A Novel Multi-view Face Detection Method Based on Improved Real Adaboost Algorithm

Wenkai Xu[1], Eung-Joo Lee[2]

[1] Department of Information Communications and Engineering, University of Tongmyong
Busan, 608-711 Korea
[e-mail: xwk6298@hotmail.com]
[2] Department of Information Communications and Engineering, University of Tongmyong
Busan, 608-711 Korea
[e-mail: ejlee@tu.ac.kr]
*Corresponding author: Eung-Joo Lee

Abstract

Multi-view face detection has become an active area for research in the last few years. In this paper, a novel multi-view human face detection algorithm based on improved real Adaboost is presented. Real Adaboost algorithm is improved by weighted combination of weak classifiers and the approximately best combination coefficients are obtained. After that, we proved that the function of sample weight adjusting method and weak classifier training method is to guarantee the independence of weak classifiers. A coarse-to-fine hierarchical face detector combining the high efficiency of Haar feature with pose estimation phase based on our real Adaboost algorithm is proposed. This algorithm reduces training time cost greatly compared with classical real Adaboost algorithm. In addition, it speeds up strong classifier converging and reduces the number of weak classifiers. For frontal face detection, the experiments on MIT+CMU frontal face test set result a 96.4% correct rate with 528 false alarms; for multi-view face in real time test set result a 94.7 % correct rate. The experimental results verified the effectiveness of the proposed approach.

Keywords: Multi-view face detection, real Adaboost, Haar feature, pose estimator

1. Introduction

Over the past decade, face recognition has emerged as an active research area in computer vision with numerous potential applications including biometrics, surveillance, human-computer interaction, video-mediated communication and content-based access of images and video databases.

Face detection is the first step in automated face recognition. Its reliability has a major influence on the performance and usability of the entire face recognition system. Given a single image or a video, and ideal face detector should be able to identify and locate all the present faces regardless of their position, scale, orientation, age, and expression. Furthermore, the detection should be irrespective of extraneous illumination conditions and the image and video content. Face detection can be performed based on several cues: skin color, motion, facial or head shape, facial appearance, or a combination of these parameters [1].

Most successful face detection algorithms are appearance-based without using other cues. The processing is done as follow: An input image is scanned at all possible locations and scales by a sub-window. Face detection is posed as classifying the pattern in the sub-window as either face or non-face. The face or non-face classifier is learned from face and non-face training examples using statistical learning methods.

Statistical methods have been widely adopted in face detection [2]. Moghaddam and Pentland [3-6] introduced the Eigenface method, where the probability of face patterns is modeled by the "distance-in-feature-space" (DIFS) and "distance-from-feature-space" (DFFS) criteria. Osuna et al. [7, 8] presented an SVM-based approach to frontal-view face detection. Unlike the Eigenface method where only the positive density is estimated, this approach seeks to learn the boundary between face and non-face patterns. After learning, only the 'important' examples located on the boundary are selected to build the decision function. Soulie et al. [9] described a system using neural networks (NNs) for face detection. They implemented a multi-modal architecture where various rejection criteria are employed to trade-off false recognition against false rejection. Sung and Poggio [10] also presented a NN based face detection system. They designed six positive prototypes (faces) and six negative prototypes (non-faces) in the hidden layer. Supervised learning is performed to determine the weights of these prototypes to the output node. Rowley et al. [11] introduced a NN based upright frontal face detection system. A retinal connected NN examines small windows of an image and decides whether each window contains a face. This work was later extended to rotation invariant face detection by designing an extra network to estimate the rotation of faces in the image plane [12, 13].

Multi-view face detection (MVFD) is used to detect upright faces in images that with ±90-degreee rotation-out-of plane (ROP) pose changes. Rotation invariant means to detect faces with 360-degree rotation-in-plane (RIP) pose changes [14].

Recently, Viola and Jones [15] presented an approach to fast face detection using simple rectangle features which can be efficiently computed from the so-called Integral Image. This system prompted the development of more general systems such as rotation invariant frontal face detection and MVFD [16]. Adaboost and cascade methods are then used to train a face detector based on these features. Li et al. [16, 17] adopted similar but more general features which can be computed from block differences. Also, Floatboost is proposed to overcome the monotonicity of the sequential Adaboost learning. Existing work on MVFD includes Schneiderman et al.'s work [18] based on Bayesian decision rule and Li et al.'s [16]

pyramid-structured detector which was reported as the first real-time MVFD system. In Hongming Zhang et al.'s work [19], they emphasized on designing efficient binary classifiers by learning informative features through minimizing the error rate of the ensemble ECOC multi-class classifier. 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.

In this paper, we propose a novel improved real Adaboost algorithm for MVFD based on Schapire and Singer's improved Adaboost classifiers [20]. LUT weak classifier is used to train processed Haar feature (mirror and rotation) by our real Adaboost algorithm. In addition, a pose estimator based on confidence of strong classifier is proposed in this paper, which used first four layers to estimate face pose. We tested our MVFD method on CMU and MIT face database, the experimental results show us that the performance of our detection system is better than discrete Adaboost and real Adaboost algorithm, in addition, it speeds up strong classifier converging, reduces the number of weak classifiers and decreases detection time. The detection results and its timeliness are satisfactory.

2. Improved Real Adaboost Learning Algorithm

2.1 Real Adaboost Algorithm

For Adaboost learning, a complex nonlinear strong classifier $H_M(x)$ is constructed as a linear combination of M simpler, easily constructible weak classifiers in the following form. The Adaboost learning procedure is aimed at learning a sequence of best weak classifiers to combine $h_m(x)$ and the combining weights α_m in:

$$H_M(x) = \frac{\sum_{m=1}^{M} \alpha_m h_m(x)}{\sum_{m=1}^{M} \alpha_m} \tag{1}$$

It solves the following three fundamental problems: (1) learning effective features from a large features set; (2) constructing weak classifiers, each of which is based on one of the selected features; and (3) boosting the weak classifiers to construct a strong classifier.

