Unconstrained Face Detection

Shengcai Liao, Anil K. Jain, *Fellow, IEEE* and Stan Z. Li, *Fellow, IEEE*

Abstract—Face detection, as the first step in automatic facial analysis, has been well studied over the past two decades. However, challenges still remain for face detection in unconstrained scenarios, such as arbitrary pose variations and occlusions. In this paper, we propose a method to address these challenges in unconstrained face detection. First, a new type of image feature, called Normalized Pixel Difference (NPD) is proposed. NPD feature is computed as the difference to sum ratio between any two pixel intensity values, inspired by the Weber Fraction in experimental psychology. Besides its computational efficiency, the NPD feature has several desirable properties, such as scale invariance, boundedness, and ability to reconstruct the original image. Second, we develop a method for learning the optimal subset of NPD features and their combinations via regression trees, so that complex face manifolds can be partitioned by the learned rules. This way, only a single cascade classifier is needed to handle unconstrained face detection. The proposed face detector is robust in handling pose, occlusion, illumination, blur and low image resolution. Experimental results on three public face datasets (FDDB, GENKI, and CMU-MIT) show that the proposed method outperforms the state-of-the-art methods reported to date in detecting unconstrained faces with arbitrary pose variations and occlusions in cluttered scenes.

Index Terms—Unconstrained face detection, normalized pixel difference, regression tree, AdaBoost, cascade classifier, pose, occlusion, blur

1 Introduction

The objective of face detection is to find and locate faces in an image. It is the first step in automatic face recognition applications. Face detection has been well studied for frontal and near frontal faces. The Viola and Jones' face detector [1] is the most well known face detection algorithm, which is based on Haar-like features and cascade AdaBoost [2] classifier. However, in unconstrained scenes such as faces in a crowd, state-of-the-art face detectors fail to perform well due to large pose variations, illumination variations, occlusions, expression variations, out-of-focus blur, and low image resolution. For example, the Viola-Jones face detector fails to detect most of face images in the FDDB database [3] (examples shown in Fig. 1) due to the difficulties mentioned above. In this paper, face detection with arbitrary facial variations is called the unconstrained face detection problem. We are interested in face detection in unconstrained scenarios such as video surveillance or images captured by hand-held devices.

Fig. 1. Face images annotated (red ellipses) in the FDDB database [3].

Numerous face detection methods have been developed following Viola and Jones' work [1], mainly focusing on extracting different types of features and developing different cascade structures. Various complex features [4], [5], [6], [7], [8], [9], [10], [11], [12], [13] have been proposed to replace the Haar-like features used in [1]. While these methods can improve the face detection performance to some extent, they generate very large number (hundreds of thousands) of features and the resulting systems take too much time to train. Another development in face detection has been to learn different cascade structures for multiview face detection, such as parallel cascade [14], pyramid architecture [15], and Width-First-Search (WFS) tree [16]. All these methods need to learn one cascade classifier for each specific facial view (or view range). In unconstrained scenarios, however, it is not easy to define all possible views of a face, and the computational cost increases with increasing number

- *Shengcai Liao was with the Dept. of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA. He is now with the National Laboratory of Pattern Recognition and the Center for Biometrics and Security Research, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: scliao@nlpr.ia.ac.cn*
- *Anil K. Jain is with the Dept. of Computer Science and Engineering, Michigan State University, East Lansing, MI 48824 USA. He is also affiliated with the Dept. of Brain & Cognitive Engineering, Korea University, Anamdong, Seongbukgu, Seoul 136-713, Republic of Korea. E-mail: jain@cse.msu.edu*
- *Stan Z. Li is with the National Laboratory of Pattern Recognition and the Center for Biometrics and Security Research, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: szli@cbsr.ia.ac.cn*

of classifiers in complex cascade structures. Moreover, these approaches require manual labeling of face pose in each training image.

While previous methods [14], [15], [16] handled multiview faces alone, they did not simultaneously consider other challenges such as occlusion. In fact, since these methods require to partition multiview data into known poses, occlusion is not easy to handle in this way. On the other hand, while several studies addressed face detection under occlusion [17], [18], [19], [20], [21], they constrained themselves to detect frontal faces under occlusion. As discussed in [22], a robust face detection algorithm should be effective under arbitrary variations in pose and occlusion, which is still an unresolved challenging problem.

In this paper, we are interested in developing effective features and robust classifiers for unconstrained face detection with arbitrary facial variations. First, we propose a simple pixel-level feature, called the Normalized Pixel Difference (NPD). An NPD is computed as the ratio of the difference between any two pixel intensity values to the sum of their values, in the same form as the Weber Fraction in experimental psychology [23]. The NPD feature has several desirable properties, such as scale invariance, boundedness, and ability to reconstruct the original image. Besides, it is easy to compute, involving only one addition, one subtraction, and one division between two pixel values per feature computation.

Secondly, we develop a method to construct a single cascade classifier that can effectively deal with complex face manifolds and handle arbitrary pose and occlusions. While the individual NPD feature may have "weak" discriminative ability, our work indicates that a subset of NPD features can be optimally selected by AdaBoost learning and combined to construct more discriminative features in a regression tree. This is a "divide and conquer" strategy to tackle unconstrained face detection in a single classifier, without pre-labeling of views in the training set of face images. The resulting face detector is robust to variations in pose, occlusion, and illumination, as well as to blur and low image resolution. The robustness and performance of the proposed face detector come from the advantages of the NPD features and the way the classifier is constructed.

The novelty of this work is summarized as follows:

- A new type of feature, called NPD is proposed, which is efficient to compute and has several desirable properties, such as scale invariance, boundedness, and enabling reconstruction of the original image.
- A subset of NPD features is automatically learned and combined in regression trees to boost their discriminability. In this way only a single cascade AdaBoost classifier is needed to handle unconstrained faces with occlusions and arbitrary viewpoints, without pose labeling or clustering in the training stage.

The advantages of the proposed approach include:

- A total of $(20 \times 20) \times (20 \times 20 - 1)/2 = 79,800$ NPD features are computed in a 20×20 face template for feature representation. While the Viola-Jones face detector [1] requires image normalization and integral image computation, our feature representation does not need such computation.
- No complex cascade structure is required for pose invariant face detection; pose labeling or clustering in the training stage is also not required.
- A single face detector is able to handle illumination variations, pose variations, occlusions, out-of-focus blur, and low resolution face images in unconstrained scenarios.

The remainder of this paper is organized as follows. In Section 2 we review the related work. In Section 3 we introduce the NPD feature space. The proposed NPD based face detection method is presented in Section 4. Experimental results are provided in Section 5. Finally, we summarize the contributions in Section 6.

2 Related work

As indicated in a recent survey of face detection methods [24], most popular face detection methods are appearance based[1], which use local feature representation and classifier learning like boosting or SVM. Viola and Jones's work [1] was the first one to apply rectangular Haar-like features in a cascaded AdaBoost classifier for real-time face detection. Since then, many approaches have been proposed to advance the state of the art in face detection. Several researchers have tried to extend the original Haar-like features proposed in [1]. Lienhart and Maydt [4] proposed an extended set of Haar-like features, where 45° rotated rectangular features were introduced. Li et al. [5] proposed another extension of Haar-like features, where the rectangles can be spatially set apart with a flexible distance. A similar feature, called the diagonal filter was also proposed by Jones and Viola [6]. Various other local texture features have been introduced for face detection, such as the modified census transform [7], local binary pattern (LBP) [8], MB-LBP [11], LBP histogram [10], and the locally assembled binary feature [12]. These features have been shown to be more robust to illumination variations. Mita et al. [9] proposed the joint Haar-like features to capture the co-occurrence of effective Haar-like features. Huang et al. [16] proposed a sparse feature set in a granular space, where granules were represented by rectangles, and each individual sparse feature was learned as a combination of granules. A problem with the approaches in [9] and [16] is that the joint feature space is very large, making the

1. Skin color based face detection [22], [25], [26] is another promising method. Please refer to [22], [26] for surveys along this direction.

optimal combination a difficult task. Recently, Jain and Learned-Miller [27] proposed an online domain adaption approach to improve the Viola-Jones face detector's performance on the FDDB database [3]. Li et al. [13] proposed the use of SURF feature [28] in an AdaBoost cascade for face detection, which achieved impressive results on the FDDB database [3].

