8%	-	-
	Without PE	-	-	84.1%	86.2%	-	-	-	91.3%	-
Schneiderman		-	75.2%	-	-	85.5%	-	-	-	92.7%

7.2. Multi-View Face Detection

Our dataset contains about 10,000 half profile and 14,000 full profile faces. The first six layers of the nesting detector are used to estimate the face pose. We have tested our multi-view detectors on CMU profile face test set [6]. Schneiderman et al. [6] reached a 85.5% pass rate with 91 false alarms on this set. Li et al. [5] also tested their pyramid detector on this set, but did not report any statistic on the results. For the faces in the set that are all nearly upright, we only use the detectors with 60-degree, 90-degree and 120-degree RIP angle when testing. The results of our MVFD are shown in Table 2 and some detection results are in Fig.8. The running time of MVFD is about 80 ms for a 320×240 image on a Pentium 4 2.4 GHz PC. Using pose estimation can increase performance by a speedup factor of 1.7.

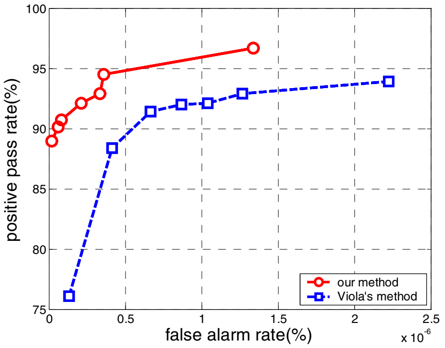


Figure 6. ROC of frontal face detection.

7.3. Rotation Invariant Multi-View FD

For rotation invariant MVFD, there is no standard test set available. Therefore we have provided detection results on pictures collected from WWW, see Fig.9. The running time is about 250 ms for a 320×240 image on a Pentium 4 2.4 GHz PC.

8. Conclusions

In this paper, we presented a method for rotation invariant MVFD. The view-based detectors have been trained with an extensive dataset that includes faces collected from real-life photos, which implies our view based detectors are able to work in more general situations.

The main contributions of this paper are: 1) Look-Up-Table (LUT) type weak classifier is proposed and Real Adaboost is used to boost them. 2) A novel nesting-structured detector is proposed and a corresponding pose estimation method for improving overall performance is developed. 3) A fast multi-view face detection system which can deal with profile and 360-degree rotated faces based on the above structure. One thing we want to mention is that in review of the latest ICCV'03 proceedings, we found Xiao et al.'s [8] "boosting chain" while originating from the same concept of inheriting previous training results, it resulted in different techniques, of which theirs is a complete chain structure for Discrete Adaboost framework and ours is a nested structure for Real Adaboost. We argue that a nested structure for Real Adaboost is more powerful, and comparatively much more efficient.

We have tested our method on two public test sets, the CMU+MIT frontal face and CMU profile face test set. The latter is particularly difficult. While the results of previous methods on these sets are far from satisfying, our MVFD system has achieved a high detect rate and an acceptable number of false alarms. This proves that it is possible to use MVFD for real-life conditions.

9. Acknowledgements

This work is supported mainly by a grant from OMRON corporation. It is also supported in part by National Science Foundation of China under grant No.60332010.

10. References

[1] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions", Machine Learning, 37, 1999, 297-336.

[2] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, 2001.

[3] H. A. Rowley, "Neural Network-based Human Face Detection", Ph.D. thesis, Carnegie Mellon University, May 1999.

[4] Shihong Lao, Toshiyuki Kozuru, et al., "A fast 360-degree rotation invariant face detection system", in ICCV2003 Demo program.

[5] S. Z. Li, L. Zhu, Z. Q. Zhang, et al., "Statistical Learning of Multi-View Face Detection". In Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark. May, 2002.

[6] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars". In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA, 2000.

[7] Y. Freund and R. E. Schapire, "Experiments with a New Boosting Algorithm". In Proc. of the 13-th Conf. on Machine Learning, Morgan Kaufmann, 1996, 148-156.

[8] Rong Xiao, Long Zhu, Hongjiang Zhang, Boosting Chain Learning for Object Detection, ICCV 2003, 709–715.

[9] B. Moghaddam and A. Pentlan, “Beyond Linear Eigenspaces: Bayesian Matching for Face Recognition”, Face Recognition from Theory to Application, edited by H. Wechsler et al. Springer 1998, 230-243.

[10] R. Feraud, O.J. Bernier, Jean-Emmanuel Viallet, and Michel Collobert, “A Fast and Accurate Face Detector Based on Neural Networks”, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 23, No. 1, 2001, 42-53.