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 [20], [21]:

● Given dataset $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$;

Where $(x_1, y_1) \in \mathcal{X} \times \{-1, +1\}$, the weak classifier pool $\mathcal{H}$ and the number of weak classifiers to the selected $\mathcal{T}$.

● Initialize the sample distribution $D_1(i) = 1/m$.

● For $t = 1, \dots, T$

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

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

　　b. Under the distribution D_t calculate

$$W_l^j = P(x_i \in X_j, y_i = l) = \sum_{i:x_i \in X_j \wedge y_i = l} D_t(i) \tag{2}$$

where $l = \pm 1$.

c. Set the output of h on each X_j as

$$\forall x \in X_j, h(x) = \frac{1}{2} ln \left(\frac{W_{+1}^j + \varepsilon}{W_{-1}^i + \varepsilon} \right) \tag{3}$$

where ε is a small positive constant.
  d. Calculate the normalization factor

$$Z = 2 \sum_j \sqrt{\frac{W_{+1}^j}{W_{-1}^i}} \tag{4}$$

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

$$Z_t = min_{h \in \mathcal{H}} Z \tag{5}$$

$$h_t = arg\, min_{h \in \mathcal{H}} Z \tag{6}$$

(3). Update the sample distribution

$$D_{t+1}(i) = D_t(i) exp[-y_i h_t(x_i)] \tag{7}$$

  And normalize D_{t+1} to a probability distribution function.
● The final strong classifier H is

$$H(x) = sign[\sum_{t=1}^T h_t(x)] \tag{8}$$

It can be seen that (2) and (3) define the output of the weak classifier, so all that is left to the weak learner is to partition the domain $\mathcal{X}$.

2.2 Improved Real Adaboost Algorithm

The real Adaboost algorithm proposed by Schapire et al. [20] extended the criterion of discrete two-value to output of continuous confidence. This algorithm selected Equation (3) as the output; all of each weak classifier will have this judgment function as well; the core objective is to confirm a partition on these datasets. So, we do some improvement on partition of datasets and the control of weights adjustment.
● Given dataset $S = \{(x_1, y_1), \dots, (x_m, y_m)\}$;
  Where $(x_1, y_1) \in \mathcal{X} \times \{-1, +1\}$, the weak classifier pool $\mathcal{H}$ and the number of weak classifiers to the selected $\mathcal{T}$, the number of features for classifier is s, we give three weights parameters as $w_{base}, w_{factor}, w_{index}$.
● Initialization
  (1) Initialize the sample distribution.

$$D_1' = \begin{cases} (w_{base} + w_{factor})/m & x_i\, is\, real\, face\, sample \\ 1/m & others \end{cases} \tag{9}$$

  (2) Initialize the Gray value interval distribution of Haar feature.

Preprocess each Haar feature corresponding to weak classifier, divide range evenly into n sub-ranges. A partition on the range corresponds to a partition on $\mathcal{X}$: $X_1{}^k, ..., X_n{}^k$, $k=1,2,...,s$.
● For t = 1, ... , T
(1) For each weak classifier h' in $\mathcal{H}'$ do:
a. Under the distribution D_t' calculate

$$W'^j_l = P(x_i \in X_j^k, y_i = l) = \sum_{i:x_i \in X_j \wedge y_i=l} D_t'(i) \tag{10}$$

where $l = \pm 1$, $k=1,...,s$, $j=1,...,n$.
b. Set the output of h on each X_j under sample weights as:

$$\forall x \in X_j, h'(x) = \frac{1}{2} ln \left(\frac{W'^j_{+1} + \varepsilon}{W'^i_{-1} + \varepsilon} \right) \tag{11}$$

where ε is a small positive constant.
c. Calculate the normalization factor

$$Z' = 2 \sum_j \sqrt{\frac{W^{j'}_{+1}}{W^{i'}_{-1}}} \tag{12}$$

(2). Select the h'_t minimizing Z', i.e.

$$Z'_t = \min_{h' \in \mathcal{H}'} Z' \tag{13}$$

$$h'_t = arg \min_{h' \in \mathcal{H}'} Z' \tag{14}$$

(3). Update the sample distribution

$$D'_{t+1}(i) = D'_t(i)exp[-y_i h_t(x_i) * \beta_t] \tag{15}$$

For training weight value β_t as

$$\beta_t = \begin{cases} 1 & others \\ w_{base} + w_{factor} * d_{rate}^{w_{index}} & d_{rate} \ is \ the \ detection \ rate \end{cases} \tag{16}$$

(4). Normalize D'_{t+1} to a probability distribution function
● The final strong classifier H is

$$H'(x) = sign[\sum_{t=1}^{T} h'_t(x) - b] \tag{17}$$

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

$$Conf_H(x) = \left| \sum_t h'_t(x) - b \right| \tag{18}$$

2.3 Analysis of Improved Real Adaboost Algorithm

Original real Adaboost algorithm used local optimum criteria, the convergence is better than discrete Adaboost algorithm; but the time complexity of real Adaboost algorithm is still large. The number of training samples is defined as N, the number of selecting class features as M, and the number of weak classifiers for strong classifier as T. The time complexity of weak classifier is $O(M * N * N)$, and the time complexity of strong classifier is $O(T * M * N * N)$.

The improved real Adaboost algorithm we proposed used weak classifiers h to divide sample space X into n sub-ranges evenly; each h corresponding to space division $X_1{}^k, ..., X_n{}^k$ can be obtained, and no longer to redistribute training sample space while selecting the best weak classifier each step, by only using the updated weights distribution D_t', and calculating W'^j_l under D_t' by addition. So the complexity for each weak classifier is only depended on the calculated amount of training sample weights by addition statistics. The time complexity of weak classifier is $O(M * N)$, and the time complexity of strong classifier is $O(T * M * N)$, the training speed is enhanced by $O(N)$.

