Face Detection on Distorted Images Augmented by Perceptual Quality-Aware Features

Suriya Gunasekar, Joydeep Ghosh, *Fellow, IEEE*, and Alan C. Bovik, *Fellow, IEEE*

***Abstract*— Motivated by the proliferation of low-cost digital cameras in mobile devices being deployed in automated surveillance networks, we study the interaction between perceptual image quality and a classic computer vision task of face detection. We quantify the degradation in performance of a popular and effective face detector when human-perceived image quality is degraded by distortions commonly occurring in capture, storage, and transmission of facial images, including noise, blur, and compression. It is observed that, within a certain range of perceived image quality, a modest increase in image quality can drastically improve face detection performance. These results can be used to guide resource or bandwidth allocation in acquisition or communication/delivery systems that are associated with face detection tasks. A new set of features, called qualHOG, are proposed for robust face-detection that augments face-indicative Histogram of Oriented Gradients (HOG) features with perceptual quality-aware spatial Natural Scene Statistics (NSS) features. Face detectors trained on these new features provide statistically significant improvement in tolerance to image distortions over a strong baseline. Distortion-dependent and distortion-unaware variants of the face detectors are proposed and evaluated on a large database of face images representing a wide range of distortions. A biased variant of the training algorithm is also proposed that further enhances the robustness of these face detectors. To facilitate this research, we created a new distorted face database (DFD), containing face and non-face patches from images impaired by a variety of common distortion types and levels. This new data set and relevant code are available for download and further experimentation at www.live.ece.utexas.edu/research/Quality/index.htm.**

***Index Terms*— Face detection, no reference image quality, spatial NSS, surveillance.**

I. Introduction

THE advent of affordable digital storage devices and powerful, network pervasive visual data sharing websites such as Flickr, Facebook, Instagram etc., has caused an explosion of visual data that is being generated and shared at an exponentially growing rate. While the principal consumers of visual data are humans, practical machine vision deployments are becoming more commonplace. In both realms, automated methods for culling, sharing, organizing, and understanding large volumes of visual content is highly desirable. Computer vision algorithms that aim to understand visual content are being increasingly employed in real life applications such as image search, automated surveillance, human computer interfaces, etc. A primary component of many computer vision algorithms is some form of an object detection/recognition system. Such systems are often prone to performance degradation when the quality of the input images deteriorates. One such task that has resulted in successful commercial embodiment is automatic face detection. Face detection in inexpensive mobile or outdoor devices commonly used for surveillance is often highly unconstrained and subject to quality–destructive distortions, that can adversely affect detection performance. Since face detection is usually a precursor to advanced tasks of recognition, expression tracking, etc., understanding the relationship between face quality and detectability is important.

Substantial research efforts have recently focused on the development of automated image quality algorithms (IQA) that aim to accurately predict end–user quality–of–experience. These include *Full Reference (FR)* algorithms [1], [2], in which the fidelity of a test image to a presumed undistorted reference version is evaluated, *No Reference (NR)* algorithms [3]–[6], which do not use any information from reference images, and the intermediate *Reduced Reference (RR)* algorithms [7], [8], which use partial information available about reference images. Among these, NR algorithms have the greatest potential for many practical settings, since references are seldom available. General purpose NR frameworks that rely on models of natural statistics of images have been recently shown to provide state–of–the–art performance in predicting perceived image quality [4], [5], [9].

Another exciting direction of inquiry is the interaction between visual quality and visual tasking. A small body of work exists on how quality affects biometric tasks (iris, face, fingerprint detection and recognition) [10]–[13]. These papers study various image factors that affect the detection or recognition performance. For example, ISO/IEC 19794-5 [14] specifies a list of factors such as spectacles, pose, centering, occlusion, expression, head shape, etc., that affect "face quality". While these do affect detection and recognition, there is no clear distinction between scene–dependent challenges like occlusion, illumination, etc., and the challenges imposed by traditional notions of "quality impairments" from capture, compression, processing, transmission, etc. In this paper, we are concerned with the latter interpretation of "quality" as it affects face detection performance. This is an important line of work as in many facial acquisition and communication channels, the effects of such quality impair-

Manuscript received March 17, 2014; revised July 28, 2014 and September 3, 2014; accepted September 3, 2014. Date of publication September 29, 2014; date of current version November 10, 2014. This work was supported by the National Science Foundation Office of the Director under Grant IIS-1116656. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Aly A. Farag.

The authors are with the Department of Electrical and Computer Engineering, University of Texas at Austin, Austin, TX 78712 USA (e-mail: suriya@utexas.edu; ghosh@ece.utexas.edu; bovik@ece.utexas.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2014.2360579

ments on detection/recognition can often be mitigated, e.g., by reallocating resources such as bandwidth. Further, in spite of numerous algorithms proposed in the past few decades [15], [16], face detection methods that are robust to image distortions have not been widely explored.

This work is motivated by the fact that the IQA models described above were designed to predict the perceptual quality of digital images but have not been applied to visual task models involving faces. We explore the important question of whether the perceptual quality of facial images is a good predictor of the success of algorithm performance on visual tasks. The question is quite relevant given that the human visual apparatus is remarkably well adapted to analyzing faces. Early works by Rouse et. al. in this direction [17]–[19] show that perceptual FR IQA algorithms, including VIF and SSIM, correlate strongly with "recognizing thresholds" of human observers. However, the effects of quality on machine vision algorithms have not been evaluated, and moreover FR algorithms are of limited use in this regard due to the unavailability of reference images in most practical scenarios.