While complex features may provide better discrimination power than Haar-like features for the face detection task, they generally increase the computational cost. In contrast, ordinal relationships among image regions are simple yet effective image features [29], [30], [31], [32], [33], [34]. Sinha [29] studied several robust ordinal relationships in face images and developed a face detection method accordingly. Liao et al. [32] further showed that ordinal features can be effectively learned by AdaBoost classifier for face recognition. Sadr et al. [30] showed that pixelwise ordinal features (ordinal relationship between any two pixels) can faithfully encode image structures. Similar ideas for exploiting ordinal relationships between pixels were also proposed in [31], [33], [34]. Baluja et al. [31] showed that simple pixelwise ordinal features are good enough for discriminating between five facial orientations, a relatively simpler task than face detection. Wang et al. [34] applied the random forest classifier together with pixelwise ordinal features for facial landmark localization. Abramson and Steux [33] proposed a pixel control point based feature for fast face detection, where each feature is associated with two sets of pixel locations (control points). However, it is not easy to learn these optimal control point based features because of the huge number of control points combinations.

Besides different feature representations, some researchers have also tried different AdaBoost algorithms and weak classifiers. For weak classifiers utilized in boosting, Lienhart et al. [35] and Brubaker et al. [36] have shown that classification and regression trees (CART) [37] work better than simple decision stumps. In this paper, we show that optimal ordinal features and their combinations can be learned by integrating the proposed NPD features in a regression tree to represent the intrinsic object structure, and in this way, arbitrary pose variations can be automatically partitioned into different leaves of the learned regression tree.

Given that the original Viola-Jones face detector has limitations in multiview face detection [24], various cascade structures have been proposed to tackle multiview face detection [6], [14], [15], [16]. Jones and Viola [6] extended their face detector by training one face detector for each specific pose. To avoid evaluating all face detectors on each scanning subwindow, they developed a pose estimation step (similar to Rowley et al. [38]) before face detection, and then only the face detector trained on that estimated pose was applied. In this two-stage detection structure, if the pose estimation is not reliable, the face is not likely to be detected in the second stage. Wu et al. [14] proposed a parallel cascade structure for multiview face detection, where all face detectors of different views have to be evaluated for each scanning window; they did use the first several cascade layers of all face detectors to estimate the pose for speedup. Li and Zhang [15] proposed a coarse-to-fine pyramid architecture for multiview face detection, where the whole range of face poses was divided into increasingly smaller subranges, resulting in a more efficient detection structure. Huang et al. proposed a WFS tree based multiview face detection approach, which also works in a coarse-to-fine manner. They proposed the Vector Boost algorithm for multiclass learning, which is well suited for multiview pose estimation. However, all these methods have to learn one cascade classifier for each specific view (or view range) of face, requiring an input face image to go through different branches of the detection structure. The computational cost generally increases with increasing number of classifiers in complex cascade structures. Moreover, these approaches require manual labeling of the face pose in each training image.

Instead of designing detection structures, Lin and Liu [19] proposed that multiview face detector can be learned in a single cascade classifier. They derived a multiclass boosting algorithm, called MBHBoost by sharing features among different classes. This is a more straightforward approach for multiview face detection than designing complex cascade structures. Nevertheless, it still requires manual labeling of poses. In uncontrolled environments, however, it is not easy to define specific views of a face by discretizing the pose space, because a face could be in arbitrary pose simultaneously in yaw (out-of-plane), roll (in-plane), and pitch (up-and-down) angles. To avoid manual labeling, Seemann et al. [39] suggested learning viewpoint clusters automatically for object detection. However, for human faces, Kim and Cipolla [40] showed that clustering by traditional techniques like K-Means does not result in categorized poses. They hence proposed a multiple classifier boosting (MCBoost) for human perceptual clustering of object images, which was shown to be promising for clustering face poses. However, it is not necessary that their clustering output be only related to pose variations. For example, in addition to different pose clusters, they also obtained clusters with various illumination variations.

Face detection in presence of occlusion is also an important issue in unconstrained face detection, but it has received less attention compared to multiview face detection. This is probably because, compared to pose, it is more difficult to categorize occlusions into several predefined classes; faces in unconstrained scenarios can be occluded in arbitrary ways. Hotta [17] proposed a local kernel based SVM method for face detection, which was shown to be more reliable than global kernel based SVM in detecting frontal faces occluded by sunglasses or scarf. Lin et al. [18] tried to handle 8 kinds of manually defined facial occlusions by training 8 additional cascade classifiers besides the standard face detector, which led to improved performance in detecting occluded faces. Lin and Liu [19] further proposed the MBHBoost algorithm to handle faces with one of 12 in-plane rotations or one of 8 types of occlusions, where each kind of rotation or occlusion is treated as a different class. But, in their study pose and occlusion were not simultaneously present in the face. Chen et al. [20] proposed a modified Viola-Jones face detector, where the trained detector was divided into sub-classifiers related to several predefined local patches, and a classifier fusion was considered for the outputs of sub-classifiers. Goldmann et al. [21] proposed a component-based approach for face detection, where four components (the two eyes, nose, and mouth) were detected separately, and further connected in a topology graph to achieve face detection. However, none of the above methods considered face detection with both occlusions and pose variations simultaneously in unconstrained scenarios. As discussed in [22], a robust face detection algorithm should be effective under arbitrary variations in pose and occlusion, which is still an unresolved challenging problem.

3 Normalized Pixel Difference Feature Space

The Normalized Pixel Difference (NPD) feature between two pixels is defined as

$$f(x,y) = \frac{x-y}{x+y}, \tag{1}$$

where $x, y \geq 0$ are intensity values of the two pixels[2] in an image patch, and $f(0,0)$ is defined as 0 when $x = y = 0$.

The NPD feature measures the relative difference between two pixel values. The sign of $f(x,y)$ tells the ordinal relationship between the two pixels x and y (see Lemma 2 below), and the magnitude of $f(x,y)$ measures the relative difference (as a percentage of the joint intensity $x+y$) between x and y. Note that the definition of $f(0,0) \triangleq 0$ is reasonable between two pixels because in this case there is no difference between the two pixels x and y. Compared to the absolute difference $|x-y|$, NPD is invariant to scale change of the pixel intensities (see Lemma 3 below).

Weber, a pioneer in experimental psychology, stated that the just-noticeable difference in the magnitude change of a stimulus is proportional to the magnitude of the stimulus, rather than its absolute value [23]. This is known as the Weber's Law. In other words, the human perception of difference in stimulus is often measured as a fraction of the original stimulus, that is, in a form $\Delta I/I$, which is called the Weber Fraction. Chen et al. [41] proposed a local image descriptor, called Weber's Law Descriptor for face recognition, which was computed from Weber Fractions of pixels in a 3×3 local image region. The proposed measurement of the difference to sum ratio in Eq. (1) has also been used in other fields such as remote sensing, where the Normalized Difference Vegetation Index (NDVI) [42] is defined as the difference to sum ratio between the visible red and the near infrared spectra to estimate the green vegetation coverage.

The NPD feature has a number of desirable properties, as stated in the following Lemmas.

Lemma 1 (Antisymmetry): The NPD feature is antisymmetric with respect to exchange of variables, that is,

$$f(x,y) = -f(y,x), \ \forall x, y. \tag{2}$$

Given this property, either $f(x,y)$ or $f(y,x)$ is adequate for feature representation, resulting in a reduced feature space. Therefore, in an $s \times s$ image patch (vectorized as $p \times 1$, where $p = s \cdot s$), NPD feature $f(x_i, x_j)$ of every pixel pair $1 \leq i < j \leq p$ is computed to form the feature set, resulting in $d = p(p-1)/2$ features. For example, in a 20×20 face template, there are $(20\times20)\times(20\times20-1)/2 = 79,800$ NPD features in total. We call the resulting feature space NPD feature space, denoted as Ω_{npd} ($\in \mathbb{R}^d$).

Lemma 2 (Ordinal Relationship): $\forall x, y$, $f(x,y) > 0$ if and only if $x > y$; $f(x,y) < 0$ if and only if $x < y$; and $f(x,y) = 0$ if and only if $x = y$.