Real Adaboost algorithm is improved by weights adjustment and partition of datasets. Reference [22] had proved that pre-partition on datasets is practicable; we will prove that the control of weights adjustment is as well.

$h_t(x)$ is on the partition $S = S_1 \cup S_2 \cup ... \cup S_n, \forall x \in S_j, h_t(x) = \alpha_t$, the definition of W_l^j is as before. We define a group of stochastic variable as:

$$R_t = \begin{cases} \alpha_j & if \ y = +1 \quad x \in S_j \\ -\alpha_j & if \ y = -1 \quad x \in S_j \end{cases} \quad j = 1, \ldots, n \tag{19}$$

Let $R = \sum_{t=1}^T R_t$.

The strong classifier of real Adaboost algorithm is the sum of all weak classifiers, now we improve it by adding weights as a sum of weights added, the strong classifier is defined as:

$$H(x) = sign\left[\sum_{t=1}^T \beta_t h_t(x)\right] \tag{20}$$

β_t is given as (16), its mean value and variance can be expressed as:

$$\mu_t = \frac{1}{2}\sum_{j=1}^n ((W_+^{jt} - W_-^{jt})ln(\frac{W_+^{jt}}{W_-^{jt}})) \tag{21}$$

$$\sigma_t^2 = \sum_{j=1}^n \{W_+^{jt}[\frac{1}{2}ln\left(\frac{W_+^{jt}}{W_-^{jt}}\right) - \mu_t]^2 + W_-^{jt}[\frac{1}{2}ln\left(\frac{W_+^{jt}}{W_-^{jt}}\right) + \mu_t]^2\} \tag{22}$$

Let $R = \sum_{t=1}^T \beta_t R_t$, which mean value and variance are μ and σ_t^2, the theorem below could be proved.

THEOREM: When $h_t(x)$ is independent, $\mu > 0$ and combination coefficients $\beta_t = \mu_t/\sigma_t^2$,

the error rate of $H(x)$ is $\varepsilon \leq 1/\sum_{t=1}^{T}(\mu_t^2/\sigma_t^2)$; when T is very large, if μ_t^2/σ_t^2 is bounded but $\sum_{t=1}^{T}(\mu_t^2/\sigma_t^2)$ is large, these combination coefficients approximately are the best.

PROOF: The error rate of $H(x)$ can be regarded as the probability of $R \leq 0$, namely $\varepsilon = P_r[R \leq 0]$. As $H(x)$ has a error if and only if $y f(x) \leq 0$, $yf(x) = \sum_{t=1}^{T} y_t h(x)$, and it is equivalent to $\sum_{t=1}^{T} R_t \leq 0$, so $\varepsilon = P_r[R \leq 0]$.

We write the probability density function of R as $g(r)$, so:

$$\varepsilon = P_r[R \leq 0] = \int_{-\infty}^{0} g(r)dr$$

$$\leq \int_{-\infty}^{0} (r-\mu)^2/\mu^2 g(r)dr \leq \int_{-\infty}^{+\infty} (r-\mu)^2/\mu^2 g(r)dr$$

$$= \sigma^2/\mu^2 = \left(\sum_{t=1}^{T} \beta_t^2 \sigma_t^2\right) \Big/ \left(\sum_{t=1}^{T} \beta_t\mu_t\right)^2$$

And

$$\left(\sum_{t=1}^{T} \beta_t\mu_t\right)^2 = \left(\sum_{t=1}^{T} \beta_t \sigma_t \frac{\mu_t}{\sigma_t}\right)^2$$

$$\leq \sum_{t=1}^{T} (\beta_t\mu_t)^2 \sum_{t=1}^{T} \frac{\mu_t^2}{\sigma_t^2} \tag{23}$$

Namely σ^2/μ^2 can obtain the minimum when $\beta_t = \mu_t/\sigma_t^2$, input (21) into Equation (20), we can obtain:

$$\varepsilon \leq 1/\sum_{t=1}^{T} (\mu_t^2/\sigma_t^2) \tag{24}$$

So far, first half of the the theorem is proved. Next step, we will prove that when $T \to \infty$, the minimum point of σ^2/μ^2 also is the minimum point of $\varepsilon = P_r[R \leq 0]$. When $\beta_t = \mu_t/\sigma_t^2$, both of the mean value and variance of $\beta_t R_t$ are μ_t^2/σ_t^2, we write it as $\theta = \sum_{t=1}^{T}(\mu_t^2/\sigma_t^2)$, and then $\sigma^2/\mu^2 = 1/\theta$. As μ_t^2/σ_t^2 is bounded, on the basis of limit theorem; when $T \to \infty$, $R/T = 1/T \sum_{t=1}^{T} \beta_t R_t$ is the normal distribution with mean value $= \theta/T$ and variance $= \theta / T^2$, so $(-R/T + \theta/T)/(\sqrt{\theta}/T) = (-R + \theta)/\sqrt{\theta}$ is standard normal distribution. For stochastic variable Y, which is belonging to standard normal distribution, when $v \to \infty$, $P_r[Y \geq v] \approx (v\sqrt{2\pi})^{-1} exp(-v^2/2)$ is the monotone function about v. When $T \to \infty$ and $\theta \to \infty$,

$$\varepsilon = P_r[R \leq 0] = P_r\left[(-R + \theta)/\sqrt{\theta} \geq \sqrt{\theta}\right]$$
$$\approx exp(-\theta/2)\,(\sqrt{2\pi\theta})^{-1} \tag{25}$$

So ε is minimal when $\sigma^2/\mu^2 = 1/\theta$ is minimal.