A trade–off exists between ground truth image distortion and face detection performance. In many image/video communication channels, the distortion levels could be adjusted using channel parameters to obtain a required level of face detection performance. Therefore, we begin by investigating the effects of different types and magnitudes of distortion on face detection performance. However, in most practical scenarios accurate measures of distortion types and levels is not available. Therefore, we resort to using an easily obtainable proxy for actual distortions, namely, the human visual-quality aware NIQE score. Empirical studies reveal that this proxy yields qualitatively similar results and also retains relative performance results when compared to those provided by using the actual distortion measures. We then show that as with true distortion levels, over a range of objective quality scores delivered by a high–performance NR image quality model, moderate improvements in predicted quality can significantly aid face detection performance.

Secondly, we show that the use of easily computable "quality–aware" *spatial Natural Scene Statistics (NSS)* features [6] has the potential to greatly assist the design of more robust face detection algorithms. The widely–used *Histogram of Oriented Gradients (HOG)* based detection algorithm [20] is used as the baseline in our experiments. We use this model because it is flexible and easily reconfigured to enable the inclusion of features related to image quality.

Finally, existing face detection datasets[1] consist of samples of face and non–face patches. However, our goal is to investigate the performance of face detectors on images corrupted by common distortions such as gaussian blur and JPEG compression, whose effects are not pixel-wise but more global. So distortions isolated on local patches can exert a different effect on detection performance as compared to distortions on the entire image. Thus, we curated a new Distorted Face Database (DFD), from the web for our experiments. This new dataset consists of face and non–face patches from images that were (globally) distorted with known distortion types and levels. The dataset is available for download and further experimentation at www.live.ece.utexas.edu/research/Quality/index.htm.

[1] http://www.facedetection.com/facedetection/datasets.htm

The main contributions of this paper are as follows:

1) The performance degradation of a widely used HOG based face detector [20] with respect to the response of a high–performance NR image quality algorithm called NIQE is studied on images distorted by three common distortions: additive white gaussian noise, gaussian blur and JPEG compression. We experimentally show that over a certain range of NIQE scores, a modest improvement in image quality can significantly improve detection performance.
2) We show that the readily computable NIQE score is a valid and suitable proxy for actual distortion in the absence of knowledge about the original (reference) images or the actual types and/or levels of distortions, in terms of studying the effect of such distortions on quality of algorithmic face detectors.
3) A new set of *QualHOG* features are proposed that augments face–indicative *HOG* features with perceptual "quality–aware" *spatial–NSS* features. Face detectors learned on these features provide improved tolerance against image distortions. We experimentally quantify the degree of resulting improvements.
4) A modification to the cost function used by the classifier (an SVM) is proposed which further enhances the robustness of the QualHOG based face detector.
5) A new Distorted Face Database (DFD) was created that has face and non–face patches from images that were distorted using known distortion types and levels.

In Section II, we review relevant literature on image quality assessment and face detection algorithms. The distortion types investigated and the proposed model for robust face detection are discussed in Section III. In Sections IV and V we describe the experimental setup and the results, respectively. We conclude with directions for future work in Section VI.

II. Related Work

In this paper we combine ideas from two problems in vision science and computer vision: image quality assessment and face detection. *Image quality assessment (IQA)* aims at predicting the quality of a given image as perceived by human users. The performance of IQA models are assessed by measures of correlation between objective predicted quality scores and aggregated human opinions (Differential Mean Opinion Scores (DMOS)) on a set of representative test images. *Face detection* is a fundamental problem in various computer vision applications including camera focusing, and is a precursor to advanced tasks of identification, tracking, etc. Efficient and accurate algorithms for face detection have been widely developed over the past few decades. The problem of face detection involves accurately identifying the region(s) in an arbitrary image that corresponds to human face(s). In the rest of this section, we review some relevant literature pertaining to these two problems.

As stated previously, IQA algorithms can be broadly categorized as *Full Reference (FR)*, *Reduced Reference (RR)* and *No Reference (NR)* models. While the presence of a reference image or information regarding references simplifies the problem, in real–life applications FR and RR algorithms are limited in scope as the reference information is generally unavailable at nodes where quality computation is undertaken. Hence, we concentrate only on NR IQA models as they are much more likely to be of use in practical vision applications. Early NR IQA models were distortion specific [21], [22]. Such algorithms extract distortion–specific features that relate to loss of visual quality, such as ringing, blur, edge–strength at block boundaries, etc. While these provide satisfactory performance for specific distortion types, often the distortion type that is actually encountered is unknown beforehand or is poorly modeled. Thus, a few distortion–independent approaches to the NR IQA problem have been proposed recently [4]–[6]. These models are based on the hypothesis that natural images follow regular statistical properties that are modified by the presence of distortions. Deviations from the regularity of these *natural scene statistics (NSS)* are indicative of perceptual quality of images. Hence, models based on the quantification of the *naturalness* of an image are useful for creating distortion–independent measures of perceived quality.

For example, the *DIIVINE* index [4] deploys summary statistics derived from the NSS models of wavelet coefficients. These features are used to first identify most likely distortion types followed by distortion specific quality assessment. A similar approach named *BLIINDS–II* [5], operates in the DCT domain. A small number of features are computed from an NSS model of block DCT coefficients. These features are in turn used to train a regression model that delivers accurate quality predictions. While both DIIVINE and BLIINDS–II deliver superior performance for assessing image quality, computation of the features involved is expensive and hence deploying these models in real time is difficult.