This Lemma states that the sign of $f(x,y)$ is an indicator of the ordinal relationship between x and y. Ordinal relationship has been shown to be an effective encoding for object detection and recognition [29], [30], [32] because ordinal relationship encodes the intrinsic structure of an object image and it is invariant under various illumination changes [29]. However, simply using a threshold of zero to encode the ordinal relationship is likely to be sensitive to noise when x and y have similar values. In the next section we will show how to learn robust ordinal relationships with NPD features.

Lemma 3 (Scale Invariance): The NPD feature is scale invariant, that is, given any constant factor $a \neq 0$,

$$f(ax, ay) = f(x,y), \ \forall x, y. \tag{3}$$

With this property, the NPD feature is expected to be robust against illumination changes. This is important for image representation, since illumination change is always a troublesome issue for both object detection and recognition.

Lemma 4 (Boundedness): $\forall x, y \geq 0$, the NPD feature f(x,y) is well bounded in [-1,1]. In addition, $f(x,y) = 1$ if and only if $x > 0$ and $y = 0$; and $f(x,y) = -1$ if and only if $x = 0$ and $y > 0$.

2. For ease of representation, sometimes we also say x and y are two pixels. We use subscripts to differentiate between pixel and pixel values only when pixel locations are under discussion.

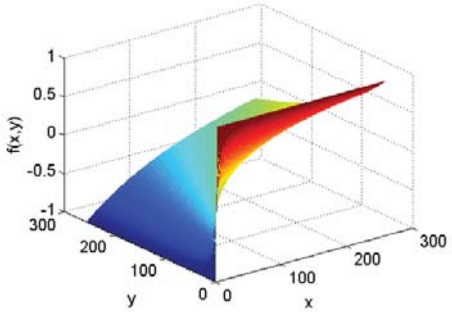


Fig. 2. A plot of the NPD function $f(x, y)$.

Fig. 3. Learning and combining ordinal features in a regression tree. Left: four pixelwise ordinal features are automatically selected in the learning process. Right: the four features are optimally combined in a regression tree for face/nonface prediction.

Appendix A contains a proof of Lemma 4. The bounded property makes the NPD feature amenable to some popular operations like histogram binning or threshold learning in tree-based classifiers [1]. Fig. 2 shows that $f(x, y)$ is a bounded function and it defines a nonlinear surface.

***Theorem** 1 **(Reconstruction)**:* Given the NPD feature vector $\mathbf{f} = (f(x_1, x_2), f(x_1, x_3), \ldots, f(x_{p-1}, x_p))^T \in \Omega_{npd}$, the original image intensity values $I = (x_1, x_2, \ldots, x_p)^T$ can be reconstructed up to a scale factor.

The proof of Theorem 1 is shown in Appendix B. The proof also gives a linear-time approach to reconstruct the original image subject to a scale factor. Theorem 1 states that there exists a preimage of a point in Ω_{npd}, and the preimage is a one-dimensional subspace of the original pixel intensity space. In contrast, Lemma 3 says that every one-dimensional subspace (it is in fact half of the subspace given the nonnegative constraint) in the original pixel intensity space is "compressed" to a point in the bounded feature space Ω_{npd}. Therefore, Ω_{npd} is a feature space which is invariant to scale variations, but it carries all the necessary information from the original space.

In practice, the NPD feature set is still large, and a feature selection procedure needs to be applied to obtain an optimal subset of NPD features for face detection, as described in the next section.

4 NPD FOR FACE DETECTION

4.1 Learning Object Structures

Ordinal relationship [29] is a very simple and basic concept: it compares the brightness of any two image regions, and encodes the result with 1 (brighter) or 0 (darker) accordingly. Sinha [29] showed that ordinal features can represent the intrinsic structure of objects such as a human face, and they are insensitive to illumination changes. Instead of encoding ordinal relationship between two image regions, in this paper, we learn robust ordinal relationships from pairwise pixel values via the NPD feature. For a face pattern which is well structured, automatically learned combinations of ordinal features may represent it better than manual configurations. Therefore, we propose to learn the combination of simple ordinal features by boosted regression trees [37]. By providing a training set of face and nonface images, a weak classifier is learned by a regression tree. Regression tree is a simple classifier that is well suited for ordinal features. At each node, the tree checks the optimal ordinal feature value, and then sends the input data to the next branch accordingly. Fig. 3 demonstrates the idea of learning and combining binary ordinal features automatically in regression trees. Regression tree is also well suited for face detection with free pose variations, since similar views can be clustered in the same leaf node of the regression tree.

Ordinal relationship can always be generated by the default threshold 0, which is likely to be sensitive to noise when the two pixels to be compared have similar values. In this paper, we propose to learn robust ordinal relationships and their combinations by learning regression trees with NPD features. In this way, regression trees not only learn optimal NPD features at each branch node, but also learn optimal thresholds for splitting. Generally, either of the two cases below are leaned for each NPD feature at a branch node to represent the object structure. These two cases are

$$f(x, y) = \frac{x - y}{x + y} < \theta_1 < 0, \tag{4}$$

$$f(x, y) = \frac{x - y}{x + y} \geq \theta_2 > 0, \tag{5}$$

where θ_1 and θ_2 are the two thresholds. Eq. (4) applies if the object pixel x is notably darker than pixel y, while Eq. (5) covers the case when pixel x is notably brighter than pixel y. This way, the optimally learned thresholds make the ordinal encodings in the learned regression trees better represent the intrinsic object structure. To learn such regression trees, we apply the CART algorithm [37] with the NPD features.

4.2 Face Detector

Given that the proposed NPD features contain redundant information, we also apply the AdaBoost algorithm to select the most discriminative features and construct strong classifiers [1]. We adopt the Gentle AdaBoost algorithm [2] to learn the NPD feature based regression trees.

As in [1], a cascade classifier is further learned for rapid face detection. We only learn one single cascade classifier for unconstrained face detection robust to occlusions and pose variations. This implementation has an advantage that there is no need to label the pose of each face image manually or cluster the poses before training the detector. In the learning process, the algorithm automatically divides the whole face manifold into several sub-manifolds by regression trees, so that face images with different poses are separated.

Below is a summary of how the proposed method is effective in handling various aspects of the unconstrained face detection problem.

- **Pose**. Pose variations can be handled by learning NPD features in boosted regression trees, where different views can be automatically partitioned into different leaves of the regression trees.
- **Occlusion**. As indicated in [18], Haar-like features or similar rectangle based features are sensitive to occlusions, because the encodings affected by occluded regions are no longer reliable. In contrast, each of the proposed NPD feature is computed by only two pixel values, hence, the number of affected NPD features by occlusion is much less than that of rectangular features which generally cover large areas.
- **Illumination**. The proposed NPD features are scale invariant, and they are converted to ordinal features with optimally learned threshold in regression trees. Therefore, NPD features are robust to illumination changes.
- **Blur or low image resolution**. Because the NPD feature involves only two pixel values, it does not require rich texture information on the face. Therefore, the proposed method is also effective in handling blurred or low resolution face images.

5 Experiments

In the following, we evaluate the performance of the proposed NPD face detector on three public face databases, FDDB [3], GENKI [43], and CMU-MIT [38]. We also provide an analysis to explain how the proposed method works, and report unconstrained face detection performances under illumination variations, pose variations, occlusions, and blur, respectively.

5.1 Implementation of NPD Face Detector

A subset of the training data[3] in [13] was used to train our detector, including 12,102 face images and 12,315 nonface images (some private face images and the Corel5k nonface images were not available so could not be used). Fig. 4 shows some example face and nonface images from this training dataset. For a composition of this training dataset, please refer to [13]. The detection template was 20×20 pixels. The detector cascade contained 15 stages, and for each stage, the training goal of false accept rate was 0.5, with a detection rate of 0.999. For the depth of the regression trees, we set a constraint that each leaf node must contain at least (1/16)th of the total number of training samples. Under this constraint, the tree depth is at most 5, and in the test phase at most 4 NPD features need to be computed for each regression tree. The final detector contains 176 weak classifiers (regression trees) constructed by 2,035 NPD features. In contrast, the Viola-Jones detector [1] contains 6,061 Haar-like features in total. The first five stages of our detector include 3, 4, 6, 7, 9 weak classifiers, respectively. On average, each 20×20 scanned subwindow needs to evaluate 34 NPD features. Notice that our method does not require any image preprocessing. Fig. 5 shows the learned NPD features contained in the three regression trees in the first stage. It can be observed that most of the learned features are around eyes, eyebrows, and nose. Besides, the learned features in the three regression trees are distributed in different parts of the facial region. This is because in the boosting scheme all samples are reweighted when a weak classifier is learned, so that the next weak classifier can focus on the training samples that can not be correctly classified in the current step. The face shown in Fig. 5 is a frontal face, but it should be kept in mind that the face can have pose variations, and some learned features may be only effective for a specific pose.