This shows that the ratio of mean-variance could be regarded as a normalization, because that of $y_i\beta_t h_t(x_i)$ is 1; the theorem indicates that the confidence of weak classifiers are better under normalization condition. The weight should adjusted by

$$\omega_i^{t+1} = \omega_i^t/Z_t exp(-y_i\beta_t h_t(x_i)) \tag{27}$$

3. Haar Feature for Multi-View Face Detection

The rectangular masks used for visual object detection are rectangles tessellated by black and white smaller rectangles. Those masks are designed in correlation to visual recognition tasks to be solved, and known as Haar-like wavelets. By convolution with a given image they produce Haar-like features. Viola [15] has used four features (**Fig. 1**(a)~(d)) for face detection and these features performed well on face. Fasel [23] has proposed a new ones (see **Fig. 1**(e)).

(a)   (b)   (c)   (d)   (e)
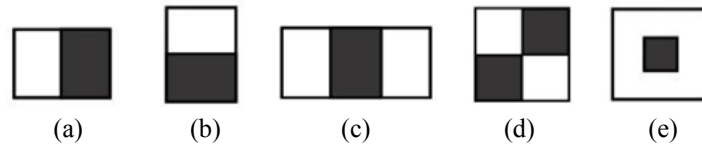
Fig.1. Face detection feature that proposed by Viola and Fasel; (a) , (b) edge features, (c) Line feature, (d) diagonal feature, (e) center surround features.

This study aims to solve multi-view face detection (MVFD), namely to detect upright faces in images that with $\pm 90^{\circ}$ rotation-out-of-plane (ROP) changes and rotation invariants, which means to detect faces with 360° rotation-in-plane (RIP) pose changes. According to ROP angle, 180-degree is divided into $[-90^{\circ},-75^{\circ}]$, $[-75^{\circ},-45^{\circ}]$, $[-45^{\circ},-15^{\circ}]$, $[-15^{\circ},+15^{\circ}]$, $[+15^{\circ},+45^{\circ}]$, $[+45^{\circ},+75^{\circ}]$, $[+75^{\circ},+90^{\circ}]$. According to the RIP angle, faces are divided into twelve categories, each covering 30 degree. So there are total 7*12 view categories corresponding to 84 detectors. Based on Adaboost algorithm, 84 categories face detectors should be designed for 84 categories multi-view faces, it is an enormous work for training. Since the Haar features can be flipped horizontally and rotated 90 degree, some of others can be generated from the original ones, see **Fig. 2** and **Table 1**.

Table 1. Geometric Transformation of Haar Feature ("R" means Rotate 90° and "M" means Mirror Transform)

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, which are the 60-degree, 90-degree and 120-degree ones. Full profile situation is the same as half profile.

According to the analysis above, it should build 7*12=84 categories classifiers originally, but it only needs 4*3=12 categories classifiers now. "4" means four angle regions of seven regions based on flipping horizontally, which cannot be obtained by transformation, and "3" means containing at least 0°, 30° and 60° in twelve categories according to RIP angle. Other situations can be obtained by transformation of Haar feature
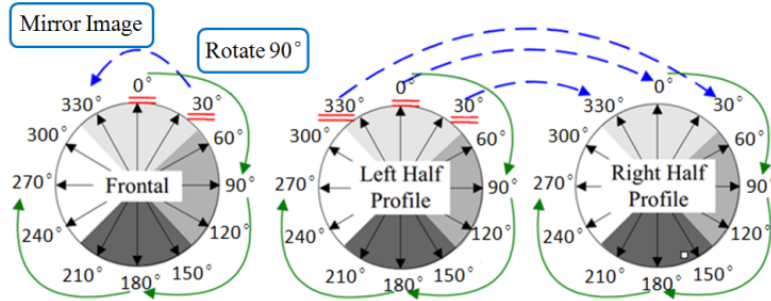
as we introduced.

Fig. 2. Mirror and rotate detectors.

Adaboost algorithm is a learning algorithm that selects a set of weak classifier from a large hypothesis space to construct a strong classifier. Basically speaking, the performance of the final strong classifier originates in the characteristics of its weak hypothesis space. In [12], threshold weak classifiers are used as the hypothesis space input to boosting procedure, the main disadvantage of the threshold model is too simple to fit complex distributions. So, in this paper we use LUT weak classifier [24] in order to use real Adaboost algorithm. Therefore we use a real-valued LUT weak classifier.

Assuming f_{Haar} is the Haar feature and f_{Haar} has been normalized to [0, 1], the range is divided evenly into n sub-ranges, then the j-th LUT item corresponding to a sub-range:

$$bin_j = [(j-1)/n, j/n), j = 1, \dots, n \tag{28}$$

A partition on the range corresponds to a partition on $\mathcal{X}$. Thus, the weak classifier can be defined as: if $f_{Haar}(x) \in bin_j$, then

$$h(x) = \frac{1}{2} ln(\frac{\overline{W}_{+1}^j + \varepsilon}{\overline{W}_{-!}^j + \varepsilon}) \tag{29}$$

Where $\overline{W}_l^j = P(f_{Haar}(x) \in bin_j, y = l), l = \pm 1, j = 1, \dots, n$. Given the characteristic function

$$B_n^j(u) = \begin{cases} 1 & u \in [j - 1/n, j/n) \\ 0 & otherwise \end{cases} \tag{30}$$

Where $j = 1, \dots, n$.
The LUT weak classifier can be formally expressed as:

$$h_{LUT}(x) = \frac{1}{2} ln(\frac{\overline{W}_{+1}^j + \varepsilon}{\overline{W}_{-1}^j + \varepsilon}) B_n^j(f_{Haar}(x)) \tag{31}$$

LUT classifier almost can simulate any kinds of probability distribution. When size of sample sub-window is confirmed, all of candidate Haar feature will be confirmed as well, each Haar feature corresponds to one LUT weak classifier, which make up our weak classifiers space $\mathcal{H}$.

Each layer of cascade is strong classifier training by improved real Adaboost algorithm we proposed. We set up a threshold of each layer as b as Equation (17), let 99.9% of face can pass it, and try best to throw out counterexample. The classifiers on the later place are more complexity, which contains more LUT-based classifiers; therefore, they have stronger classification capability.