Scalable transform–free (spatial) models for NR IQA were recently developed by Mittal et al. [6], [9]. The *BRISQUE* and *NIQE* indices proposed in these works operate directly on multiscale spatial pixel data and hence are inexpensive to compute. These models are based on the statistics of locally debiased and divisive normalized luminance coefficients that quantify the deviation from *naturalness* of an image due to the presence of distortions. The debiasing and divisive normalization of spatial pixels are motivated by well–accepted models of front end coding by the human visual apparatus. BRISQUE uses quality–aware spatial features to train a regression model for IQA, while NIQE develops a model for undistorted "pristine" images and measures deviations of the statistics of the test image from the pristine image model. Despite using purely spatial features, these models show performance comparable to DIIVINE and BLIINDS–II at a small fraction of the computation. Going forward, we will use the spatial–NSS features used by BRISQUE and NIQE as quality–aware features.

Some of the early work on face detection was surveyed by Hjelmas et al. [16], Yang et al. [23], and more recently by Zhang et al. [15]. Early face detection algorithms have been categorized as *knowledge based methods*, which use predefined rules to detect faces, or as *feature invariant methods*, which use pose and lighting invariant features, or as *template matching methods*, which detects faces by matching against pre–stored templates, or finally as *appearance based methods*, which model faces from a set of representative training faces.

Most of the recent algorithms for face detection could be categorized as appearance based methods. A typical practice is to collect certain indicative features from a training set of face and non–face image patches and use machine learning algorithms to learn a classifier for detecting other faces. The two key variants among these algorithms are the type of features used and the kind of classifier employed.

Boosting algorithms have been a popular choice in the literature. AdaBoost, RealBoost, and GentleBoost are some of the popular methodologies in this framework and they have been compared by Lienhart et al. [24] and Brubaker et al. [25]. The Viola–Jones algorithm [26] for face detection has had a large impact on face detection research because of its low testing time that has made face detection feasible in real time. The algorithm uses simple Haar–like features to train weak classifiers in a multi–stage boosting algorithm [27]. However, while the computation required for testing an image for faces is real–time, the training of the cascaded classifier in the Viola–Jones face detector requires exorbitant computation. For example, using the implementation in OpenCV, training a Haar cascade classifier takes about a week. Moreover, the cascaded classifier structure works efficiently only with a highly restrictive set of Haar–like features which limits accuracy. Finally, Viola-Jones does not provide a mechanism to investigate the trade-off between true and false positive, so that AUROC/AUPR based comparisons are not possible.

More recently, regional image statistics features are being used increasingly for face detection. With the advent of more complex features, various single stage classifiers such as Bayesian classifiers and support vector machines (SVMs) have gained popularity. An extensive survey of various other features used by recent face detection algorithms is provided by Zhang et al. [15]. Dalal et al. [20], introduced a popular regional statistics based feature called the *Histogram of Oriented Gradients (HOG)* and used a linear SVM classifier to detect humans in an image. These features are invariant to 2D rotations and illuminations. The baseline used for comparison in this paper is an adaptation of the human detector proposed by Dalal et al. [20], for the problem of face detection [28].

Studying the effects of quality impairments on detection and recognition tasks is of interest as it can be exploited to mitigate the effects of such impairments on relevant tasks. Some work can be found in the literature that study the effects of image quality on object detection/recognition performance [17]–[19], [29]. Rouse et al. take a broad view of quality vs. tasking [17]–[19]. Recognizing the importance of perceptual principles in both visual tasks and in quality assessment, the authors study human "recognition thresholds" of objects as a function of objective image quality as measured by the FR algorithms, multiscale SSIM and VIF.

They find that perception–driven FR IQA indices can indeed successfully predict image recognizability [18], [19]. Likewise, Gala et al. [29] find that the SSIM metric can be used to predict the performance of tracking algorithms with a high degree of confidence. However, as mentioned previously, FR and RR algorithms are limited in their applicability and hence we investigate face detection performance as a function of image quality predicted by a state–of–the–art NR algorithm. This is of particular importance for oncoming wireless vision applications, where intelligent, robust blind algorithms are needed, where severe distortions occur, and where facial images are becoming increasingly important in both consumer and/or security applications.

III. QualHOG Based Face Detectors

We first describe the types of image distortions that we consider. A new quality–aware face detector called QualHOG, which uses face–indicative HOG features and quality–indicative spatial–NSS features, is then discussed. We finally motivate and propose a modification to the cost function of the SVM classifier to further enhance QualHOG.

A. Image Distortions

We consider three basic types of distortions that commonly occur in digital devices and over communication channels. The image is denoted by a matrix I, such that $I(i, j)$ represents the $(i, j)^{\text{th}}$ pixel in the image I.

1) AWGN (σ_N^2), Additive White Gaussian Noise: This is a local distortion, in which a zero mean gaussian noise of variance parameter σ_N^2 is added independently to each pixel.

$$\tilde{I}(i, j) = I(i, j) + N_{ij}, \text{ such that } N_{ij} \sim \mathcal{N}(0, \sigma_N^2) \quad (1)$$

where $\mathcal{N}(\mu, \sigma^2)$ is a gaussian distribution with mean μ and variance σ^2. This is a common model for a broadband device or channel noise.