In the test stage, a scale factor of 1.2 was set for the pyramid detection. A postprocessing method similar to the OpenCV face detection module was implemented, which merges nearby detections by the disjoint set algorithm. For each detected face, we summarized scores of AdaBoost classifiers in all stages of the cascade to be the final score, and used this score to generate the Receiver Operating Characteristic (ROC) curve as a measure of performance.

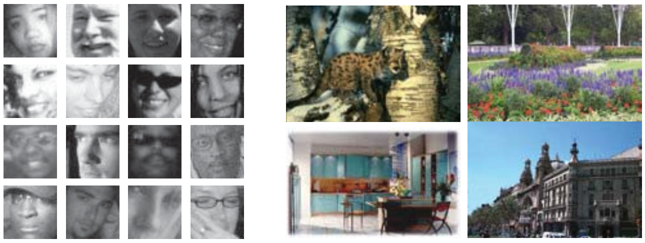


Fig. 4. Example face (left) and nonface (right) images from [13] for face detector training.

3. https://sites.google.com/site/leeplus/publications/facedetectionusingsurfcascade

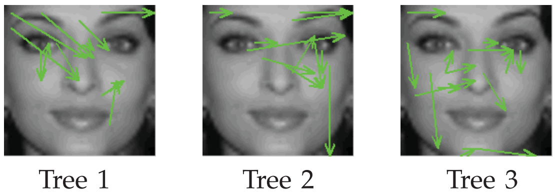


Fig. 5. The learned NPD features by boosting regression trees in the first stage.

Our face detector is implemented in MATLAB, with a C-Mex function for NPD feature extraction and classifier evaluation. For processing VGA image frames (640×480) containing 1 to 8 faces per frame, the current implementation runs at 21.6 frames per second (FPS), on average, on a PC with Intel Quad 2.66GHz CPU. We have also tested four different implementations of the Viola-Jones face detector provided in OpenCV2.4.1. These four detectors run at 21.8 FPS, 19.9 FPS, 17.0 FPS, and 16.1 FPS, respectively, for the same VGA frames on the same computer. All the five detectors tested were configured with a scaling factor of 1.2 and a minimal face of 48×48 to detect, and with parallel computing enabled. Therefore, our NPD face detector implemented in MATLAB can achieve a comparable speed to the fastest Viola-Jones face detector in OpenCV that has been well optimized for real-time performance.

We used three public face databases, FDDB [3], GENKI [43], and CMU-MIT [38], to evaluate our face detection algorithm. The details of these databases are presented below.

5.2 Evaluation on FDDB Database

Face Detection Data set and Benchmark (FDDB) is a face detection database for challenging scenarios, developed by Jain and Learned-Miller [3] at the University of Massachusetts, Amherst. The FDDB database contains 2,845 images with 5,171 faces, with a wide range of challenging scenarios including arbitrary pose, occlusions, different lightings, expressions, low resolutions, and out-of-focus faces. All faces are annotated with elliptical regions. Fig. 1 shows some examples of the annotated faces from the FDDB database.

For benchmark evaluation, Jain and Learned-Miller [3] also provided an evaluation code for a comparison of different face detection algorithms. There are two metrics for performance evaluation based on ROC: discrete score metric and continuous score metric, which correspond to coarse match (similar to previous evaluations in the face detection literature) and precise match, respectively, between the detection and the ground truth. Two experimental setups are proposed in [3]. The first experiment requires a 10-fold cross-validation, while the second experiment allows unrestricted training, which means that training data outside of FDDB can be used for face detector

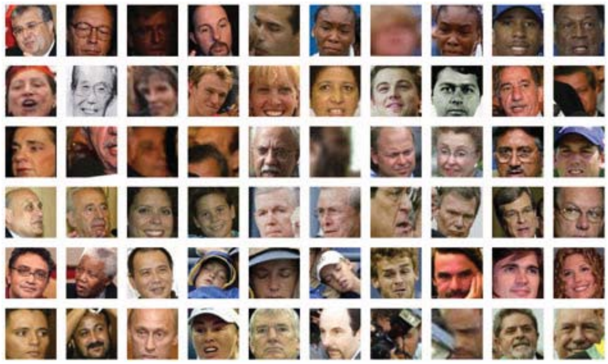


Fig. 6. Face images cropped from the FDDB database [3].

Fig. 7. Modified images from the FDDB database [3] for bootstrapping nonface samples.

training. We followed both experimental protocols. For Experiment 1, we trained 10 face detectors, with the same training settings described in Section 5.1, and tested the 10 subsets separately in 10-fold cross-validation manner. On average, we used about 4,500 face images annotated in FDDB to train a single face detector. Fig. 6 shows some face images that were cropped from the FDDB database for training our face detectors. It can be observed that the face appearance has large variations. Since FDDB does not provide a set of nonface images, we replaced all annotated face regions with black patches in the FDDB images and then used the resulting images to bootstrap nonface samples. Fig. 7 illustrates such modified images.

For Experiment 2, we used the detector trained with data outside FDDB, as described in the previous subsection. For evaluation, this detector was applied on each subset of the FDDB database separately, and an average performance was obtained.

We compared our method with state-of-the-art results reported on the FDDB website[4]. The ROC curves of various algorithms are depicted in Fig. 8 for the discrete score metric and in Fig. 9 for the continuous score metric. In both figures, "NPD-FDDB" is the proposed detector for Experiment 1, while "NPD-Outside" is the proposed detector for Experiment 2. "Viola-Jones" is the Viola-Jones face detector [1] implemented in OpenCV (reported in [3]). "SURF Cascade" is the SURF descriptor based cascade method proposed by Li et. al. in [13], which is the best published method to date. "Olaworks, Inc." is a commercial face detector. Note that the proposed NPD-FDDB is the only detector that follows the Experiment

4. http://vis-www.cs.umass.edu/fddb/results.html

1 protocol, while all others report results only for the Experiment 2 protocol. Note also that "NPD-Outside" used a sub training set that was previously used for "SURF Cascade".

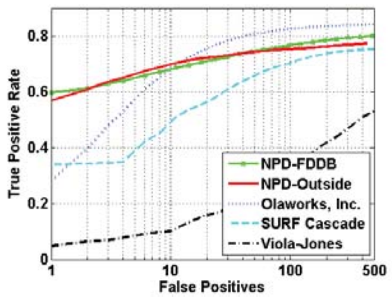


Fig. 8. ROC curves for face detection on the FDDB database [3] with the discrete score metric.

Fig. 8 shows that the NPD detector outperforms both the SURF cascade and the Viola-Jones detector with the discrete score metric. The NPD detector is also much better than the Olaworks' detector when the number of false positives (FP) is lower than 10. For example, when FP=0 (cannot be shown in log scale), our detector can detect about 60% of the annotated FDDB faces in coarse sense (50% overlap with ground truth), while the detection rates of all other detectors are below 40%. It is also observed that the performance of NPD-FDDB is comparable to NPD-Outside.

As for the continuous score metric (see Fig. 9), the NPD detector outperforms both the SURF cascade method and the Viola-Jones detector. NPD-FDDB outperforms the commercial detector Olaworks, Inc., especially when FP$<$ 20. In this test, FP=285 generally means one false detection per image, on average. In addition, NPD-Outside does not perform as good as NPD-FDDB, though the training data size for NPD-Outside is several times larger than that for NPD-FDDB. This result indicates that FDDB contains sufficient training data for unconstrained face detection. However, it is not easy to handle all this data in training a single detector (recall the large appearance variations as seen in Fig. 6). The proposed method makes a success in that generic NPD feature is learned in regression trees to divide and conquer the complex face manifolds.