The improved real Adaboost algorithm performs better than the Discrete Adaboost algorithm in our cascade classifier since the real Adaboost algorithm can output a more continuous confidence value (as Equation (18)) than Discrete Adaboost algorithm, and our LUT weak classifier functions especially well with the Real Adaboost algorithm. The training algorithm can be explained as follows:

● Setting the maximum false positive rate per layer as f, the minimum passing rate per layer as d, and the overall false positive rate as F_{target}. Given the positive training set as Pos, and the negative training set as Neg.
● Initialize: $F_{all} = 1$, $i = 1$.
● While $F_i > F_{target}$
  1) Training i-th layer by using Pos and Neg; setting threshold value b to let false positive rate f_i lower than f, and let passing rate higher than d.
  2) $F_{i+1} = F_i \times f_i$; $i = i + 1$; $Neg \leftarrow \infty$.
  3) If $F_{i+1} > F_{target}$, scanning non-face images using current cascade classifier, collecting all the false positive set Neg.

4. Pose Estimation based on Cascade Classifier

Ideally, it needs to scan all of candidate sub-windows by trained face detector in linear detection process. But it will take too long time for performing in real time. So we present a strategy for improving timeliness-pose estimation.

Rowley firstly presented pose estimation in his multi-view face detection based on ANN [25], he designed some sub-classes according to different face poses, and then to train an ANN for these classes, called estimator. When the system decides whether sub-windows belong to face or not in the input image, it should estimate the pose firstly using pose estimator and then sent these sub-windows to corresponding face detectors for processing again. Obviously, the pose estimator has a direct influence on detection results.

In our study, we do not train a pose estimator separately, instead to use the confidence of strong classifier, which defined by Equation (18) as before. Suppose there are *d* detectors multi-view-based and each detector has *n* layers. Writing the confidence of the *j-th* layer of *i-th* detector as $Conf_i^{(j)}$, *i=1,...,d, j=1,...,n*. The confidence of the first *k* layers can be expressed as:

$$Conf_i^{[1,k]} = \begin{cases} \prod_{j=1}^{k} Conf_i^{(j)} & x\ pass\ the\ first\ k\ layers \\ 0 & otherwise \end{cases} \qquad (32)$$

Then the pose estimator of first *m* layers can be defined as:

$$\widehat{pose} = arg\ max_{1 \le i \le d}[conf_i^{[1,m]}(x)] \qquad (33)$$

For this processing, x will be regarded as non-face if it had been thrown out at first m layer, and it needn't pose estimate again. Unlike Rowley's method, in our method, pose estimator is not separate from face detecting; therefore it will not introduce extra computation problem. **Fig. 3** illustrates the pose estimation procedure we proposed.
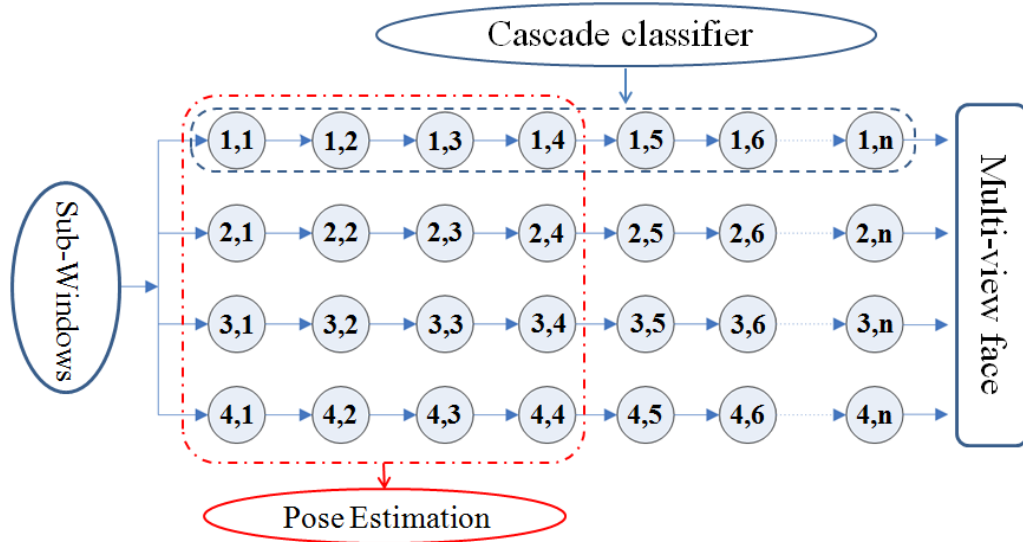

Fig. 3. Pose estimation frame

In this study, there are four view-based detectors; cascade classifier contains 16 layers and first four layers are used to estimate the pose of face.

5. Experimental Results and Analysis

The multi-view face detection system we designed is running on the hardware environment of Intel (R) Core (TM) 2 (2.93GHz), a Web camera, and the software environment of Windows 7 and Visual Studio 2008.

Our face dataset include MIT-CBCL database and CMU face database, which contains frontal faces and profile faces.

5.1 Comparison of Convergence on Adaboost algorithms

A. Experiment for Convergence of Single Strong Classifier

To compare the convergence performance of original real Adaboost and improved real Adaboost algorithm in this paper, we design an experiment on single strong classifier; training set includes 3500 faces and 10000 non-faces, when the detection rate achieve 99.9%, training will be stopped. The coefficients of dynamic weights of improved real Adaboost algorithm are set as: $w_{base} = 1.5, w_{factor} = 0.5, w_{index} = 2.0$, experimental result is shown as **Fig. 4**; the "Detection Rate" is the data of the training set.