2) Gblur(σ_B), Gaussian Blur: This is a global distortion in which each pixel is blurred through convolution with a gaussian low pass filter of standard deviation σ_B. For computational ease the gaussian kernel is truncated at $6\sigma_B$. The discrete truncated gaussian filter in two dimensions is given as follows:

$$G(x, y) = \frac{1}{2\pi\sigma_B^2} e^{-\frac{x^2+y^2}{2\sigma_B^2}} \quad (2)$$

where $-\lceil 3\sigma_B \rceil \leq x \leq \lceil 3\sigma_B \rceil$ and $-\lceil 3\sigma_B \rceil \leq y \leq \lceil 3\sigma_B \rceil$. An image with gaussian blur distortion is given by $\tilde{I} = I * G$

$$\tilde{I}(i, j) = \sum_{x=-\lceil 3\sigma_B \rceil}^{\lceil 3\sigma_B \rceil} \sum_{y=-\lceil 3\sigma_B \rceil}^{\lceil 3\sigma_B \rceil} I(i + x, j + y) G(x, y) \quad (3)$$

This is a common model for lens blur.

3) JPEG(Q), JPEG Compression: This is the most commonly used lossy compression method for digital photography. The trade–off between storage size and image fidelity is controlled by a "quality factor", $0 \leq Q \leq 100$, where $Q = 100$ corresponds to no compression while lower values of Q lead to higher compression and lower image quality. Note that while Q is generally monotonic with the perceived quality of a compressed image, it is a poor predictor of perceptual image quality. This compression scheme first converts the spatial image into the frequency domain using a discrete cosine transform (DCT). In the DCT domain the DCT coefficients are quantized to reduce storage requirements. The degree of quantization is controlled by the Q factor. If G is the DCT matrix of image I, the quantized DCT matrix, $\tilde{G}$ is given by:

$$\tilde{G}(i, j) = \text{round}\left(\frac{G(i, j)}{\mathbb{Q}(i, j)}\right) \quad (4)$$

where the quantization matrix, $\mathbb{Q}$ (dependent on Q) which is of the same size as G, is designed to provide higher resolution in frequencies that are hypothesized to be perceptually more important.

B. QualHOG Face Detector

The QualHOG patch descriptor consists of two components:

1) Spatial–NSS: The spatial–NSS features used in QualHOG were proposed by Mittal et al. [6] to accomplish blind IQA and consist of parameters describing the natural scene statistics of spatial components. The image patch, I, is preprocessed using local mean removal and divisive normalization:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (5)$$

where (i, j) are spatial indices, $\mu(i, j)$ and $\sigma(i, j)$ are the mean and variance, respectively, of neighborhood pixels weighted by a truncated symmetric 2-D gaussian, and C is the saturation constant (typically $C = 1$) that that stabilizes the division.

The motivation for these NSS features lies in statistical models of photographs and in low–level models of visual perception. It is well established that the early stages of human vision process images locally. These processes have evolved to encode images using natural statistics for efficient neural transmission and representation in higher–level visual tasks [30], [31]. Ruderman [32] hypothesized that the neural channel for transmitting visual signals were constrained by the variance of the signals and hence the optimal coding of images could be attained using gaussian statistics. He established that local mean subtracted and divisive normalized pixel values of natural images (as in Equation 5) regularly obey gaussian histograms. The mean subtraction in the numerator of the equation results from a center–surround band pass operation that approximates post–retinal ganglion processing to obtain residual images with lower entropy; apparently to accomplish predictive coding [33]. The divisive normalization by σ in the denominator models the adaptive gain control process (AGC) in the visual cortex that accomplishes contrast masking as a byproduct [34], [35].

A white gaussian model of (5) is quite regular across good–quality photographic images. However, when images are distorted, histograms of pixels after preprocessing using (5), are generally no longer gaussian. Extensive experimentation with IQA models has shown that the distorted image histograms subject to (5) can be fit using a generalized gaussian

distribution (GGD) and that the deviation of an image from the "true naturalness" can be used to effectively predict distortion types and levels. A zero mean GGD is parameterized by (α, β) and is given as:

$$f(x;\alpha,\beta) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(\frac{-|x|}{\beta}\right)^{\alpha} \tag{6}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$ is the Gamma function.

Moreover, it has been observed that distortions typically introduce unnatural spatial dependencies, which can be measured by examining the distributions of local image correlations [36]. A set of directional (horizontal,vertical and diagonal) spatial features are computed as:

$$\begin{aligned} H(i,j) &= \hat{I}(i,j)\hat{I}(i,j+1) \\ V(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j) \\ D1(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j+1) \\ D2(i,j) &= \hat{I}(i,j)\hat{I}(i+1,j-1) \end{aligned} \tag{7}$$

The histograms of each directional components, $\{H(i,j)\}$, $\{V(i,j)\}$, $\{D1(i,j)\}$ and $\{D2(i,j)\}$ are fit using a zero mode asymmetric generalized gaussian distribution (AGGD), which is parameterized by $(\gamma, \beta_l, \beta_r)$ as:

$$f(x;\alpha,\beta) = \begin{cases} \frac{\gamma}{2(\beta_l+\beta_r)\Gamma(1/\gamma)} \exp\left(\frac{-|x|}{\beta_l}\right)^{\alpha}, & \text{if } x \le 0 \\ \frac{\gamma}{2(\beta_l+\beta_r)\Gamma(1/\gamma)} \exp\left(\frac{-|x|}{\beta_r}\right)^{\alpha}, & \text{if } x > 0 \end{cases} \tag{8}$$

Finally, the statistical mean of each AGGD fit is computed as:

$$\eta = (\beta_r - \beta_l)\frac{\Gamma(2/\gamma)}{\Gamma(1/\gamma)} \tag{9}$$

The parameters of AGGD $(\gamma, \beta_l, \beta_r, \eta)$ and GGD (α, β) are estimated using moment–matching based approaches proposed by Sharifi et al. [37] and Lasmar et al. [38], respectively. The same approaches were also adopted by Mittal et al. [6].

