Research Article

A Joint Learning Approach to Face Detection in Wavelet Compressed Domain

Szu-Hao Huang[1] and Shang-Hong Lai[2]

[1] *Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu 300, Taiwan*
[2] *Department of Computer Science, National Tsing Hua University, Hsinchu 300, Taiwan*

Correspondence should be addressed to Szu-Hao Huang; shhuang@ie.nthu.edu.tw

Face detection has been an important and active research topic in computer vision and image processing. In recent years, learning-based face detection algorithms have prevailed with successful applications. In this paper, we propose a new face detection algorithm that works directly in wavelet compressed domain. In order to simplify the processes of image decompression and feature extraction, we modify the AdaBoost learning algorithm to select a set of complimentary joint-coefficient classifiers and integrate them to achieve optimal face detection. Since the face detection on the wavelet compression domain is restricted by the limited discrimination power of the designated feature space, the proposed learning mechanism is developed to achieve the best discrimination from the restricted feature space. The major contributions in the proposed AdaBoost face detection learning algorithm contain the feature space warping, joint feature representation, ID3-like plane quantization, and weak probabilistic classifier, which dramatically increase the discrimination power of the face classifier. Experimental results on the CBCL benchmark and the MIT + CMU real image dataset show that the proposed algorithm can detect faces in the wavelet compressed domain accurately and efficiently.

1. Introduction

Automatically detecting specific objects from images has been a popular research topic for intelligent image analysis and understating with many applications, including face recognition, face tracking, expression cloning, face pose estimation, and 3D head model reconstruction from images. These applications usually assume the face regions are detected correctly as the first step. A lot of researchers from computer vision and image processing have proposed many different approaches for this problem.

Most previous face detection methods focused on detecting faces from a single gray-scale image. The survey paper [1] by Yang et al. classified the face detection methods into four categories, namely, knowledge-based methods, feature-based methods, template matching methods, and appearance-based methods. The appearance-based approach has evolved to a major stream in the face detection research. Since it is very hard to describe a general face in an image by some explicit characterization or feature description, the appearance-based

approach learns a face classifier from a large number of face and nonface examples. The training stage in this approach is to decide a two-class classifier from training examples. After collecting a large number of training face images, most researchers focus on finding a suitable feature representation and a powerful classifier for face detection.

In recent years, with the popularity of digital camera and camcorder, the demand for real-time face detection is increasing. Detecting faces directly in a compressed domain, instead of the original image, is an interesting approach that can save time in the decompression process and reduce the complexity of hardware and software design, especially that most digital images in the world are stored in a compressed form. However, not much previous research work was focused on detecting faces in a compressed domain [2, 3]. Detecting faces directly from a compressed domain can skip parts of the decompression and feature extraction process. In this paper, we propose a novel joint feature representation based on the wavelet coefficients and improve the AdaBoost-based learning for fast and accurate face detection.
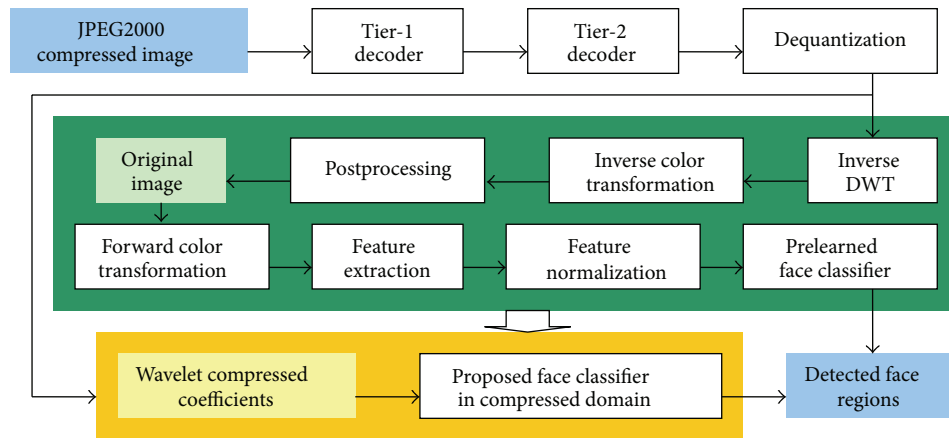
FIGURE 1: The block diagram of face detection in the original image space and the wavelet compressed domain for JPEG2000 [15, 16] compressed images. Note that the decompression and feature extraction blocks in the middle green area can be skipped in the proposed face detector.

In the detection process, the proposed system accesses the corresponding wavelet coefficients and executes the cascade classifier efficiently with a sliding-window search fashion.

Some previous face detection works proved that the wavelet representation [4–7] or Haar-like features [8–14] can well describe faces. However, if we apply the famous AdaBoost face detector [8, 11] with the features replaced by the wavelet coefficients, the resulting accuracy turns out to be unsatisfactory for the following two reasons. Firstly, the feature discrimination is limited by the restricted wavelet compresses feature space. Secondly, it is difficult to implement image contrast normalization directly in the wavelet compressed domain. Thus, this work is motivated by achieving a discriminative face detector in the restricted wavelet feature space. We proposed a paired feature representation and an improved learning framework to achieve a robust classifier with high accuracy in the wavelet compressed domain.

Figure 1 shows the flow diagram of detecting faces directly in the compressed image, such as JPEG2000 [15, 16]. Note that some decompression and feature extraction processes, that is, the blocks in green zone in Figure 1, will be skipped in the proposed method.

Increasing the discrimination in the limited wavelet compressed feature space is the main goal of our learning mechanism. We propose a space warping technique to increase the representation capability of each feature via reweighted learning of the sample distributions. In addition, the joint feature representation projects learning samples onto a paired-feature plane to improve the discrimination power. We propose an improved AdaBoost system based on learning with this joint feature representation.

In order to avoid the information loss in the feature learning procedure, some modified components are developed to preserve more information. For example, the ID3-like quantization is applied after the join feature space representation. Compared with traditional quantization methods, the proposed ID3-like quantization considers the positive and negative sample distribution in 2D pair feature space to

achieve the best discrimination by separating samples into different bins with their labels. Moreover, instead of a binary classifier, Bayesian weak learners are adopted to compute the ratio between positive and negative samples for each bin as the output for the classifier. With simple prelearned look-up tables, the weak classifier can provide more detailed classification result than a hard decision.

Finally, the trained face classifier can be applied directly in the wavelet compressed domain with very efficient calculation. The learning framework preserves the essential information in quantization tree and some look-up tables. The execution process of the face detector is simplified to some low-complexity computation, such as accessing corresponding coefficients and querying look-up tables. Although the input features of the proposed algorithm are restricted to the wavelet compressed coefficients without normalization, our experimental results show the accuracy of the proposed face detector is comparable to some state-of-the-art methods which detect faces in the uncompressed image domain.