The performance curve in **Fig. 4** shows that, the convergence performance of improved real Adaboost algorithm is better than original real Adaboost algorithm. We can find that our real Adaboost has met the stopping criterion when it reached the 31 layer weak classifiers, however original need to 67 layers.
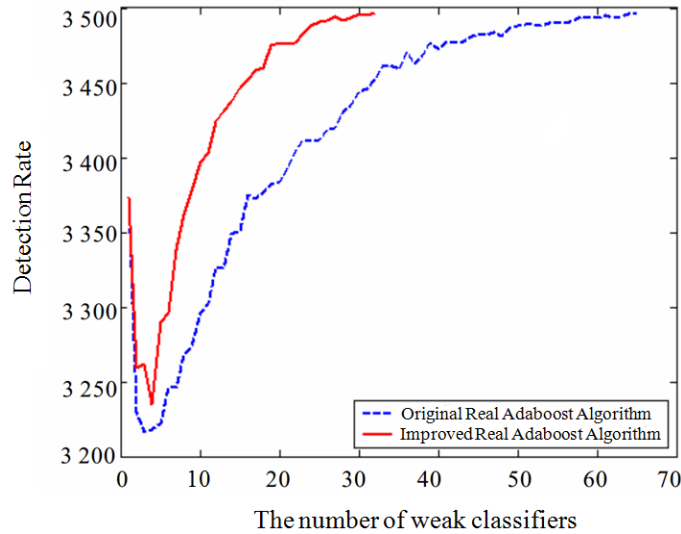
Fig. 4. Comparison of two Adaboost algorithms

B. Experiment for Cascade Classifier

Under the same training condition, we do experiments on Discrete Adaboost, original real Adaboost and ours for testing the performance of these three cascade classifiers. 3500 face pictures with resize of 20×20 are selected for training set, and we use bootstrapping method to select 10000 non-face samples from 500 pictures, which are without human face. 127 face pictures (include 332 faces) are used for testing set. Each algorithm use the same 5000 Haar-like features, Table 2 shows us the detail of performance comparison.

Table 2. Performance Comparison of Three Adaboost Algorithms

Algorithm	Discrete-Adaboost	Real Adaboost	Proposed Algorithm
Number of Weak Classifiers	4297	1784	742
Number of Layers	62	9	16
Average Time (ms)	61	23	15
Positive Rate	198/332	227/332	273/332
False Positive	109	80	83

We can find that from **Table 2**: our algorithm has better performance than Discrete Adaboost algorithm, and detection time is shorter than original real Adaboost algorithm by 8ms. The number of training layers is more than original real algorithm, because the stopping criterion is based on detection rate, each layer can reach high detection rate by using less weak classifiers; since the number of rejecting as non-face is few, more layers are necessary to reject all the non-faces. Moreover, the number of weak classifiers is less than original real Adaboost algorithm by 1042 in this paper.

5.2 Experiments of Multi-View Face Detection

A. Frontal Face Detection

We use the CMU frontal face database and MIT face database [26] which consists of a

training set with 6977 images (2429 faces and 4548 non-faces) and a test set with 24045 images (472 faces and 23573 non-faces). The images are 19×19 grayscale and we renormalize them to 20×20. With these samples, we trained them by improved real Adaboost algorithm that has 16 layers and 742 weak classifiers. Compared to the Viola's method that has 62 layers and 4297 features and Schneiderman's method [18], our method is much more efficient. **Fig. 5** is the ROC curve and **Fig. 6** show us face detection results. Obviously, our method has better convergence performance on training set.
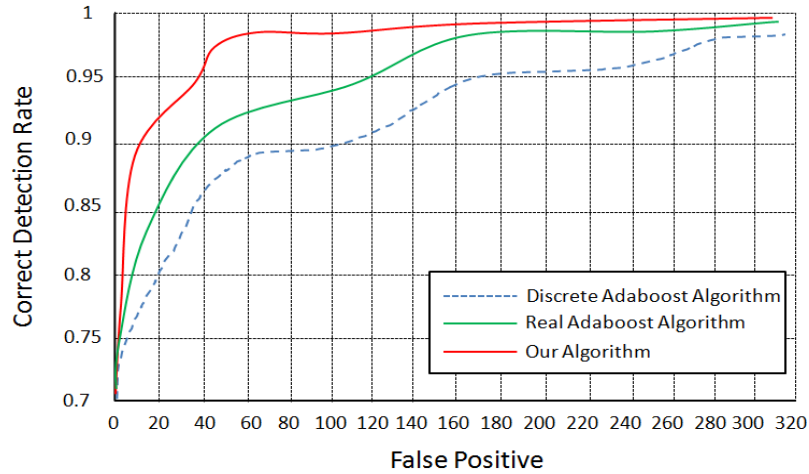

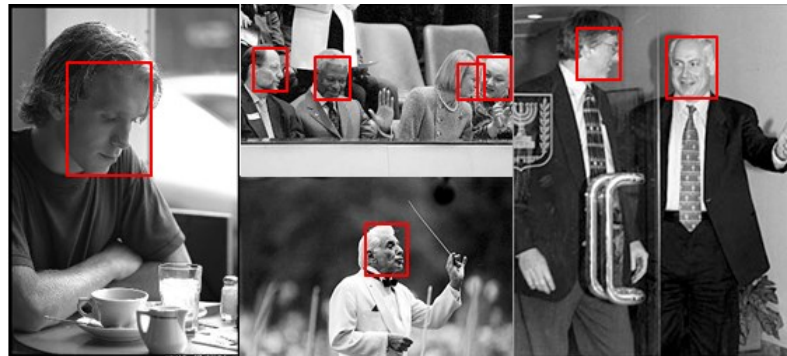**Fig. 5.** The ROC curves of our detectors on the MIT frontal face test set

Fig. 6. Frontal face detection results on CMU+MIT test set

B. Multi-view Face Detection

We tested the profile detector on CMU profile face test set with 208 images (441 faces and 307 of them are non-frontal face). For the training set, about 34000 normalized images (about 8000 half profile faces, 10000 full profile faces and 16000 frontal faces) are selected on Internet. According to the analysis in this paper, only 12 categories classifiers are trained for multi-view face detection with ROP and RIP angle. **Table 3** and **Fig. 7** show the experimental results of multi-view face detection.