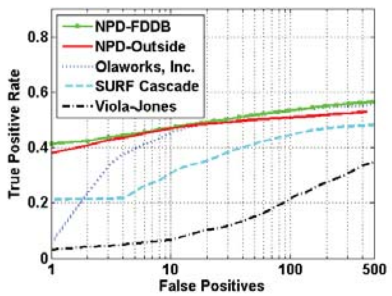


Fig. 9. ROC curves for face detection on the FDDB database [3] with the continuous score metric.

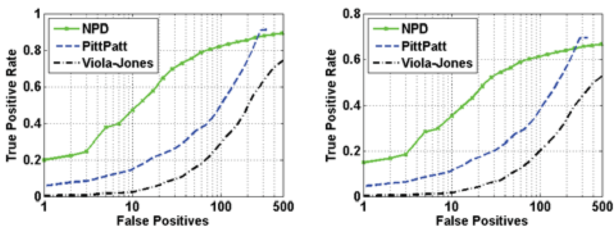
Fig. 10 shows some examples of detected faces in the FDDB database by the proposed NPD method. Rotated, occluded, and out-of-focus faces can be successfully detected by the proposed method in Fig. 10. Some occluded faces (e.g. 4th row and 2nd column of Fig. 10) and blurred faces (e.g. top-right image in Fig. 10) that are not annotated in the ground truth can still be detected by the proposed method. However, there are still a number of faces that cannot be detected by the proposed method, especially in very crowded scenes, observed in the 1st image and the 3rd image in row 1, and the bottom-right image of Fig. 10.

5.3 Evaluation on GENKI Database

The GENKI database [43] was collected by the Machine Perception Laboratory, University of California, San Diego. We evaluated the current release of the GENKI database, GENKI-R2009a, on its SZSL subset, which contains 3,500 images collected from the Internet. These images include a wide range of backgrounds, illumination conditions, geographical locations, personal identity, and ethnicity. Some examples of face images from the GENKI database are shown in Fig. 12, with labeled detections by the proposed NPD method. Most images in the GENKI dataset contain only one face. Therefore, the GENKI dataset is not as challenging as the FDDB dataset. Some of the images in the GENKI-SZSL dataset contain faces that are not labeled, therefore they are not suitable for the face detection evaluation task. After removing such unlabeled images, we are left with 3,270 images for face detection evaluation. For performance evaluation, it is not fair to apply the learned detector described in Section 5.1, because the training data used for that detector contained face images from the GENKI database[5]. Therefore, we evaluated the NPD face detector trained on the first fold of the FDDB 10-fold cross validation. We also evaluated the Viola-Jones face detector implemented in OpenCV2.4.1, and a state-of-the-art commercial face detector PittPatt [44]. We used the benchmark evaluation code by Jain and Learned-Miller [3] for performance evaluation, but slightly modified the code for allowing ground truth annotations of rectangles. The ROC curves of the three methods are shown in Fig. 11 for both the discrete and continuous score metrics. The results show that the proposed NPD face detector performs much better than both the Viola-Jones and PittPatt face detectors, indicating that the proposed algorithm is more suited for the general unconstrained face detection task.

5. This training data is provided by the authors of [13]. We cannot remove the GENKI face images from this training data because we can only access the raw face images in binary format, without knowing the corresponding filenames and sources.

Fig. 10. Detected faces in the FDDB database [3] by the proposed NPD method. Green boxes are detections by the proposed method, while red ellipses are ground truth annotations.



(a) discrete (b) continuous

Fig. 11. ROC curves for face detection on the GENKI-SZSL dataset [43] with (a) discrete and (b) continuous score metrics.

5.4 Evaluation on CMU-MIT Database

The CMU-MIT face dataset [38] is one of the early benchmark dataset for face detection and it has been routinely used for evaluation of face detection algorithms. The CMU-MIT frontal face data set contains 130 gray-scale images with a total of 511 faces, most of which are not occluded. We applied the same NPD detector described in Subsection 5.1 on this database. We also used the modified benchmark evaluation code from Jain and Learned-Miller [3] with the discrete score metric for performance evaluation. Fig. 13 shows the ROC curves for the proposed NPD face detector, the Soft cascade method [45], the SURF

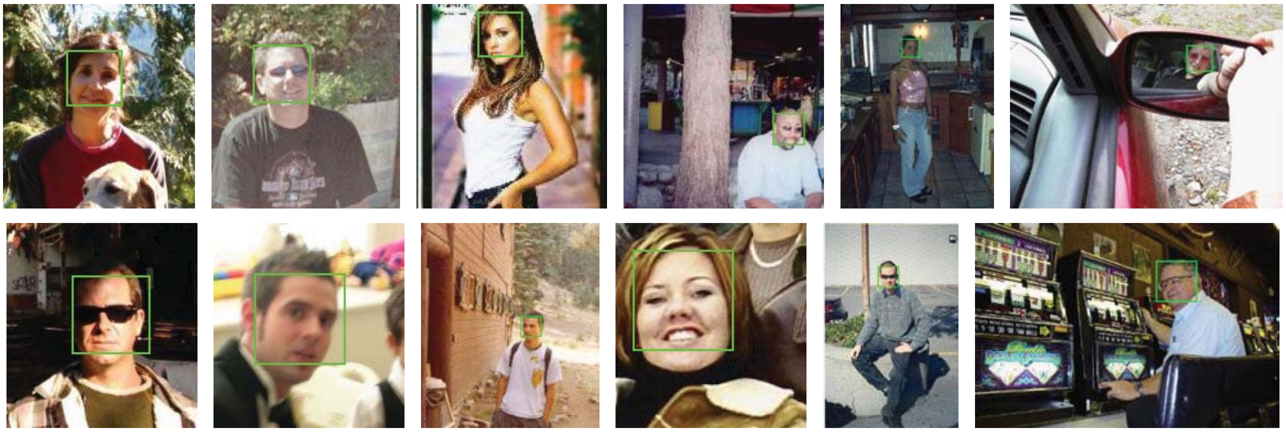


Fig. 12. Detected faces in the GENKI-SZSL dataset [43] by the proposed NPD method.

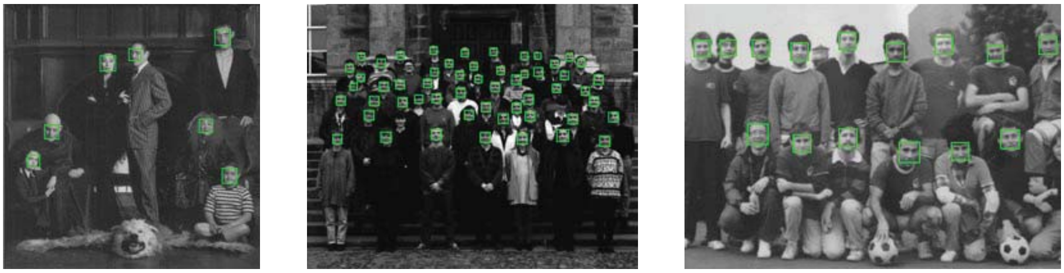


Fig. 14. Detected faces in the CMU-MIT dataset [38] by the proposed NPD method.

cascade method [13], and the Viola-Jones detector [1]. The results show that, compared to the Viola-Jones frontal face detector, the NPD detector performs better when the number of false positives, FP < 50, while it is slightly worse than Viola-Jones at higher FPs. Compared to the SURF cascade detector, the NPD detector is better when FP < 3, but SURF cascade method outperforms NPD at higher FPs. Note that the SURF cascade method used a face template of size 40×40 pixels, which is four times larger than our face detection template (20×20 pixels). Generally, a larger face template contains more features for face description, but is computationally more expensive and may have a limitation in detecting blurred faces. In addition, the proposed method is not as good as the Soft cascade, the state-of-the-art method on the CMU-MIT dataset. The proposed NPD method can detect about 80% of the frontal faces without false positives, which is promising since we did not focus on training a frontal face detector. Some detected faces on the CMU-MIT dataset by the proposed NPD method are shown in Fig. 14.