The rest of this paper is organized as follows. Section 2 reviews the previous related works and the AdaBoost learning algorithm [8, 11], since our learning framework is based on this algorithm. Section 3 gives the details of the proposed learning framework for our face detector, including feature space warping, joint feature representation, ID3-like quantization, and weak probabilistic classifier. Subsequently, we describe the execution flow of the proposed face detector applied in the wavelet compressed domain in Section 4. In Section 5, we show several experimental results on two popular benchmarking databases. Finally, we conclude this paper in Section 6.

2. Related Works

Automatic object detection is an important issue in computer vision and image. In addition to face detection, researchers have proposed algorithms for car detection [5], pedestrian detection [6, 17–20], and even generic object detection [21]. In this section, we will first briefly review some previous face

detection techniques. Then, we will describe the AdaBoost learning algorithm [22, 23].

2.1. Previous Face Detection Techniques.

Our focus in this paper is on frontal face detection from still gray-scale images. The survey paper [1] by Yang et al. reviewed some face detection methods in the early period, such as manually established facial rules and predefined symmetric attributes. These methods are intuitive but lack robustness, because the natural sense includes too much variety and the simple heuristic rules or models can not cover all possible variations well. The appearance-based methods became the mainstream of face detection research. It normally consists of collecting a lot of training samples, projecting these samples onto an appropriate feature space, and applying machine learning techniques to form a classifier from the distribution of samples in the feature space. Most initial appearance-based methods used the pixel brightness values from the sliding windows in an image as the features. Then, some well-developed learning algorithms, such as support vector machine [24], neural network [25], eigen space analysis [26], and Sparse Network of Winnows [27], were applied to develop the face detectors.

In order to tolerate more scale and pose variations, Fleuret and Geman [28] proposed a coarse-to-fine face detector based on an edge feature descriptor. Schneiderman and Kanade [5] employed the histogram of wavelet features for face and car detection with out-of-plane rotation. In the meantime, Papageorgiou and Poggio [6] developed their multipurpose object detection system by using the wavelet features with the support vector machine classifier. In addition, Heisele et al. [29] partitioned a face region into several local patches and applied support vector machines to develop a component-based face detector. In 2001, Viola and Jones [8] presented the first real-time face detection system based on AdaBoost learning in conjunction with block sum difference features easily computed with an integral image. The efficient computation and acceptable accuracy of this system bring the face detection into real applications. More details in the AdaBoost learning will be discussed in the next subsection. Later, Liu [30] applied Bayesian Discriminating Features (BDF) technique to develop an accurate face detection system with a very low false detection rate.

Although detecting faces in gray-scale images is the most general approach, some researchers also employed the color information to simplify the face detection problem. By using the color information, the face detection system can extract more discriminative information and increase the speed and accuracy dramatically. Traditional works [4, 31] collected a large number of pixels of skin regions to determine a skin color distribution and filter. Hsu et al. [32] proposed a face detector that contains a lighting compensation step and eye/mouth color maps. Huang and Lai [7] developed a color face classifier by learning face appearance in the color feature space. Tsalakanidou et al. [33] used extra 3D range data acquired by a 3D sensor to improve the performance of color face detection under illumination and expression variations.

Detecting faces in video [7, 34, 35] has been another interesting face detection approach in recent years. Combining video tracking techniques and the face appearance models can extend the face detection from still images to video sequences. These methods proved that they can recover missed face detections and eliminate false positives by temporally integrating the face detection results.

Furthermore, multipose face detection techniques have been researched to extend the previous frontal face detection methods in recent years. Schneiderman and Kanade [5] first applied the statistical histogram representation for detecting faces in profile as well as frontal views. Then, a convolutional neural network architecture [36] and an AdaBoost method with pose estimator [10] are proposed to extend the upright and frontal face classifier to detect faces with large pose variations, with rotation up to ±30 degrees in image plane (RIP) and up to ±60 degrees for out-of-plane rotation (ROP). More recently, Huang et al. [14] proposed the width-first-search structure and vector boosting algorithm to accomplish the face detection with arbitrary RIP angles and the ROP angles up to ±90 degrees.

In addition, some more related works were developed for different problem settings and different applications. For example, detecting small faces from degraded images [13] focused on detecting low-resolutional faces. There were previous methods proposed to detect faces in the DCT compressed domain [2, 3], which is somewhat related to the problem setting of this paper. The major difference between the previous works and the proposed method is that our method can be applied directly in the wavelet compressed domain without wavelet decomposition or intensity normalization, and it can still achieve high accuracy comparable to the state-of-the-art face detection methods.

In addition to the face detection, face identification and recognition is another challenging problem which had been widely discussed in computer vision research field. After the face region is detected precisely from the input image, the face recognition system would analyze the frontal facial image patch and determine or verify the identity of the person. Zhao et al. [37] had extensively reviewed early machine recognition systems and surveyed several psychological studies which focused on human faces. These works can be roughly categorized into two types: face recognition from single still image and face recognition from video sequences. Wright et al. [38] proposed a new classification framework based on sparse representation techniques and provided new insights into two crucial issues: feature extraction and robustness to occlusion. To solve the face identification under uncontrolled environments or with a lack of training samples, Schwartz et al. [39] employed a large and rich set of feature descriptors and used partial least squares regression model to increase the discriminant ability of recognizer across varying conditions.

2.2. AdaBoost Learning.

AdaBoost, short for adaptive boosting, is a meta-algorithm that can cooperate with other machine learning techniques to improve their performance. The AdaBoost algorithm was originally proposed by Freund and Schapire [22], and this original algorithm is listed in Algorithm 1.

As shown in Algorithm 1, after a series of labeled data put into the learning machine, we need to initialize the weights of each learning sample. In most two-class classification

Input: sequence of N labeled examples $\langle (x_1, y_1), \ldots, (x_n, y_n) \rangle$
       Distribution D over N examples
       Weak learning algorithm **WeakLearn**
       integer T specifying number of iterations
Initialize the weight vector $w_i^1 = D(i)$ for $i = 1, \ldots, N$
Do for $t = 1, 2, \ldots, T$
    (1) Set $p^t = w^t / \sum_{i=1}^{N} w_i^t$
    (2) Call **WeakLearn**, providing it with the distribution p^t; get back a hypothesis $h_t: X \rightarrow [0, 1]$
    (3) Calculate the error of h_t: $\varepsilon_t = \sum_{i=1}^{N} p_i^t |h_t(x_i) - y_i|$
    (4) Set $\beta_t = \varepsilon_t / (1 - \varepsilon_t)$
    (5) Set the new weight s vector to be $w_t^{t+1} = w_i^t \beta_t^{1 - |h_t(x_i) - y_i|}$
Output the hypothesis

$$h_f(x) = \begin{cases} 1 & \text{when } \sum_{t=1}^{T} \left(\log \frac{1}{\beta_t} \right) h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \log \frac{1}{\beta_t} \\ 0 & \text{otherwise} \end{cases}$$

ALGORITHM 1: The AdaBoost algorithm [22].

problems without prior knowledge of the training examples, the weight summations of all positive and negative data will be set equally and each learning sample belonging to the same category has the same weight. Another issue is about an adequate quantity of training samples, which is very difficult to determine for a practical machine learning problem. Bootstrap learning architecture provides a solution to resolve this problem.