[11] E. Osuna, R. Freund and F. Girosi, “Training Support Vector Machines: an Application to Face Detection”, In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Puerto Rico, 1997.

[12] V. P. Kumar and T. Poggio. “Learning-Based Approach to Real Time Tracking and Analysis of Faces”, MIT A.I. Memo No.1672, Sept, 1999.

[13] Y. Li, S. Gong, and H. Liddell. “Support Vector Regression and Classification based multi-view Face Detection and Recognition”. In IEEE Conf. on Automatic Face and Gesture Recognition, March 2000.

[14] R.E. Schapire, Y. Freund, P. Bartlett and W.S. Lee, “Boosting the margin: A new explanation for the effectiveness of voting methods”, The Annals of Statistics, 26(5), 1998, 1651–1686.

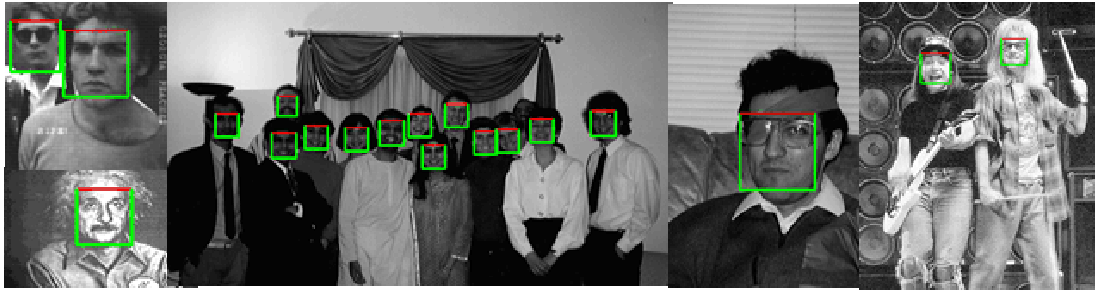


Figure 7. Some frontal face detection results on the CMU+MIT frontal face test set.

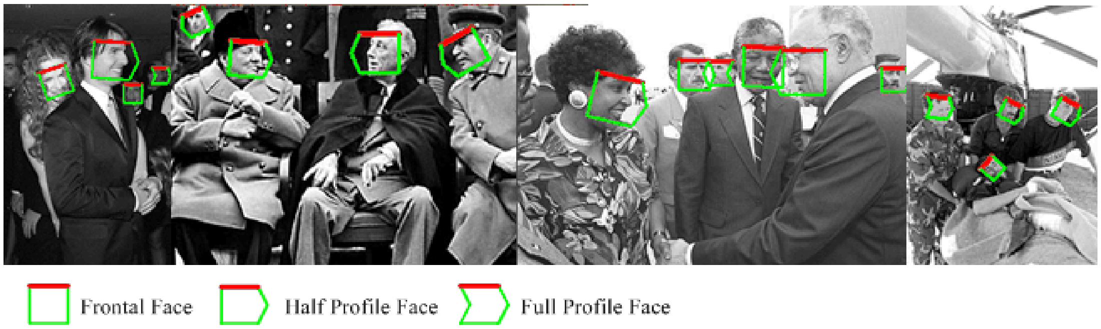


Figure 8. Some MVFD results on the CMU profile face test set.

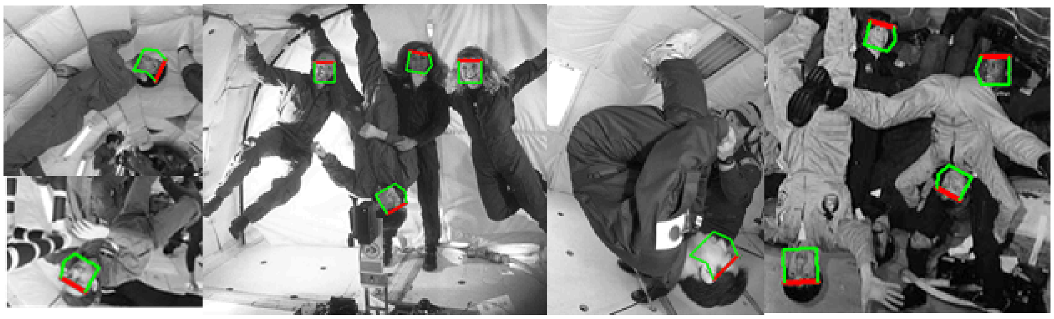


Figure 9. Some Rotation Invariant MVFD results