5.5 Analysis of Contributions

Since the proposed face detector is a combination of regression tree and the NPD features, it is instructive to determine the contribution of each of these two components. First, we trained a detector based on the NPD features, but with the stump classifier [1], which can be regarded as a basic tree classifier with

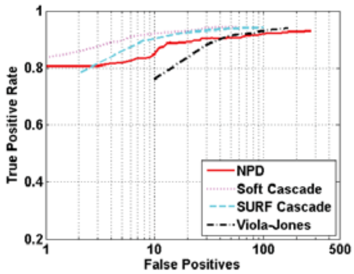
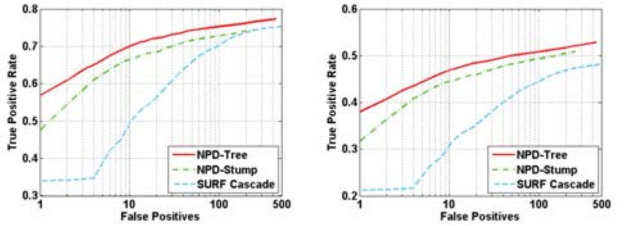


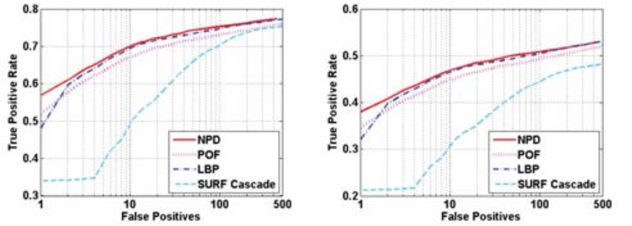
Fig. 13. ROC curves for face detection on the CMU-MIT dataset [38].

only one splitting node. This stump classifier based detector contains 1,597 weak classifiers. In contrast, the regression tree based detector contains 176 weak classifiers, indicating that combining NPD features in a regression tree is much more effective in constructing a weak classifier for AdaBoost learning. Furthermore, in cascade processing, each scanning subwindow needs to evaluate 37 NPD features, on average, for the stump classifier based detector. On the other hand, for the regression tree based detector, 34 NPD features need to be evaluated, which means that for the whole detector, using regression tree as a weak classifier does not increase the computation cost. The face detectors based on the stump classifier and the regression tree were tested on the FDDB database. The ROC curves of these two detectors are shown in Fig. 15 for both the discrete score metric and



(a) discrete (b) continuous

Fig. 15. ROC curves for face detection on the FDDB database [3] with (a) discrete and (b) continuous score metrics.



(a) discrete (b) continuous

Fig. 16. ROC curves for face detection on the FDDB database [3] with (a) discrete and (b) continuous score metrics.

continuous score metric. We include the SURF cascade method [13] as a baseline in these two figures, which also used the same training set. As illustrated, using regression trees instead of stump classifier improves the performance by about 2% – 10% for discrete metric and 1% – 7% for continuous metric. The improvement is larger at smaller false positives. Interestingly, the proposed method with either regression tree or stump based weak classifier clearly outperforms the SURF cascade method, which represents the best published result on the FDDB database to date.

Next, we fixed the regression tree based weak learner, but tried two other local features, namely pixelwise ordinal feature (POF) [34] and LBP [46]. Since LBP is a discrete label, we treated it as a categorical variable in the regression tree learning, that is, for branching each tree node, the algorithm finds the optimal criterion that splits the discrete LBP codes into two groups. In [11], it is shown that LBP feature is better than Haar-like feature for face detection. Using the same training set as in Section 5.1, we trained the two detectors using POF and LBP, respectively. The POF detector learned 276 weak classifiers with 3,082 POF features, while the LBP detector learned 108 weak classifiers with 1,269 LBP features. In contrast, the NPD detector learned 176 weak classifiers with 2,035 NPD features. However, it should be noted that each LBP feature needs to compare 8 pairs of pixels and convert the resulting binary string to the corresponding decimal number. The three detectors were tested on the FDDB database, and the corresponding ROC curves are shown in Fig. 16 for both the discrete and continuous score metrics. The SURF cascade method [13] trained on the same dataset is also included for comparison. It can be observed that the NPD detector performs better than both the POF detector and the LBP detector with the regression tree based weak learners. NPD is better than POF, because with NPD features the regression tree learns optimal thresholds to form more robust ordinal rules. At low false positives, both NPD and POF are better than LBP, indicating that combining optimal pixel-level features in regression trees provides better discrimination between real faces and difficult nonface samples. Besides these findings, the experimental results show that, the NPD, POF, and LBP detectors in our learning framework perform much better than the SURF cascade method in detecting unconstrained faces.

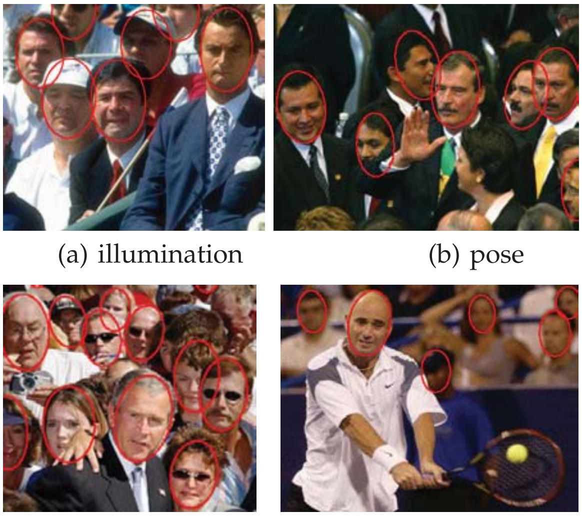
5.6 Evaluation with Major Challenges

In the following, we evaluate how the proposed NPD face detector performs under four major challenges for unconstrained face detection, namely, illumination variation, pose variation, occlusion, and blur (or low resolution). Note that these four challenges are often encountered simultaneously, which means that each testing image may involve more than one type of challenge. In our selection of the four subsets, we focused on the main challenge of each image. For each challenge, we selected 100 images from the FDDB database [3] (examples are shown in Fig. 17), and ran the NPD detector described in Subsection 5.1 on these four subsets separately, resulting in ROC curves shown in Fig. 18. It can be observed that the NPD face detector performs the best on the illumination subset, which is due to the proposed NPD feature's robustness against illumination variations. Further, the NPD method performs better for face images with pose variation than with occlusion or blur. These results indicate that occlusion and blur are the two major challenges for unconstrained face detection, which have not been well addressed in the literature.

The NPD face detector is also compared with the Viola-Jones face detector implemented in OpenCV2.4.1, and the commercial face detector PittPatt on the four subsets of FDDB discussed above and shown in Fig. 17. The resulting ROC curves with the discrete score metric are shown in Fig. 19. These plots show that the proposed NPD face detector outperforms both the Viola-Jones and the PittPatt face detectors on the four subsets. The reasons for the superior performance of the proposed method under illumination variations, pose variations, occlusions, and blur, were discussed in Subsection 4.2.

6 Summary and Future Work

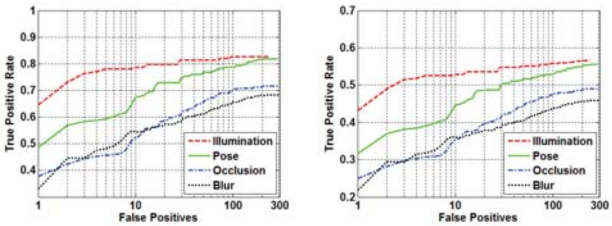
We have proposed a method for unconstrained face detection in cluttered scenes, which is based on the



(a) illumination (b) pose

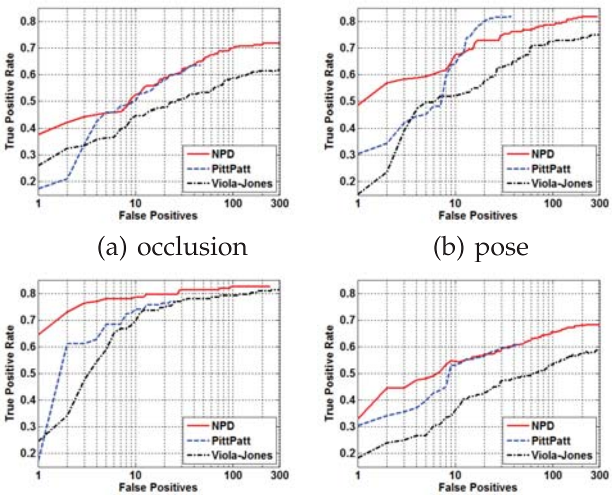
(c) occlusion (d) blur

Fig. 17. Example images and annotated faces for four subsets from the FDDB database [3].