In the AdaBoost algorithm depicted in Algorithm 1, WeakLearn is a function or an algorithm that performs the hypothesis to classify the input samples into different categories by considering the current sample weights. The word "weak" means the hypothesis is not expected to be very powerful since it only uses very simple features and calculation in the weak classifier. In most applications, the WeakLearn function is normally designed in a simple way, such as a binary function of a feature value. The basic idea is that the WeakLearn classification functions are very easy to calculate and at least slightly better than random guess. Thus, the AdaBoost learning algorithm is applied to select a set of discriminating and complementary weak classifiers to form a final strong classifier.

The input integer T specifies the number of iterations in the learning system. One obvious advantage of AdaBoost is that it did not need any tuning parameters except T. The selection of T depends on different applications. Selecting a larger T value will decrease the error measure in the training data, but it may lead to the overfitting problem. The value h_t is decided in each iteration and the sample weights are updated from the error measure. The basic idea in the AdaBoost is to assign more weighting to the samples misclassified in the previous iterations to achieve a global optimization process.

AdaBoost has been very popular in computer vision and image processing research fields since the first real-time face detection method proposed by Viola and Jones [8]. For face detection, there were some improved versions, such as Kullback-Leibler Boosting [40], FloatBoost [9], and asymmetric AdaBoost [41] algorithm, to increase the accuracy and efficiency of learning performance. Some modifications of

the learning framework extended Viola and Jones' method to different applications, such as detecting faces in video [7], detecting faces in degraded images [13], and detecting pedestrians via motion and appearance patterns [17]. A similar technique of AdaBoost was also applied in different feature space to solve other two-class learning problems. The image retrieval with relevance feedback [42, 43] was also an important application of AdaBoost.

At the end of this section, we list some improved versions of the AdaBoost algorithm in Table 1. Although most applications applied the AdaBoost algorithm to solve the two-class classification problems, there were some extensions of the AdaBoost algorithm to the regression [22] and multiclass classification problem [23].

3. Proposed Learning Method

The proposed learning system is an improved version of Viola and Jones' face detector [8, 11] to adapt to our requirement, that is, detecting faces directly from wavelet compressed domain. The fundamental structure is similar to most appearance-based learning methods. An initial training data set, including 4916 face image blocks and 7872 nonface (negative) image blocks, is prepared for the learning of the face classifier. When the trained AdaBoost classifier can separate samples in the training dataset well, the current classifier is applied to a large image database to accumulate false positive blocks as the negative samples in the next training dataset. In the bootstrap learning system, the growing negative learning samples are extracted from 100 different categories of Corel PhotoStock database, totally 10000 natural images.

For the proposed face detector, a 3-leveled wavelet transformation is applied for each training image to obtain the 576-dimensional features. The LL band of highest level from 24×24 images is skipped because this part cannot be recovered when we execute the face detection directly from compressed domain without any decompressed process.

TABLE 1: Variants of the AdaBoost algorithm.

| Algorithm	Authors and references	Description
AdaBoost	Freund and Schapire [22]	Original algorithm
RealBoost	Schapire and Singer [23]	A real version of weak hypothesis
AdaBoost.M1/M2	Freund and Schapire [22]	Multiple class extension of original AdaBoost algorithm
AdaBoost.R	Freund and Schapire [22]	Solving regression problems
AdaBoost.MO/MH/MR	Schapire and Singer [23]	Multiclass, multilabel extensions
Kullback-Leibler Boosting	Liu and Shum [40]	Incorporate Kullback-Leibler divergence into AdaBoost
FloatBoost	Li and Zhang [9]	Replace exponential error function by backtrack mechanism

(a)

(b)
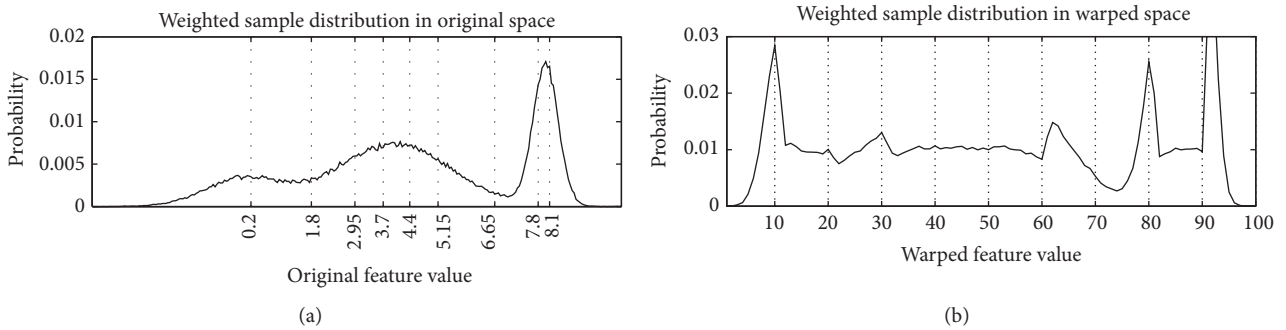
FIGURE 2: The feature space warping representation: (a) the landmarks in the original distribution and (b) the weighted distribution in the warped feature space.

3.1. Learning System Overview.

The goal of the improved AdaBoost learning algorithm is to learn an efficient face detector from a restricted wavelet feature space. Without wavelet decomposition and intensity-based normalization processes, the feature discrimination power for face detection is weak. The improved AdaBoost learning system contains four major improvements based on two principles: higher discrimination and more information preservation. Algorithm 2 gives an overview of the proposed system. More details of each step will be described in the following parts of this section.

3.2. Feature Space Warping.

The function of feature space warping is similar to the histogram equalization for image enhancement. The basic idea is that we should use more levels in the area with dense data distribution and less levels in the area with sparse data for the feature quantization. In actual implementation, the distribution of the training data samples should be reweighted by the current weights. We need a nonlinear transformation for each feature to increase the feature representation capability.

A discrete cumulative density function is estimated to find some landmark points, such as the feature value located in 50% weighted distribution. After we have these landmarks, a simple space warping, which linearly interpolates samples between these points, is applied. Figure 2(a) shows the landmarks of original weighted data distribution, and Figure 2(b) shows the weighted distribution after space warping. After the space warping process, the distance measure of a single feature between two different samples is driven by the current sample weights and distributions.

3.3. Joint Feature Representation.

Schneiderman and Kanade [5] have first adopted the idea of joint distribution of a pair of features to represent objects. Mita et al. [12] simply extended the AdaBoost detector proposed by Viola and Jones and combined three binary weak classifiers to the three-digit code learning. In the proposed method, we map all learning samples onto paired wavelet feature spaces and estimate the corresponding 2D distributions before selecting the weak classifiers. Because of the feature space warping step, the data distribution appears approximately uniform when only considering a single feature. After the joint feature representation, mapping data to higher dimension is a strategy to increase the feature variety instantly. With the processes of permutation and combination, it provides higher possibility to explore more discrimination.