Using, estimates of $(\gamma, \beta_l, \beta_r, \eta)$ along the four directions and (α, β) from the GGD fit to the histogram of $\{\hat{I}(i,j)\}$, $18D$ features are computed at two scales leading to a $36D$ *spatial–NSS* feature vector.

Fast Spatial–NSS: Since QualHOG is intended to be used in a scanning window approach, we first implemented a fast algorithm using integral images to allow efficient spatial–NSS feature computation within rectangular windows in an image. By using this *Fast Spatial–NSS* implementation, it is only necessary to first compute integral images at each scale in an image pyramid. Computation of spatial–NSS features for any rectangular window is near–instantaneous thereafter.

2) HOG: The HOG descriptor was first introduced by Dalal et al. [20]. It is a widely used feature descriptor for various object detection tasks [39]. To compute the HOG features, a detection window is divided into dense overlapping blocks of size 16×16 with a stride of 8×8 pixels. Each block is further divided into 2×2 cells and a histogram of gradients in 9 orientations is computed within each cell. All the histograms within a patch are concatenated to form the HOG feature descriptor.

This feature descriptor quantifies the gradient structure within a block which characterizes local edge directions. The appearance of an object in a detection window can be largely captured by the edge directions within indexed blocks. The local intensities are initially contrast normalized (before computation of the gradients) to provide illumination invariance. Thus, a discriminative classifier trained on histograms of oriented gradients extracted from dense set of local blocks in a detection window is capable of generalizing to other objects.

QualHOG: The quality aware *QualHOG* descriptor is obtained by simply concatenating the HOG and spatial–NSS features. In our experiments, the detection windows are of size 80×64, which gives a HOG feature vector of length 2268, which combined with the $36D$ spatial–NSS features yields a 2304 dimensional QualHOG feature vector. The motivation behind appending perceptually relevant quality–aware features to conventional object detection features is that the optimal decision boundary in the HOG vector space varies non–trivially as a function of input image/video quality. By appending spatial–NSS features to the HOG feature vector and passing this to a linear SVM, we effectively model a quality dependent boundary shift in the space spanned by the HOG features.

QualHOG Based Face Detector: Linear support vector machines (SVMs) [40] were trained using QualHOG features from face and non–face patches. Specifically, we use a soft–margin SVM with a slack penalty that simultaneously maximizes the margin while minimizing the training error. SVMs with non–linear kernels were also tried in the initial experiments, however, they require much longer computational time and did not provide significant improvements in the results.

Soft–margin SVM is trained using a set of n annotated samples, $\{(X_i, y_i) : i = 1, 2, \ldots, n\}$, where X_i are the discriminating features of the training samples, and y_i are the class label, $+1$ for face and -1 for non–face samples. Training a linear SVM involves solving the following optimization problem:

$$\begin{aligned} \min_{W,b,\{\xi_i\}} \quad & \frac{1}{2}\|W\|_2^2 + \lambda \sum_{i=1}^{n} \xi_i \\ \text{such that} \quad & y_i(\langle W, X_i\rangle + b) \ge 1 - \xi_i \quad \forall i \end{aligned} \tag{10}$$

where, λ controls the penalty for slack variables $\{\xi_i\}$.

For robust face detection, we train linear SVM with the QualHOG features, i.e. $X_i = [X_i^{\text{HOG}},\ X_i^{\text{NSS}}]$. The baselines are trained using only $\{X_i^{\text{HOG}}\}$. In the pre–processing step, the features are scaled so that they take values in the range of $[-1, 1]$.

C. Biased–QualHOG Face Detector

In the above formulation of SVM, the linear predictor for a sample with feature X is $\hat{y} = \text{sign}(W^T X + b)$, where $W = [W^{\text{HOG}},\ W^{\text{NSS}}]$. Further, the weights, w_j corresponding to each feature, x_j are regularized equally. With such an uniform regularization, the weights corresponding the 36 dimensional spatial–NSS features W^{HOG}, could be unfairly penalized in comparison to weights of the 2268 dimensional HOG feature W^{NSS}. This might potentially undermine the importance of quality–aware spatial–NSS features. To overcome this

we propose the following biased SVM formulation of the QualHOG based face detector.

$$\begin{aligned} \min_{W,b,\{\xi_i\}} \quad & \frac{1}{2}\|W^{\text{HOG}}\|_2^2 + \frac{1}{2C_s^2}\|W^{\text{NSS}}\|_2^2 + \lambda\sum_{i=1}^{n}\xi_i \\ \text{such that} \quad & y_i(\langle W, X_i\rangle + b) \geq 1 - \xi_i \quad \forall i \end{aligned} \tag{11}$$

In training the above model, C_s and λ are set using cross–validation.

IV. Experiments

As discussed in Section III, the distortions considered in this paper (except for additive white noise) are global and hence we cannot use existing face databases that have only face and non-face patches. Instead we require full images which are first distorted by known distortion types and levels and then segmented into face and non–face patches. Thus, as a first step, we created a new Distorted Face Database (DFD) of facial images from images available freely on the internet. To keep the task simple, we chose images with mainly frontal faces. A total of 215 images were crawled, each with one or more frontal faces. These images were manually ensured to be of high quality with no visible distortions. This set of 215 images was divided into 150 training images and 65 test images. The faces in these images were manually annotated. A total of 1231 faces were present in the training set of images and 393 were present in the test set.

For simplicity, we demonstrate our model at a single scale and hence we designed a system that detects faces in patches of size 80×64. In order to obtain training and testing face samples of the required dimensions, we resized the images so that the average sizes of the faces within an image are 80×64. Also, in the image selection process, care was taken to ensure that in case of multi–face images, the sizes of the faces were not widely different. In this way, on the resized images, a 80×64 sized bounding box centered at the faces captures the facial content accurately.