(a) discrete (b) continuous

Fig. 18. ROC curves of the proposed NPD face detector on four subsets from the FDDB database [3] with (a) discrete and (b) continuous score metrics.



(a) occlusion (b) pose

(c) illumination (d) blur

Fig. 19. ROC curves for face detection on four subsets from the FDDB database [3] with the discrete score metric.

normalized pixel difference (NPD) feature in conjunction with boosted regression trees. We have shown that the proposed NPD features are discriminative and robust for the unconstrained face detection task. An analysis of NPD feature shows its property of scale invariance, boundedness, and reconstruction ability. We have developed a method for learning optimal NPD features and their combinations by boosted regression trees to handle complex distribution of faces in unconstrained conditions. It is shown that, a single cascade AdaBoost classifier is able to achieve promising results for unconstrained face detection with large pose variations and occlusions. Evaluations on the FDDB, GENKI, and CMU-MIT datasets show that the proposed method outperforms state-of-the-art methods for unconstrained face detection. The speed of the proposed NPD face detector implemented with MATLAB is also shown to be comparable to that of the Viola-Jones face detector implemented in OpenCV. The reported results also show that occlusions and blur are two big challenges for unconstrained face detection. Our future work will use the NPD feature and the classifier learning method for other applications such as face attribute classification (e.g. pose estimation, age estimation, and gender classification) and pedestrian detection.

Appendix A
Proof of Lemma 4

From the definition of NPD we know that $x \geq 0$, $y \geq 0$, and $f(0,0) = 0 \in [-1,1]$. When either x or y is nonzero, for example, $y \geq 0$ but $x > 0$, Eq. (1) can be reformulated as

$$f(x,y) = \frac{x-y}{x+y} = \frac{2x}{x+y} - 1 = \frac{2}{1+\frac{y}{x}} - 1 \leq 1. \tag{6}$$

The inequality is because of $y \geq 0$, and the last equality holds if and only if $x > 0$ and $y = 0$. Similarly, when $x \geq 0$ but $y > 0$, Eq. (1) can be reformulated as

$$f(x,y) = \frac{x-y}{x+y} = 1 - \frac{2y}{x+y} = 1 - \frac{2}{\frac{x}{y}+1} \geq -1. \tag{7}$$

The inequality is because of $x \geq 0$, and the last equality holds if and only if $x = 0$ and $y > 0$. □

Appendix B
Proof of Theorem 1

Denote $f_{ij} = f(x_i, x_j)$. From Eq. (1) we have

$$f_{ij}(x_i + x_j) = x_i - x_j. \tag{8}$$

Equivalently,

$$(f_{ij} - 1)x_i + (f_{ij} + 1)x_j = 0. \tag{9}$$

Therefore, we have the following linear equations

$$\mathbf{Fx} = \mathbf{0}, \tag{10}$$

where

$$\mathbf{F} = \begin{pmatrix} f_{12}-1 & f_{12}+1 & 0 & \cdots & 0 \\ f_{13}-1 & 0 & f_{13}+1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{1p}-1 & 0 & 0 & \cdots & f_{1p}+1 \\ 0 & f_{23}-1 & f_{23}+1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & f_{p-1,p}+1 \end{pmatrix} \tag{11}$$

is a sparse $d \times p$ matrix with each row containing at most two nonzero entries. Furthermore, from the formulation of $\mathbf{F}$ we know that each row of $\mathbf{F}$ contains at least one nonzero entry, because $f_{ij}-1 \neq f_{ij}+1$ always holds for all i and j. Without loss of generality, let's assume $f_{12}+1 \neq 0$. Then it follows that $f_{1j}+1 \neq 0, \forall j$. Because if $\exists j$ such that $f_{1j}+1=0$, then from Lemma 4 we know that $x_1 = 0$. This will further lead to $f_{12}+1=0$, a conflict with the assumption. Therefore, the first $p-1$ rows are linearly independent of each other.

We will further prove that $rank(\mathbf{F}) = p-1$. In fact, any other row of the matrix $\mathbf{F}$ can be linearly expressed by the first $p-1$ rows. To show this, let's denote the row containing $f_{ij}-1$ and $f_{ij}+1$ by $\mathbf{r}_{ij}$. We will show that

$$\mathbf{r}_{ij} = -\frac{f_{ij}-1}{f_{1i}+1}\mathbf{r}_{1i} - \frac{f_{ij}+1}{f_{1j}+1}\mathbf{r}_{1j}, \tag{12}$$

holds for all $i > 1$ and $j > i$. In fact, it is easy to verify that the above equation holds for all columns of $\mathbf{r}_{ij}$, $\mathbf{r}_{1i}$, and $\mathbf{r}_{1j}$ after the first column. So, we only need to show that, for the first column, we have

$$-\frac{(f_{1i}-1)(f_{ij}-1)}{f_{1i}+1} - \frac{(f_{1j}-1)(f_{ij}+1)}{f_{1j}+1} = 0, \tag{13}$$

which is equivalent to

$$f_{1i}f_{1j}f_{ij} - f_{1i} + f_{1j} - f_{ij} = 0. \tag{14}$$

This can be verified by substituting each feature with its definition in Eq. (1).

Given $rank(\mathbf{F}) = p-1$, we know that the nullspace of $\mathbf{F}$ contains only one nonzero vector, which is a solution to Eq. (10). Furthermore, from Lemma 4 we can infer that $(f_{ij}-1)(f_{ij}+1) \leq 0$, hence Eq. (9) tells that $x_i x_j \geq 0, \forall i, j$. Consequently, Eq. (10) always has a nonnegative solution $\hat{\mathbf{x}}$, and all solutions to Eq. (10) must be $c\hat{\mathbf{x}}$, where c is a scale factor. □

We make four observations below:

- For a solution, c can be any real value, but to satisfy the constraint that all pixel intensity values are nonnegative, c should be positive.
- The solution to Eq. (10) spans a one-dimensional subspace (the nullspace).
- A specific solution can be obtained by assigning $x_1 = 1$ and solving for the other variables from the first $p-1$ rows of Eq. (10) in linear time.
- When the original image is $\mathbf{x} = \mathbf{0}$, it can also be reconstructed by $c\hat{\mathbf{x}}$ where $\hat{\mathbf{x}}_i = 1, \forall i$, and $c = 0$. However, in this case a solution with $c > 0$ is not generally regarded as a scaled version of the original image $\mathbf{x} = \mathbf{0}$.

References

[1] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.

[2] J. Friedman, T. Hastie, and R. Tibshirani, "Additive logistic regression: a statistical view of boosting," *The Annals of Statistics*, vol. 28, no. 2, pp. 337–374, April 2000.

[3] V. Jain and E. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," University of Massachusetts, Amherst, Tech. Rep. UM-CS-2010-009, 2010.

[4] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proceedings of the IEEE International Conference on Image Processing*, 2002.

[5] S. Li, L. Zhu, Z. Zhang, A. Blake, H. Zhang, and H. Shum, "Statistical learning of multi-view face detection," in *Proceedings of the 7th European Conference on Computer Vision*, 2002.

[6] M. Jones and P. Viola, "Fast multi-view face detection," *Mitsubishi Electric Research Lab TR-2003-96*, 2003.

[7] B. Froba and A. Ernst, "Face detection with the modified census transform," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004.

[8] H. Jin, Q. Liu, H. Lu, and X. Tong, "Face detection using improved LBP under bayesian framework," in *Proceedings of the 3rd International Conference on Image and Graphics*, 2004.

[9] 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*, vol. 2, 2005, pp. 1619–1626.

[10] H. Zhang, W. Gao, X. Chen, and D. Zhao, "Object detection using spatial histogram features," *Image and Vision Computing*, vol. 24, no. 4, pp. 327–341, 2006.