The original feature dimension for a three-level wavelet transform of a 24-by-24 block image is 576. After the cross-level and cross-band combinations of a pair of wavelet coefficients, the feature dimension is increased to 160,461 dramatically without extracting any additional information. Instead of the paired feature representation, a cubic or even higher dimensional mapping is another aggressive possibility. However, the efficiency of the AdaBoost learning system should be considered, especially when the computational cost for the AdaBoost learning from the paired feature spaces is already very high.

The major design principle for the feature space warping and the joint feature representation is to explore high discrimination from the limited set of wavelet coefficients. After the joint feature representation is created, the following steps are designed for preserving more information in the

(i) Given example images $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_n, y_n)$, where y_i takes the value 0 for negative examples or 1 for positive examples, respectively.

(ii) Initialize weights $w_{1,i} = 1/(2m)$ for $y_i = 0$ or $w_{1,i} = 1/(2l)$ for $y_i = 1$, where m and l denote the total numbers of negative and positive images, respectively.

(iii) For $t = 1, \ldots, T$

(1) Estimate feature space warping function $W_j(x_i)$ via the sample distribution $[x_{1,j}, x_{2,j}, \ldots, x_{n,j}]$ and weights $\mathbf{w}_t$

(2) For each possible feature pair (j, k), map all training sample onto paired plane via warped feature value $W_j(x_i)$ and $W_k(x_i)$

(3) Apply ID3-like tree method to each axis of paired feature plane in rotation, and find the quantization function $f_{j,k}(x_i)$ which will try to separate positive and negative samples into different bins.

(4) Compute the conditional probability as the Bayesian classification result for each weak classifier $C_{j,k}(x_i)$.

(5) Estimate the error $\varepsilon_{j,k}$ for each feature pair (j, k) as follows:

$$\varepsilon_{j,k} = \sum_{i=1}^{n} w_{t,i} \times \left| y_i - C_{j,k}(x_i) \right|$$

(6) Select the paired feature $h(t)$ with minimum error $\varepsilon_{j,k}$

$$h(t) = \arg \min_{(j,k)} \varepsilon_{j,k}$$

(7) Update the weights for all training samples as follows:

$$w_{t+1,i} = w_{t,i} \times \beta_t^{1 - |y_i - c_{h(t)}(x_i)|},$$

where $\beta_t = \varepsilon_{h(t)} / \left(1 - \varepsilon_{h(t)}\right)$.

(8) Normalize the weights by

$$w_{t+1,i} = \frac{w_{t+1,i}}{\sum_{i=1}^{n} w_{t+1,i}}$$

(iv) The final classifier is given by

$$SC(x) = \begin{cases} \text{face} & \text{when } \sum_{t=1}^{T} \alpha_t C_{h(t)}(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t, \\ \text{non-face} & \text{otherwise} \end{cases}$$

where $\alpha_t = \log\left(1/\beta_t\right)$.

ALGORITHM 2: The proposed AdaBoost-based learning algorithm for face detection.

feature quantization before going into the AdaBoost learning procedure.

3.4. ID3-Like Plane Quantization. After the joint feature representation, we can estimate the positive and negative data distributions for each feature pair from the training data. To develop weak classifiers for all possible pair features for AdaBoost learning, we quantize the paired feature plane for computing the conditional probabilities of the joint features for positive and negative cases. Our strategy is to segment the paired feature plane into several representative regions and use the ratio of the conditional probabilities in each quantized region for classification. In other words, we want to quantize the continuous paired feature plane such that each quantized segment has its own dominant sample label. With this strategy, the system can separate the positive and negative into different segments as good as possible to achieve high discrimination capability.

To achieve the above goal, we employ the ID3-like plane quantization on the paired feature space. This quantization for each feature is determined based on the distribution of the training data with the current weight function. Compared to other traditional quantization methods, our ID3-like approach can retain more information with a little bit more

computational effort. The main algorithm of ID-3 decision tree [44] is to select the best boundary in each node such that it can divide the data passing through this node into two classes with the highest information gain. It means that the boundary is selected such that each branch contains as much data of the same class as possible. In other words, we want to find appropriate boundaries to divide the data into intervals of maximal uniformity.

In the ID3-decision tree, we first define the entropy and information gain as follows:

$$\begin{aligned} \text{Entropy}(S) = &-p\left(y_i = 1 \mid x_i \in S, W_t\right) \\ &\times \log p\left(y_i = 1 \mid x_i \in S, W_t\right) \\ &- p\left(y_i = 0 \mid x_i \in S, W_t\right) \\ &\times \log p\left(y_i = 0 x_i \in S, W_t\right), \end{aligned}$$

$$\text{Gain}(S, A) \equiv \text{Entropy}(S) - \sum_{v \in \text{leaf\_nodes}} \frac{|S_v|}{|S|} \text{Entropy}(S_v),$$

(1)

where W_t is the weighting function for each training sample at the tth iteration and the symbol A denotes a function that can classify samples from a parent node S to a different leaf node S_v. Because our ID3-like balance tree quantization is
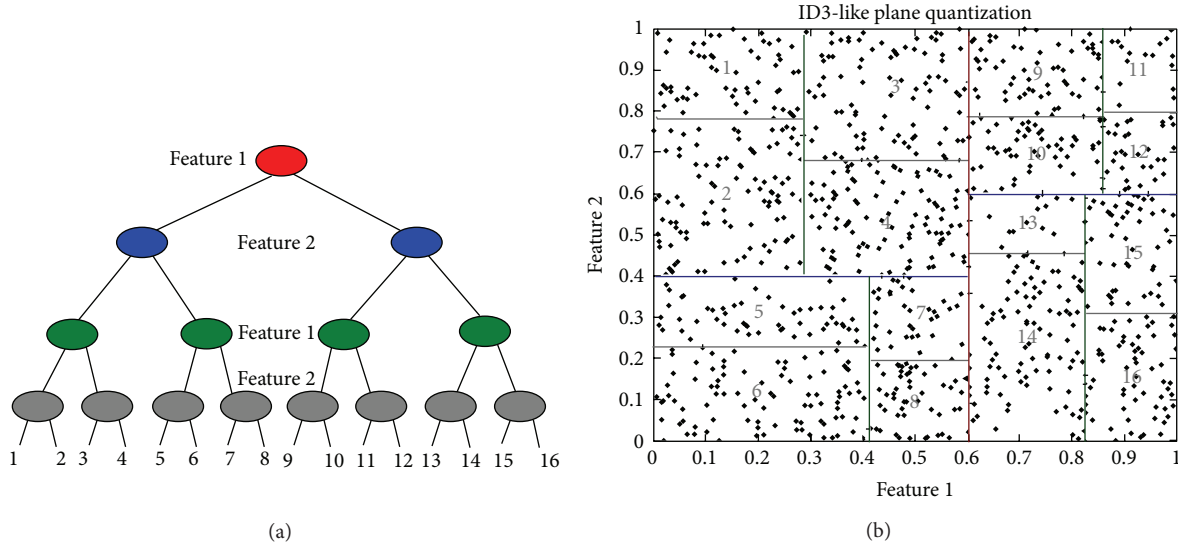
FIGURE 3: ID3-like plane quantization representation: (a) a 4-layer ID3 tree structure and (b) the corresponding quantization regions and boundaries in the feature plane.