Next, the images were modified in various ways to create distortions. The following distortion types were introduced at different levels on the training and test datasets.

- AWGN: The *imnoise()* function in MATLAB was used to introduce additive white gaussian noise to the images. 10 levels of AWGN were added with the noise variance parameters varying over a log scale, $\sigma_N^2 = \{4.5 \times 10^{-5}, 0.0001, 0.0003, 0.0009, 0.0025, 0.0065, 0.02, 0.05, 0.15, 0.36\}$.
- GBlur: The *imfilter()* function in MATLAB was used to introduce gaussian blur at 10 levels. The standard deviation of the gaussian filter was varied over a log scale, $\sigma_B = \{0.4, 1.0, 2.3, 3.6, 4.5, 6.0, 7.4, 12.0, 20.0, 32.0\}$.
- JPEG: The *imwrite()* function in MATLAB was used to produce JPEG compressed images at 10 levels of distortion. The Q factor controlling the quality of the image was also varied on a log scale, $Q = \{90, 60, 40, 25, 15, 10, 7.5, 5.0, 3.0, 2.0\}$.

A. Training the Face Detector

From each of the above sets of training images, the 1231 manually annotated faces were cut out to provide positive samples for each dataset. A random subset set of 1500 negative patches were initially selected from the non–face parts of the images in each training dataset.

Soft–margin linear SVM (10) and its biased variant (11), were trained using QualHOG features extracted on the positive and negative samples from different combinations of the training datasets described above. As baselines, analogous classifiers were trained using just the HOG features. Hereafter, we use the following terminology. A face detector trained on QualHOG and HOG features of samples from pristine images alone will be called *QualHOG–Prist* and *HOG–Prist* respectively. Similarly, face detectors trained on QualHOG and HOG features of samples from pristine images and images from $L1$ to Ln levels of distortion of distortion type D, are denoted as *QualHOG–D–L1-n* and *HOG–D–L1-n* respectively (for example *QualHOG–AWGN–L1-4* refers to the face detector trained on QualHOG features of training samples from pristine images and images distorted with AWGN of variances $\{4.5 \times 10^{-5}, 0.0001, 0.0003, 0.0009\}$). Finally, analogous biased linear SVM variants (learned using (11)) of QualHOG based face detectors are referred using the notation *Biased–QualHOG–D–L1-n*.

To train the face detectors based on QualHOG and HOG features, implementation of soft–margin linear SVM from LIBLINEAR [41] was used in the experiments. For each classifier, a preliminary detector was first trained using a small sub–sample of non–face patches of the training images (1500 negatives). The remainder of non–face regions of the training images were searched exhaustively for false positives (from the predictions of the preliminary detector), also referred to as "hard negatives". A maximum of 1000 hard negatives were obtained for each training dataset. The classifiers were then retrained using the augmented set of negative samples (the initial 1500 negative samples + hard negatives). This retraining process is adapted from the work by Dalal et al. [20], where the authors observed a significant improvement in the performance of each detector. Finally, for each type of face detector (QualHOG and HOG), the parameter λ for the soft–margin SVM was chosen via cross–validation by doing a grid search on the log scale.

In Biased–QualHOG face detector, we also observe that the optimization problem in (11) can be re–written as follows by substituting $\widehat{W}^{\text{NSS}} = \frac{W^{\text{NSS}}}{C_s}$:

$$\begin{aligned} &\min_{W^{\text{HOG}},\widehat{W}^{\text{NSS}},b,\{\xi_i\}} \frac{1}{2}\|W^{\text{HOG}}\|_2^2 + \frac{1}{2}\|\widehat{W}^{\text{NSS}}\|_2^2 + \lambda\sum_{i=1}^{n}\xi_i \\ &\text{s.t. } y_i\big(\langle W^{\text{HOG}}, X_i^{\text{HOG}}\rangle + \langle \widehat{W}^{\text{NSS}}, C_s X_i^{\text{NSS}}\rangle + b\big) \geq 1 - \xi_i, \ \forall i \end{aligned}$$

The above optimization problem can be solved via conventional SVM learners using the scaled set of features, $\widehat{X}_i = [X_i^{\text{HOG}}, C_s X_i^{\text{NSS}}]$. In this setting, the parameters λ and C_s are independently selected via cross–validation.

B. Testing

As mentioned earlier, the 393 faces annotated on each of the test datasets were cut out to obtain positive test samples and an exhaustive set of ∼17500 negative samples were

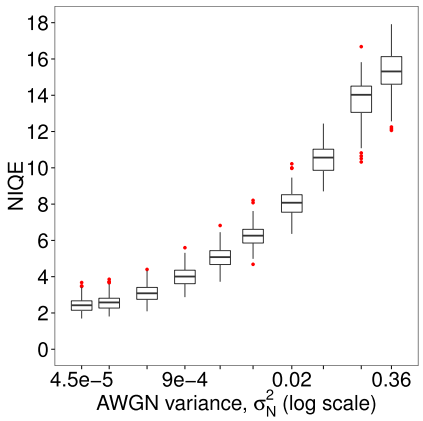


Fig. 1. NIQE vs AWGN.