Table 3. Multi-view face detection results on CMU profile face test set.

Algorithm / False Drop	Improved Real-Adaboost		Real-Adaboost
	With PE	Without PE	
5	74.3%	71.1%	—
12	—	—	75.2%
28	87.7%	87.1%	—
76	—	93.7%	—
91	—	—	85.5%
153	94.4%	—	—
221	—	94.7%	—
528	96.4%	95.9%	—
700	—	—	92.7%

Fig. 7. Multi-view face detection results on CMU profile face test set.

In reference [18] reached 85.5% pass rate with 91 false drops, but we can find that the performance of our method is much better than Schneiderman's method; it can reach a high detection rate with less false drop. On the other hand, the first layers of cascade classifiers are used to estimate the face pose, the results with PE is better than that without PE as **Fig. 8** shown.
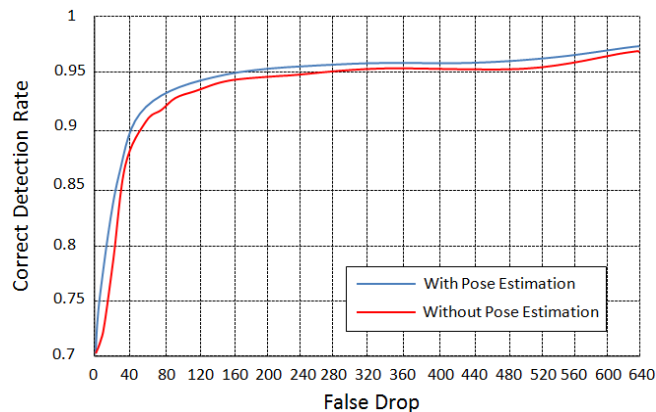
Fig. 8. The ROC curves of our algorithm on CMU profile face data set

For testing our method in real time, we design a multi-view face detection system based on VS 2008 and OpenCV library, and three persons with different poses, expression and other factors (glasses, hair blocking and so on) are selected for testing set. The detection results are shown as **Fig. 9.**
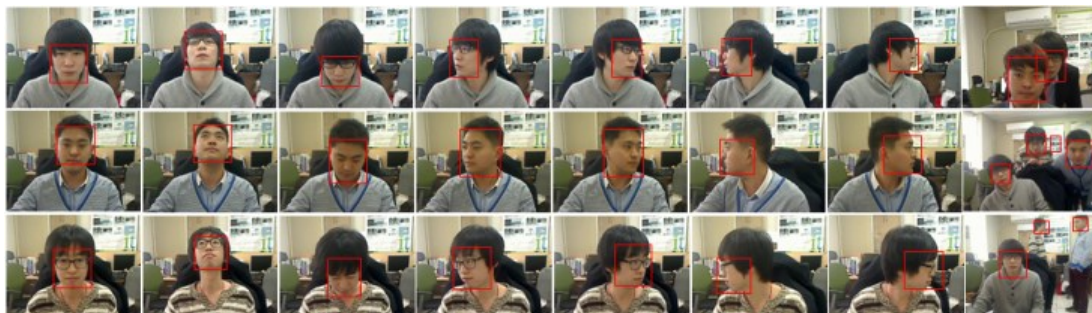

Fig. 9. Multi-view faces detection in real time

The size of each frame is set as 320×240 and the running time for detection is about 60ms with high correct detection rate. The performance of detection system is robust and effective.

6. Conclusion

In this paper, we present a novel algorithm for multi-view face detection. An improved real Adaboost algorithm for training Haar feature is proposed that used dynamic weights and pre-partitioning samples, and we prove its viability based on probability. Compared to original real Adaboost algorithm, the time complexity of weak classifier is $O(M*N)$, and the time complexity of strong classifier is $O(T*M*N)$, the training speed is enhanced by $O(N)$.

Based on our processing of Haar features for multi-view face detection (mirror image and rotation), only 12 categories classifiers are needed to build for ROP and RIP angle changing, it greatly reduces the complexity of training. After that, LUT weak classifier is used to boost Haar features in order to use real Adaboost algorithm. Moreover, instead of training a pose estimator separately, we use the confidence of strong classifier, which defined on real Adaboost algorithm in this paper. 4 of 16 layers of cascade classifiers are used to estimate face pose, there is no extra computation for pose estimating.

We tested our algorithm on CMU and MIT face database. The experimental results showed that, the convergence performance of improved real Adaboost algorithm is better than original real Adaboost algorithm. In addition, our multi-view face detection system has a high detection rate within an acceptable number of false drops and satisfactory timeliness.

References

[1]   Wenkai Xu and Eung-Joo Lee, "A Combinational Algorithm for Multi-Face Recognition," *International Journal of Advancements in Computing Technology*, vol. 4, no. 13, pp. 146-154, 2012.
http://www.aicit.org/ijact/global/paper_detail.html?jname=IJACT&q=1045

[2]   Yu-Bu Lee and Sukhan Lee, "Robust Face Detection Based on Knowledge-Directed Specification of Bottom-Up Saliency," *ETRI Journal*, vol. 33, no. 4, pp. 600-610, Aug. 2011.
http://dx.doi.org/10.4218/etrij.11.1510.0123

[3]   B. Moghaddam, A. Pentland, "Face Recognition Using View-based And Modular Eigenspaces,"

in *Proc. of Automatic Systems for the Identification and Inspection of Humans*, July, 1994.
http://dx.doi.org/10.1117/12.191877

[4] B. Moghaddam, A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 137-143, 1997.
http://dx.doi.org/10.1109/34.598227