a binary tree, the function A can be represented by a list of thresholds that divide a feature value into left or right leaf nodes recursively. Then, we can select the best seed value that maximizes the information gain as follows:

$$\text{selected}_{\text{boundary}} = \arg\underset{A}{\text{Max}}\left(\text{Gain}\left(S, A\right)\right). \tag{2}$$

First, entering all learning samples into root node can find the best boundary in one of the axes to separate the space into two parts. In these two regions, we turned the seed selection process in another axis independently. The ID3-like plane quantization involves repeating the above process recursively and alternatively along the two axes to determine a quantization function $f_{j,k}(x)$. If the processing time is a critical issue, histogram equalization can provide the initial seeds to speed up the computation with similar performance. In practical setting, a four-layer decision tree is constructed just like Figure 3(b) to quantize the pair feature plane into 16 different regions, as shown in Figure 3(a).

3.5. Weak Probabilistic Classifier.
For each pair of features, we can train a weak classifier based on the corresponding joint conditional probability determined from the training data. The AdaBoost training algorithm is then used to select some powerful and complementary weak classifiers and combine them to form a final classifier for face detection. For each weak classifier, we apply the Bayes rule based on the conditional probability density function in each interval to decide the class. By thresholding the ratio of face to nonface conditional probabilities for a given paired feature vector in a quantized interval, we obtain a number of weak classifiers that can be used in the AdaBoost training algorithm.

Making a binary decision in the weak classifier does not fully exploit the information computed in the conditional probabilities. Therefore, we replace the binary decision in the weak classifier by the conditional probability in the AdaBoost learning algorithm.

By applying the Bayes rule, we can compute the conditional probability as follows:

$$C_{j,k}\left(x_i\right)$$

$$= \frac{p\left(y_i = 1 \mid f_{j,k}\left(x_i\right), W_t\right)}{p\left(y_i = 0 \mid f_{j,k}\left(x_i\right), W_t\right) + p\left(y_i = 1 \mid f_{j,k}\left(x_i\right), W_t\right)}$$

$$= \left(p\left(f_{j,k}\left(x_i\right) \mid y_i = 1, W_t\right)\right)$$

$$\times \left(p\left(f_{j,k}\left(x_i\right) \mid y_i = 1, W_t\right) + p\left(f_{j,k}\left(x_i\right) \mid y_i = 0, W_t\right)\right.$$

$$\left. \times \frac{p\left(y_i = 0\right)}{p\left(y_i = 1\right)}\right)^{-1}, \tag{3}$$

$$p\left(f_{j,k}\left(x_i\right) \mid y_i = 1, W_t\right)$$

$$= \frac{\sum_{\text{all } m, \text{let } f_{j,k}(x_m)=f_{j,k}(x_i)} y_m W_{t,m}}{\sum_{m=1}^{n} y_m W_{t,m}}, \tag{4}$$

where $C_{j,k}(x)$ means the probability of the (j, k)th pair feature for image x to be class 1, that is, face class and $f_{j,k}(x)$, which denotes the leaf node index of ID3 tree determined from the plane quantization. Equation (4) measures the conditional distribution of $f_{j,k}(x)$ under the situation that label $y_i = 1$ and the sample weights w_t are updated after t iterations of the AdaBoost learning algorithm. Equation (3) returns a probability value between 0 and 1, which indicates the face conditional probability. We use $C_{j,k}(x)$ to be the Bayesian weak classifier in the AdaBoost system.

In our implementation, we segment the paired feature plane into 16 regions and a look-up table containing 16

(i) Given a test image represented in N-layered wavelet compressed domain
(ii) Each layered-coefficient plane L_i is composed of three sub-bands, L_i^{LH}, L_i^{HL} and L_i^{HH}
(iii) Preprocessing
    (1) Apply the bi-linear interpolation to down-sample each sub-band to 1/1.25, 1/1.5,
       and 1/1.75 scales, respectively, and form three additional wavelet layer sets.
(iv) For each of these four sets of the wavelet-layer representation, run the sliding window face
    detection with the scale i initialized to 1
    (1) Apply the AdaBoost face classifier to each sliding window which is constructed from
       the coefficients in the planes from L_i to L_{i+2}.
    (2) If the classifier determines the region is a face, calculate and save the position and
       size of the corresponding window in the original image space based on the shift,
       downsample, and layer information.
    (3) Repeat the previous two steps with the scale i incremented by one until the scale $i > N - 2$.
(v) Postprocessing
    (1) Eliminate the overlapped face regions based on the scores provided by the AdaBoost classifier.
    (2) Output the detected faces.

ALGORITHM 3: The proposed face detection algorithm in the wavelet compressed domain.

Bayesian probabilities is used for the AdaBoost learning and testing processes.

4. Detecting Faces on Wavelet Compressed Domain

In this section, we describe how the trained AdaBoost face classifier is applied for face detection from the wavelet representation of the whole image. The previous AdaBoost face detector is featured for its simple computation. Although we add several complicated components into our learning system to increase the overall discrimination for face detection from a restricted feature space, the resulting AdaBoost classifier only needs a very small amount of computation with some look-up tables and quantization tree structures. Algorithm 3 depicts the complete algorithm of the proposed face detection on the wavelet compressed domain. Sections 4.1 and 4.2 give more details of our implementation.

4.1. Face Detection in a Single Sliding Window. First, we describe how the trained AdaBoost classifier is applied in a sliding window search strategy for face detection from a wavelet compressed image. When focusing on a single block of a whole image to determine whether it is a face region, we can access the corresponding wavelet coefficients from HL, HH, and LH bands of three contiguous levels. Since the LL bands for all levels are not available without a complete decompression process, we do not include them into our feature space learning.

In each iteration t, the selected hypothesis $h_t(x)$ from our system refers to a weak classifier associated with a feature pair (j, k). The first step is to retrieve the coefficient values of features j and k. These two feature values are quantized through the corresponding ID3-like plane quantization, as described above, with only four comparison operators. Then, a simple look-up table is used to access the corresponding weak classifier output value for the quantized feature region, which can be done very efficiently. Finally, the accumulated

sum of the products of α_t, provided from the AdaBoost training, and the weak Bayesian classifier probability output is used to determine if this window is a face region or not.

4.2. Face Detection in a Whole Image. When detecting faces in a whole image, the position and scale of sliding windows should cover all the possible image blocks where faces could appear. The minimal detection window is as large as the training samples with 24×24 pixels. The shift of the sliding window in the wavelet compressed domain can be easily implemented by adding relative shift terms when seeking the corresponding feature values. The shift at one higher level is twice the shift at the previous level. Figure 4 is a simple chart that describes the corresponding spatial relationship of the window shifting. The gray solid rectangle shifts one coefficient in level 3 from the red rectangle. The relative shifts in level 2 and level 1 are 2 and 4, respectively. Recovering to the original image domain, the gray detected window shifts 8 pixels in horizontal direction from red one. This kind of scanned windows with full correspondence is not dense enough for face detection. Therefore, shifting windows of the corresponding coefficients in lower level and rounding the shifts in higher level could achieve higher accuracy.