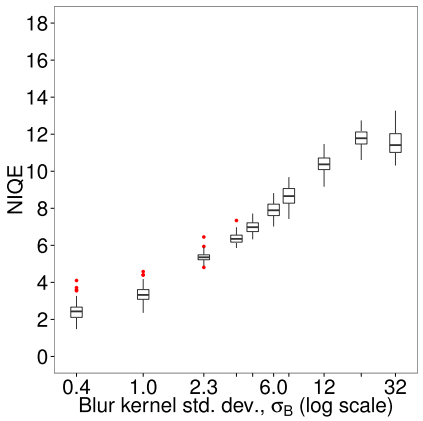


Fig. 2. NIQE vs gaussian Blur.

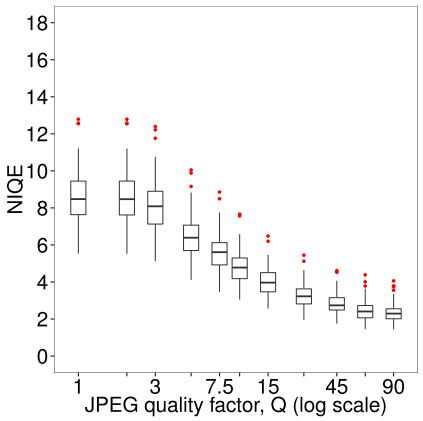


Fig. 3. NIQE vs JPEG.

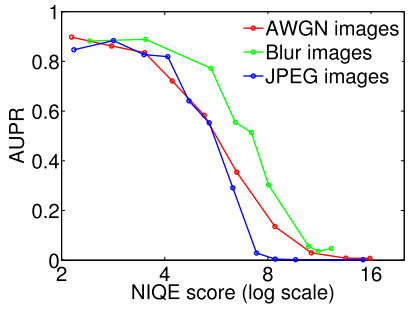


Fig. 4. Performance degradation of *HOG–Prist* with perceived quality measured as NIQE (high NIQE⇒low quality).

extracted from the non–face parts of the test images from the corresponding datasets. The *area under precision recall curve (AUPR)* was used as the evaluation metric since the test dataset is highly skewed as compared to the training dataset. Precision is defined as the fraction of detected positives that are faces, i.e., the ratio of true positives to the detected positives. Recall is defined as the fraction of actual positives that is detected, i.e., the ratio of true positives to the total number of positives. Typically, the continuous output of a classifier is thresholded to determine the discrimination boundary. Precision–recall curves for a system plot the trade–off between precision (y axis) and recall (x axis) as the discrimination threshold is varied.

V. Results

In practical settings, precise information regarding the distortion types and distortion levels afflicting an image are difficult to estimate. The NIQE image quality index, described in Section II, on the other hand, is a high performance distortion agnostic algorithm that does not rely on any form of distortion models. Further, the spatial–NSS features used to compute NIQE scores are computationally inexpensive as compared to other NR quality scores [4], [5]. We therefore use NIQE scores as surrogates for perceptual distortion levels. However, as a sanity check, we first assessed the NIQE scores of images against all of the distortion types considered. The NIQE scores of images distorted by various levels of AWGN, gaussian Blur and JPEG distortions are shown in Figs. 1–3, respectively. As expected, a strong positive correlation between degree of distortion and NIQE scores is observed for the common distortion types considered. Of course near–monotonicity against distortion severity is a minimum expectation of a perceptual image quality model.

A. Degradation of Face Detector Performance With Quality

We next studied the performance degradation of the baseline HOG based face detector, *HOG–Prist* on distorted images that are quality assessed by NIQE. In order to evaluate the performance of face detectors against NIQE scores, we first binned the images from the test datasets which were distorted by multiple degrees of each distortion type, into 10 discrete NIQE levels, then evaluated the performance of the baseline *HOG–Prist* face detector in each bin.

Fig. 4 plots the performance degradation of the *HOG–Prist* against NIQE for the three distortion types considered. Please note that binning process to create datasets of given NIQE score introduces inaccuracies in the evaluation. For example, the first bin in Fig. 4 with an average NIQE score of 2.1 has patches from images with NIQE score in the range 1.4 to 2.8. Thus, the absolute AUPR values reported in the experiments using these NIQE datasets (Figs. 4 and 5–7) are potentially inaccurate and are meant to only show the relative gain of QualHOG based face detector as compared to HOG–Prist.

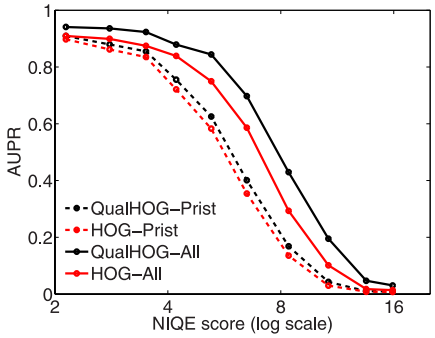


Fig. 5. Performance on images with AWGN distortion.

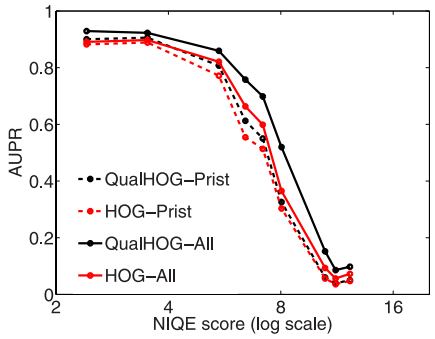


Fig. 6. Performance on images with Gaussian blur distortion.

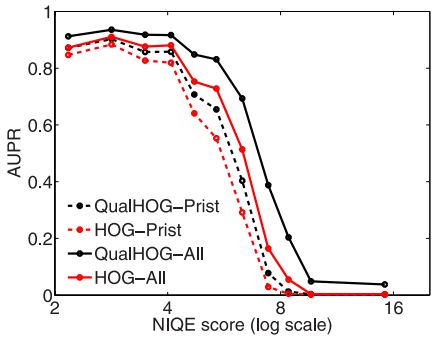


Fig. 7. Performance on images with JPEG compression.