[11] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li, "Face detection based on multi-block LBP representation," in *Proceedings of the IAPR/IEEE International Conference on Biometrics*, 2007.

[12] S. Yan, S. Shan, X. Chen, and W. Gao, "Locally assembled binary (LAB) feature with feature-centric cascade for fast and accurate face detection," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.

[13] J. Li, T. Wang, and Y. Zhang, "Face detection using SURF cascade," in *ICCV BeFIT workshop*, 2011.

[14] B. Wu, H. Ai, C. Huang, and S. Lao, "Fast rotation invariant multi-view face detection based on real AdaBoost," in *Proceedings of the 6th IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 79–84.

[15] S. 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.

[16] 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.

[17] K. Hotta, "A robust face detector under partial occlusion," in *International Conference on Image Processing*, 2004.

[18] Y. Lin, T. Liu, and C. Fuh, "Fast object detection with occlusions," in *Proceedings of the European Conference on Computer Vision*, 2004, pp. 402–413.

[19] Y. Lin and T. Liu, "Robust face detection with multi-class boosting," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.

[20] J. Chen, S. Shan, S. Yang, X. Chen, and W. Gao, "Modification of the adaboost-based detector for partially occluded faces," in *18th International Conference on Pattern Recognition*, 2006.

[21] L. Goldmann, U. Monich, and T. Sikora, "Components and their topology for robust face detection in the presence of partial occlusions," *IEEE Transactions on Information Forensics and Security*, vol. 2, no. 3, pp. 559–569, 2007.

[22] M. Yang, D. 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.

[23] E. H. Weber, "Tastsinn und gemeingefühl," in *Handwörterbuch der Physiologie*, R. Wagner, Ed. Brunswick: Vieweg, 1846, pp. 481–588.

[24] C. Zhang and Z. Zhang, “A survey of recent advances in face detection,” Microsoft Research, Tech. Rep. MSR-TR-2010-66, June 2010.
[25] R. 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.
[26] J. B. Martinkauppi, A. Hadid, and M. Pietikänen, “Skin color in face analysis,” in *Handbook of Face Recognition*, 2nd ed., S. Z. Li and A. K. Jain, Eds. Springer, Sep. 2011, ch. 9, pp. 223–249.
[27] V. Jain and E. Learned-Miller, “Online domain adaptation of a pre-trained cascade of classifiers,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011.
[28] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.
[29] P. Sinha, “Qualitative representations for recognition,” in *Proceedings of the Second International Workshop on Biologically Motivated Computer Vision*, Tubingen, Germany, 2002.
[30] J. Sadr, S. Mukherjee, K. Thoresz, , and P. Sinha, “Toward the fidelity of local ordinal encoding,” in *Proceedings of the Fifteenth Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 3-8 2001.
[31] S. Baluja, M. Sahami, and H. Rowley, “Efficient face orientation discrimination,” in *International Conference on Image Processing*, vol. 1, 2004, pp. 589–592.
[32] S. Liao, Z. Lei, X. Zhu, Z. Sun, S. Z. Li, and T. Tan, “Face recognition using ordinal features,” in *Proceedings of the 1st IAPR International Conference on Biometrics*, Hong Kong, 2006.
[33] Y. Abramson, B. Steux, and H. Ghorayeb, “Yet even faster (YEF) real-time object detection,” *International Journal of Intelligent Systems Technologies and Applications*, vol. 2, no. 2, pp. 102–112, 2007.
[34] L. Wang, L. Ding, X. Ding, and C. Fang, “2D face fitting-assisted 3D face reconstruction for pose-robust face recognition,” *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, vol. 15, no. 3, pp. 417–428, 2011.
[35] R. Lienhart, A. Kuranov, and V. Pisarevsky, “Empirical analysis of detection cascades of boosted classifiers for rapid object detection,” MRL, Intel Labs, Tech. Rep., May 2002.
[36] S. Brubaker, J. Wu, J. Sun, M. Mullin, and J. Rehg, “On the design of cascades of boosted ensembles for face detection,” Georgia Institute of Technology, Tech. Rep. GIT-GVU-05-28, 2005.
[37] L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, *Classification and Regression Trees*. Chapman & Hall/CRC, 1984.
[38] H. Rowley, S. Baluja, and T. Kanade, “Rotation invariant neural network-based face detection,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1998.
[39] E. Seemann, B. Leibe, and B. Schiele, “Multi-aspect detection of articulated objects,” in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.
[40] T. Kim and R. Cipolla, “MCBoost: Multiple classifier boosting for perceptual co-clustering of images and visual features,” *Proceedings of Neural Information Processing Systems*, 2008.
[41] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, “WLD: A robust local image descriptor,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, Sept. 2010.
[42] F. Kriegler, W. Malila, R. Nalepka, and W. Richardson, “Preprocessing transformations and their effects on multispectral recognition,” in *Proceedings of the Sixth International Symposium on Remote Sensing of Environment*, 1969, pp. 97–131.
[43] http://mplab.ucsd.edu, “The MPLab GENKI Database, GENKI-SZSL Subset.”
[44] PittPatt Software Developer Kit, Pittsburgh Pattern Recognition, Inc., http://www.pittpatt.com.
[45] L. Bourdev and J. Brandt, “Robust object detection via soft cascade,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2005, pp. 236–243.
[46] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

Shengcai Liao received the B.S. degree in mathematics and applied mathematics from the Sun Yat-sen University, Guangzhou, China, in 2005 and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences (CASIA), Beijing, China, in 2010. He was a Post Doctoral Fellow in the Department of Computer Science and Engineering, Michigan State University during 2010-2012, and he is now an Assistant Professor in CASIA. His research interests include computer vision, pattern recognition, and machine learning, with a focus on image and video analysis, particularly face recognition, object detection and recognition, and video surveillance. He has coauthored more than 30 papers in international journals and conferences including IEEE TPAMI, IEEE TIP, ICCV, and CVPR. He was awarded the Motorola Best Student Paper award and the 1st Place Best Biometrics Paper award at the International Conference on Biometrics in 2006 and 2007, respectively, for his work on face recognition.

Anil K. Jain is a university distinguished professor in the Department of Computer Science and Engineering at Michigan State University. His research interests include pattern recognition and biometric authentication. He served as the editor-in-chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (1991-1994). The holder of six patents in the area of fingerprints, he is the author of a number of books, including Handbook of Fingerprint Recognition (2009), Handbook of Biometrics (2011), Handbook of Multibiometrics (2006), Handbook of Face Recognition (2005), BIOMETRICS: Personal Identification in Networked Society (1999), and Algorithms for Clustering Data (1988). He served as a member of the Defense Science Board and The National Academies committees on Whither Biometrics and Improvised Explosive Devices. Dr. Jain received the 1996 IEEE TRANSACTIONS ON NEURAL NETWORKS Outstanding Paper Award and the Pattern Recognition Society best paper awards in 1987, 1991, and 2005. He is a fellow of the AAAS, ACM, IAPR, and SPIE. He has received Fulbright, Guggenheim, Alexander von Humboldt, IEEE Computer Society Technical Achievement, IEEE Wallace McDowell, ICDM Research Contributions, and IAPR King-Sun Fu awards.

Stan Z. Li received the B.Eng. degree from Hunan University, Changsha, China, the M.Eng. degree from the National University of Defense Technology, China, and the Ph.D. degree from Surrey University, Surrey, U.K. He is currently a Professor and the Director of Center for Biometrics and Security Research (CBSR), Institute of Automation, Chinese Academy of Sciences (CASIA). He worked at Microsoft Research Asia as a researcher from 2000 to 2004. Prior to that, he was an Associate Professor at Nanyang Technological University, Singapore. His research interest includes pattern recognition and machine learning, image and vision processing, face recognition, biometrics, and intelligent video surveillance. He has published over 200 papers in international journals and conferences, and authored and edited eight books. Dr. Li is currently an Associate Editor of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE and is acting as the Editor-in-Chief for the Encyclopedia of Biometrics. He served as a co-chair for the International Conference on Biometrics 2007 and 2009, and has been involved in organizing other international conferences and workshops in the fields of his research interest.