In addition to the shift, the various scales of the sliding windows in the compressed domain are not easy to implement without wavelet decomposition. The multilevel structure of wavelet decomposition provides a basis to find the corresponding features from different scales, but it is restricted to the detected windows with power-of-2 scales of the template face window. For example, if we can detect faces of size 48-by-48 from wavelet levels $N - 2$, $N - 1$, and N, then the coefficients related to 96-by-96 faces are positioned in levels $N-3$ to $N-1$. In order to detect faces of sizes between these two scales, we apply the bilinear interpolation to the wavelet coefficient plane. Downsampling these coefficient planes to 1/1.25, 1/1.5, and 1/1.75 of the original width and height will create three different starting scale bases. Thus, between sizes 24 and 48, we can have sizes 30, 36, and
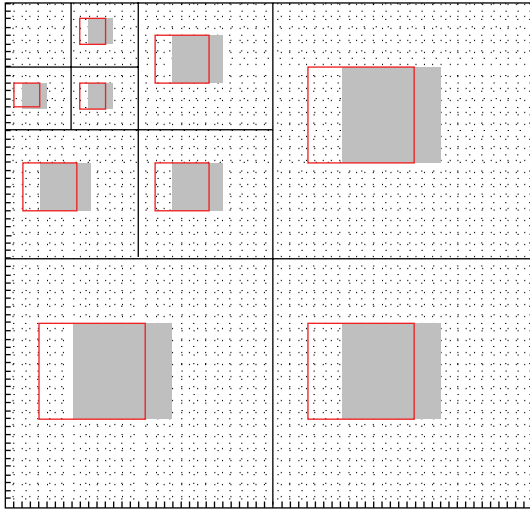
FIGURE 4: The corresponding coefficients in compressed domain of different detected window.

42 for different window widths with the same framework. The downsampled plane with 1/1.25 ratio can provide the coefficients of 30-by-30 detection windows and its higher wavelet levels should cover the sizes 60,120,..., and so forth, in the original image scale. An additional postprocessing is required to decrease the detected face regions which are overlapped with each other. The positive windows with higher scores in strong classifier will be reserved as the final decision.

5. Experimental Results

In this section, we show four sets of experiments to verify the improvement of the proposed AdaBoost learning system to demonstrate the performance and efficiency of the proposed face detector directly on the wavelet compressed domain. We first adopted CBCL face database which contains separate training and testing image datasets for evaluating the AdaBoost learning results. Totally 24,045 testing clipped images are examined after one-time learning from 6,977 training image blocks. Then, the MIT + CMU face database is used for the testing of face detection from whole images with bootstrap learning. Our experimental results show that the proposed learning system significantly improves the accuracy of face detection on the restricted wavelet compresses domain.

5.1. Learning System Improvement Benchmarks. To justify the improved AdaBoost-based learning system, we need a benchmark to evaluate the performance of the proposed method. MIT CBCL face dataset provides a fair benchmarking database to compare the performance of face classifiers. The training dataset contains 6,977 image blocks (2,429 face blocks and 4,548 nonface blocks) and the testing data set is composed of 24,045 image blocks (472 face blocks and 23,573 nonface blocks).

All the experiments in this part have three different learning system settings: AdaBoost learning with Viola and Jones' feature space [11], AdaBoost learning in conjunction with the 567-dimensional wavelet feature space, and the proposed learning system on the paired wavelet feature space.

The first benchmarking experiment is performed for each of the three learning systems with 10 weak classifiers only. The results are depicted in Figure 5(a). It is obvious that changing the feature space from the original feature space in Viola and Jones' method to the restricted wavelet feature space degrades the face detection accuracy dramatically. However, with our improved AdaBoost learning, the detection rates of the proposed face detector are better than all the other two systems under the same false positive rates. One may argue that the comparison is unfair because the proposed paired feature learning strategy adopts two features for each iteration, that is, weak classifier. Therefore, we performed another experiment, shown in Figure 5(b), which restricted each of the three final classifiers can only include 20 wavelet coefficients. The result is more reasonable from our expectation. When the false alarm is equal to 0.2, the detection rate of Viola and Jones' method is 0.79. The same learning method with wavelet features only has 0.65 detection rate, while the proposed system with paired wavelet features can improve the face detection rate to 0.76. In addition, the proposed method can obtain better results in this benchmark comparison when the false alarm rate increases to 0.35.

In order to examine the limit of the learning systems, we increase the accessing feature number to 200. In other words, the standard AdaBoost classifiers contain 200 iterations and the proposed paired feature learning system consists of 100 weak classifiers. As shown in Figure 5(c), it has similar result to that of Figure 5(b), and the proposed system outperforms Viola and Jones' method when the false alarm rate is larger than 0.185.

5.2. ROC Curve of MIT + CMU Dataset. The second experiment is mainly used to examine the performance of the proposed learning system after bootstrap learning. Our system is designed for face detection in wavelet compressed domain, so the input to our face detection system is the wavelet representation of the whole testing image.

MIT + CMU face dataset is widely used in the face detection research. There are 130 gray-scale images containing 507 faces in this dataset. Figure 6 depicts three ROC curves and detection rates with respect to different numbers of false positives, obtained by applying the three different face detection algorithms to the entire dataset. The experimental results show that the proposed method can improve the detection rate from 0.68 to 0.89 under 100 false positives, which is near the 0.92 detection rate in raw image.