[5] B. Moghaddam, W. Wahid, A. Pentland, "Beyond Eigenfaces: probabilistic matching for face recognition," *in Proc. of IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 30-35, 1998.
http://dx.doi.org/10.1109/AFGR.1998.670921

[6] A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 84-94, 1994.
http://dx.doi.org/10.1109/CVPR.1994.323814

[7] E. Osuna, R. Freund, F. Girosi, "Support vector machines: training and applications," *Technical report*, Massachusetts Institute of Technology, AI Memo 1602, 1997.
http://hdl.handle.net/1721.1/7290

[8] E. Osuna, R. Freund, F. Girosi, "Training support vector machines: an application to face detection," *in Proc. of Computer Vision and Pattern Recognition*, pp. 130-136, 1997.
http://dx.doi.org/10.1109/CVPR.1997.609310

[9] F. Soulie, F. Viennet and B. Lamy, "Multi-modular neural network architectures: applications in optical character and human face recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 7, no. 4, pp. 721-755, 1993.
http://dx.doi.org/10.1142/S0218001493000364

[10] K. Sung, T. Poggio, "Example-based learning for view-based human face detection," *Technical report,* Massachusetts Institute of Technology, AI MEMO 1521, 1994.
http://dx.doi.org/10.1109/34.655648

[11] H. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," *in Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 203 -207, 1997.
http://dx.doi.org/10.1109/34.655647

[12] H. Rowley, S. Baluja and T. Kanade, "Rotation invariant neural network-based face detection," *Technical report*, CMU-CS-97-201, 1997.
http://dx.doi.org/10.1109/CVPR.1998.698585

[13] H. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.
http://dx.doi.org/10.1109/34.655647

[14] Abraham Mathew and R. Radhakrishnan, "Versatile Approach for Feature Extraction of Image Using Haar-Like Filter," *KIIT Journal of Research & Education*, vol. 2, Issue 2, pp. 18-22, 2013.
http://www.kiit.in/images/stories/pdf/KIIT_Journal_of_Research_Education-12-a_March_2013.pdf

[15] P. Viola, M. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features", *in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 511-518, 2001.
http://dx.doi.org/10.1109/CVPR.2001.990517

[16] S. Li, L. Zhu, Z. Zhang and H. Zhang, "Statistical learning of multi-view face detection," *in Proc. of ECVC'02*, pp. 67-81, 2002.
http://dx.doi.org/10.1007/3-540-47979-1_5

[17] Z. Zhang, L. Zhu, S. Li, H. Zhang, "Real-time multi-view face detection," *in Proc. of IEEE International Conference on Automatic Face and Gesture Recognition'02*, pp. 149, 2002.
http://dx.doi.org/10.1109/AFGR.2002.1004147

[18] H. Schneiderman, T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars," *in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition'00*, pp. 546-751, 2000.
http://dx.doi.org/10.1109/CVPR.2000.855895

[19] Hongming Zhang, Wen Gao, Xilin Chen, Shiguang Shan and Debin Zhao, "Robust Multi-View Face Detection Using Error Correcting Output Codes," *in Proc. of ECCV'2006*, pp. 1-12, 2006.

http://dx.doi.org/10.1007/11744085_1

[20] R. E. Schapire and Y. Singer, "Improved Boosting Algorithms Using Confidence-rated Predictions," *Machine Learning*, vol. 37, Issue 3, pp. 297-336, 1999. http://dx.doi.org/10.1023/A:1007614523901

[21] Hendro Baskoro, Jun-Seong Kim and Chang-Su Kim, "Mean-Shift Object Tracking with Discrete and Real Adaboost Techniques," *ETRI Journal*, vol. 31, no. 3, pp. 282-291, June. 2011. http://dx.doi.org/10.4218/etrij.09.0108.0372

[22] Bo Wu, Chang Huang and Haizhou Ai, "A Multi-view Face Detection Based on Real Adaboost Algorithm," *Journal of Computer Research and Development*, vol. 42, no. 9, pp.812-817, 2005. http://dx.doi.org/10.1360/crad20050924

[23] Fasel Ian, Fortenberry B. and Movellan J., "A generative framework for real time object detection and classification," *Computer Vision and Image Understanding*, vol. 98, Issue 1, pp. 182-210, 2005. http://dx.doi.org/10.1016/j.cviu.2004.07.014

[24] Bo Wu, Haizhou Ai and Chang Huang, "LUT-Based Adaboost for Gender Classifiaction," in *Proc. of AVBPA' 03*, pp.104-110, 2003. http://dx.doi.org/10.1007/3-540-44887-X_13

[25] 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, pp. 42-53, 2001. http://dx.doi.org/10.1109/34.899945

[26] B. Weyrauch, J. Huang, B. Heisele, and V. Blanz. "Component-based Face Recognition with 3D Morphable Models," in *Proc. of First IEEE Workshop on Face Processing in Video*, 2004. http://dx.doi.org/10.1109/CVPR.2004.41
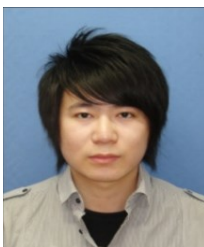
Wenkai Xu received his B. S. at Dalian Polytechnic University in China (2006-2010) and Master degree at Tongmyong University in Korea (2010-2012). Currently, he is studying in Department of Information and Communications Engineering Tongmyong University for PH. D. His main research areas are image processing, computer vision, biometrics and pattern recognition.

Eung-Joo Lee received his B. S., M. S. and Ph. D. in Electronic Engineering from Kyungpook National University, Korea, in 1990, 1992, and Aug. 1996, respectively. Since 1997 he has been with the Department of Information & Communications Engineering, Tongmyong University, Korea, where he is currently a professor. From 2000 to July 2002, he was a president of Digital Net Bank Inc. From 2005 to July 2006, he was a visiting professor in the Department of Computer and Information Engineering, Dalian Polytechnic University, China. His main research interests include biometrics, image processing, and computer vision.