The curve of Viola and Jones' method was published in their paper [11] and adopted here for comparison. Another ROC curve is obtained by applying the same AdaBoost learning algorithm with the wavelet features. It is obvious from Figure 6 that the proposed learning system with the paired wavelet features improves the ROC curve with the restricted wavelet features significantly and it is close to the ROC curve of Viola and Jones' face detector, which is based on a large number of features.
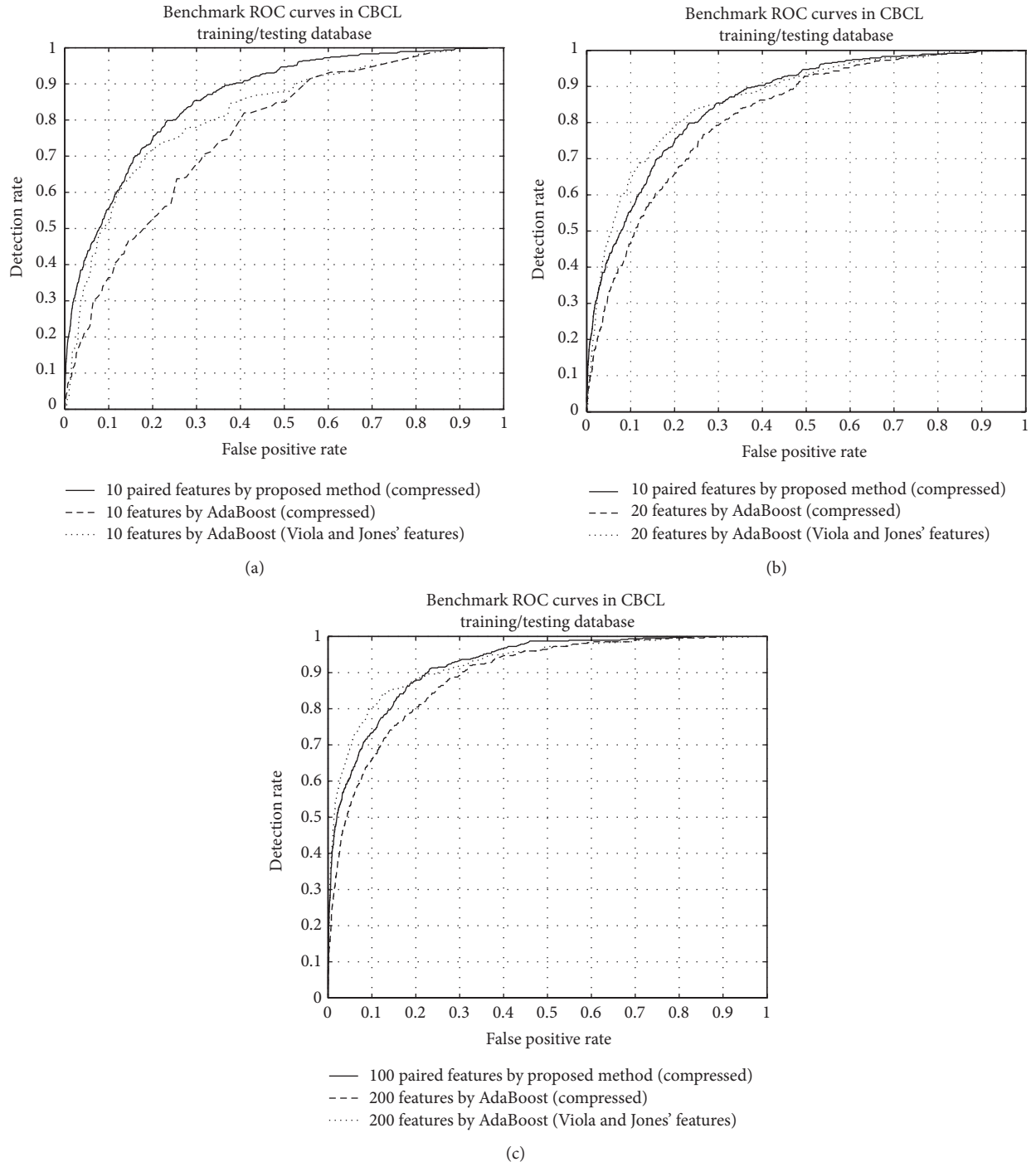
(a)

(b)

(c)

FIGURE 5: The benchmark comparison on the CBCL database: (a) three different learning systems with 10 weak classifiers adopted, (b) three different learning systems with 20 features adopted, and (c) three different learning systems with 200 features adopted.

5.3. More Comparisons on MIT + CMU Dataset. In this experiment, we provide more results and comparisons between the proposed face detection system and other systems. The experimental setting is similar to that in the previous two experiments. The detection rates under different numbers of false detection of the proposed face detector and some previous methods on the MIT + CMU face dataset are depicted in Table 2.

In this comparison, we can see that the accuracy of the proposed method is about 1%~5% lower than that of Viola-Jones' method under different rates of false positives. We think it is a reasonable accuracy decrease since the proposed face detection algorithm is restricted to the limited wavelet feature space and it can save the decompression cost. Figure 7 depicts the face detection results of the proposed algorithm in the test images of MIT + CMU dataset.

TABLE 2: The detection rates under different false detections of several different face detection systems.

Detection method	False detections							
	10	31	50	65	78	95	167	422
AdaBoost (compressed)	56.1%	62.5%	63.9%	65.9%	66.5%	67.9%	70.9%	75.9%
Proposed method (compressed)	79.8%	85.2%	87.3%	88.1%	88.5%	89.2%	90.9%	93.3%
Viola and Jones [8]	76.1%	88.4%	91.4%	92.0%	92.1%	92.9%	93.9%	94.1%
Viola and Jones (voting) [11]	81.1%	89.7%	92.1%	93.1%	93.1%	93.2%	93.7%	—
Rowley et al. [25]	83.2%	86.0%	—	—	—	89.2%	90.1%	89.9%

TABLE 3: The execution time (ms) of the proposed face detector and the original AdaBoost face detector for detecting faces from wavelet compressed images.

| Image size (pixels)	Sliding windows	Tier 1 + Tier 2	Method	IDWT	Integral images	Feature extraction	Classifiers	Processing time	Total time
3320 × 240	8,411	27.6	Original	18.4	1.8	19.5	9.7	49.4	77.0
			Proposed	—	5.4	—	14.1	19.5	47.1
640 × 480	41,243	95.7	Original	75.2	7.2	96.6	47.5	226.5	322.2
			Proposed	—	21.4	—	69.3	90.7	186.4
1024 × 768	113,789	207.4	Original	183.7	18.3	263.8	131.1	596.9	804.3
			Proposed	—	55.3	—	190.7	246.0	453.4
1600 × 1200	291,580	488.1	Original	432.9	43.7	677.1	334.2	1,487.9	1,976.0
			Proposed	—	135.7	—	492.8	628.5	1,116.6
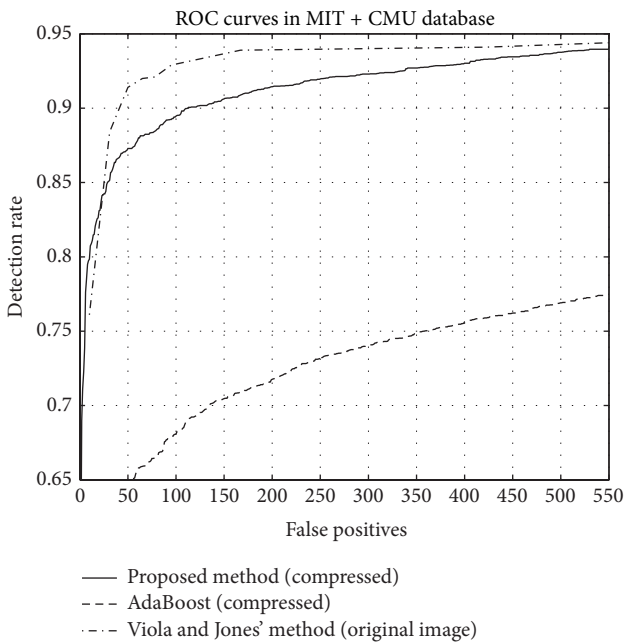
FIGURE 6: The ROC curve of the final classifier in MIT + CMU database.

FIGURE 7: Face detection results of proposed methods in MIT + CMU dataset.

FIGURE 8: False positives and missing of proposed face detection methods in MIT + CMU dataset.

The experimental results show that our system can detect face regions correctly not only in the natural photos but also in the paintings and sketches. Moreover, Figure 8 displays some false positives and missing detection of proposed method in the same database. We observe that the proposed system misses several faces with partial occlusions or higher rotation angles as shown in Figure 8. In addition, balancing the number of missing and false positive rate is critical for a binary decision system design. The false positives shown in Figure 8 can be eliminated with a tighter detector and the detection rate will also be decreased to 87.3%.

5.4. Execution Time Analysis. Table 3 depicts the execution time of the components in the proposed face detection system that operates directly in the wavelet domain and the original AdaBoost face detector [8, 11] that detects faces after wavelet decomposition. This experiment is performed on a

regular PC with Intel E6320 CPU (Dual 1.87 GHz Cores) and 2 G RAM. The face detector was applied from 36×36 blocks to full image size with 1.25 times scale increase. In the spatial domain, the sliding window is scanned on the image for face detection with step 1/8 window width in both horizontal and vertical directions for speed consideration.

The proposed wavelet-domain face detector skips the IDWT procedure and the feature extraction process to achieve more efficient detection. Our experiments show that the processing time, which skips the tier 1 and tier 2 decoding time, on a 320×240 image is only 19.5 ms. When discussing the total execution time of detecting faces from a compressed image, our face detector only requires 57% computation time of the original AdaBoost face detector [8, 11].

6. Conclusion

In this paper, we proposed a face detection system working directly in the wavelet compressed domain. The main contributions of the proposed face detection algorithm are an improved AdaBoost-based learning framework and a series of joint feature representation strategies, which can produce a strong face classifier on the restricted wavelet feature space.

The proposed face detection system involves a feature space warping process, a paired feature learning scheme, an ID3-like joint feature plane quantization method, and a weak Bayesian classifier. Although some complicated components which increase the power of classifiers are included in the face detection system, the execution of the final face classifier is quite simple. With tree structure quantization and look-up tables, the proposed face detector can work very efficiently and directly on the wavelet compressed domain. Our experimental results on the benchmarking face datasets showed that the proposed face detection system working in the compressed domain can achieve similar accuracy to that of Viola and Jones' face detector [8, 11].

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

References

[1] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: a survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[2] S. Lee and M. H. Hayes, "A simple and fast color-based human face detection scheme for content-based indexing and retrieval," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 1, pp. 701–704, October 2004.

[3] P. Fonseca and J. Nesvadba, "Face detection in the compressed domain," in *Proceedings of the International Conference on Image Processing (ICIP '04)*, vol. 3, pp. 2015–2018, October 2004.

[4] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264–277, 1999.

[5] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 746–751, June 2000.

[6] C. Papageorgiou and T. Poggio, "A trainable system for object detection," *International Journal of Computer Vision*, vol. 38, no. 1, pp. 15–33, 2000.

[7] S.-H. Huang and S.-H. Lai, "Real-time face detection in color video," in *Proceedings of the 10th International Multimedia Modelling Conference (MMM '04)*, pp. 338–345, January 2004.

[8] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I511–I518, December 2001.

[9] S. Z. Li and Z. Zhang, "FloatBoost learning and statistical face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1112–1123, 2004.

[10] R. Xiao, M.-J. Li, and H.-J. Zhang, "Robust multipose face detection in images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 31–41, 2004.

[11] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004.

[12] T. Mita, T. Kaneko, and O. Hori, "Joint Haar-like features for face detection," in *Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV '05)*, vol. 2, pp. 1619–1626, October 2005.

[13] S. Hayashi and O. Hasegawa, "A detection technique for degraded face images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 1506–1512, June 2006.

[14] C. Huang, H. Ai, Y. Li, and S. Lao, "High-performance rotation invariant multiview face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 4, pp. 671–686, 2007.

[15] ISO/IEC 15444-1, "JPEG2000 Image Coding System," 2000.

[16] D. S. Taubman and M. W. Marcellin, *JPEG2000: Image Compression Fundamentals, Practice and Standards*, Kluwer Academic Publishers, Boston, Mass, USA, 2002.

[17] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," in *Proceedings of the 9th IEEE International Conference on Computer Vision*, vol. 2, pp. 734–741, October 2003.

[18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, vol. 1, pp. 886–893, June 2005.

[19] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan, "Fast human-detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 1491–1498, 2006.

[20] Q. Zhu, S. Avidan, M.-C. Yeh, and K.-T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, pp. 1491–1498, June 2006.

[21] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 3, pp. 416–431, 2006.

[22] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of Computer and System Sciences*, vol. 55, no. 1, pp. 119–139, 1997.

[23] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, no. 3, pp. 297–336, 1999.

[24] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 130–136, June 1997.

[25] H. A. 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.

[26] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[27] D. Roth, M. H. Yang, and N. Ahuja, "A SNoW-based face detector," in *Proceedings of the Conference on Advances in Neural Information Processing Systems*, vol. 12, pp. 855–861, 2000.

[28] F. Fleuret and D. Geman, "Coarse-to-fine face detection," *International Journal of Computer Vision*, vol. 41, no. 1-2, pp. 85–107, 2001.

[29] B. Heisele, T. Serre, and T. Poggio, "A component-based framework for face detection and identification," *International Journal of Computer Vision*, vol. 74, no. 2, pp. 167–181, 2007.

[30] C. Liu, "A Bayesian discriminating features method for face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 6, pp. 725–740, 2003.

[31] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 557–563, 1999.

[32] R.-L. Hsu, M. Abdel-Mottaleb, and A. K. Jain, "Face detection in color images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 696–706, 2002.

[33] F. Tsalakanidou, S. Malassiotis, and M. G. Strintzis, "Face localization and authentication using color and depth images," *IEEE Transactions on Image Processing*, vol. 14, no. 2, pp. 152–168, 2005.

[34] R. Choudhury Verma, C. Schmid, and K. Mikolajczyk, "Face detection and tracking in a video by propagating detection probabilities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1215–1228, 2003.

[35] E. Loutas, I. Pitas, and C. Nikou, "Probabilistic multiple face detection and tracking using entropy measures," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 128–135, 2004.

[36] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1408–1423, 2004.

[37] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[38] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[39] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face identification using large feature sets," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2245–2255, 2012.

[40] C. Liu and H.-Y. Shum, "Kullback-Leibler boosting," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. I/587–I/594, June 2003.

[41] J. Wu, S. C. Brubaker, M. D. Mullin, and J. M. Rehg, "Fast asymmetric learning for cascade face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 369–382, 2008.

[42] K. Tieu and P. Viola, "Boosting image retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '00)*, pp. 228–235, June 2000.

[43] S.-H. Huang, Q.-J. Wu, and S.-H. Lai, "Improved AdaBoost-based image retrieval with relevance feedback via paired feature learning," *Multimedia Systems*, vol. 12, no. 1, pp. 14–26, 2006.

[44] T. M. Mitchell, *Machine Learning*, MIT Press, McGraw-Hill, 1997.