It is not surprising that the degradation of face detection performance with increasing NIQE score (decreasing quality) is largely monotonic. It is, however, interesting to note that a region of steep decline of face detection performance exists for images with NIQE scores in the range 5–8, in which minor enhancements to image quality can yield significant improvement in face detection performance. This trade–off between image quality and face detection performance could be immediately exploited in the design of optimum communication channel parameters in facial image systems. Moreover, it is also interesting to note that for a given level of predicted image quality, *HOG–Prist* is more tolerant of quality degradation due to gaussian blur than those due to other distortions.

B. Distortion–Unaware Face Detectors

As knowledge of the distortion types present in a system is often unavailable, we trained four distortion–unaware face detectors: *QualHOG–Prist* and *HOG–Prist* which were trained using QualHOG and HOG features, respectively, of samples from only pristine images; and *QualHOG–All* and *HOG–All* which were analogous detectors trained using training samples from various levels of all three distortion types. We evaluate the performance of face detectors against NIQE scores for each distortion type. We use the test datasets with 10 discrete NIQE levels, which was curated for the study in Section V-A, for evaluating the performance of the distortion–unaware face detectors at different NIQE levels. The performance of these distortion–independent face detectors on test images in each bin are plotted in Fig. 5–7.

It can be seen that QualHOG based face detectors show significant improvement over the HOG based ones. Training on distorted images improves the performance of both HOG and QualHOG based face detectors. The HOG based face detector is constrained to a single detection boundary in the HOG vector space to capture the discriminating characteristics across all distorted images. However, using the quality–aware spatial–NSS features, QualHOG face detectors are capable of modeling a quality dependent boundary shift in HOG feature space. Thus, as hypothesized, the improvement from training on distorted samples is significantly higher for QualHOG compared to HOG based face detectors.

C. QualHOG vs HOG

For the performance analysis on individual distortion types, we trained distortion–dependent QualHOG and HOG based face detectors using samples with increasing levels of the distortions, *QualHOG–[D]–L1*, *QualHOG–[D]–L1-2*, …, *QualHOG–[D]–L1-10* and *HOG–[D]–L1*, …, *HOG–[D]–L1-10*, respectively, where, *[D]* is a placeholder for distortion type, AWGN, GBlur, and JPEG (refer Section IV for notation).

For each distortion type, AWGN, GBlur, and JPEG, test datasets analogous to the training datasets mentioned in Section IV were created at each distortion level ($L1$–$L10$) using the held out images. The distortion–dependent face detectors were evaluated on test datasets from appropriate distortion type. To avoid clutter we report the results of only the best performing detector for each distortion type, along with the distortion–independent detectors, *QualHOG–Prist*, and *HOG–Prist*. The best performing face detectors were separately chosen for the HOG and QualHOG based detectors. These results are compared in Figs. 8–10, for AWGN, GBlur, and JPEG distortions respectively. The distortions levels are represented on a horizontal log–scale.

QualHOG based face detectors again show uniformly better robustness as compared to the baselines. When the face detectors are trained on samples form only the pristine images (in *HOG–Prist* and *QualHOG–Prist*) the improvement is marginal. This is because there is minimal information regarding the distorted faces that the spatial–NSS features used in QualHOG could deliver a benefit from. Training on samples from distorted images in general improves the tolerance to distortions. However, QualHOG face detectors are better equipped to learn quality dependent discriminating boundary in the HOG feature space, as compared to learning

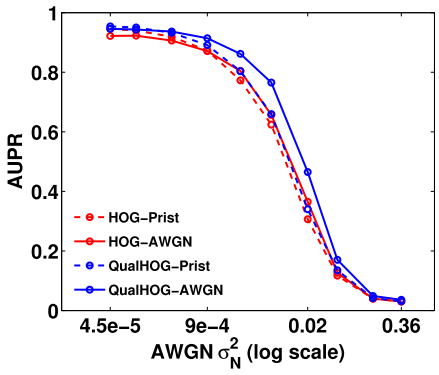


Fig. 8. AUPR vs AWGN σ_N^2.

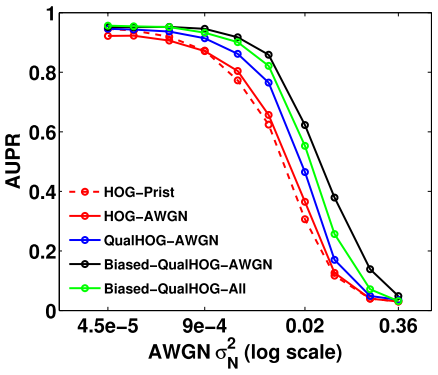


Fig. 11. AUPR vs AWGN σ_N^2.

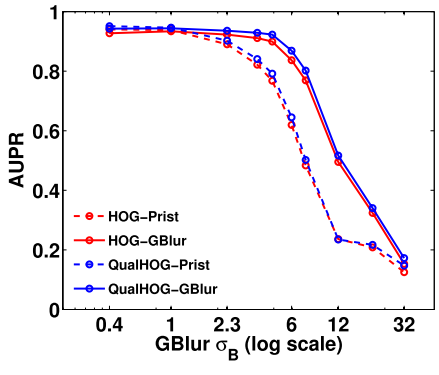


Fig. 9. AUPR vs Gaussian blur kernel σ_B.

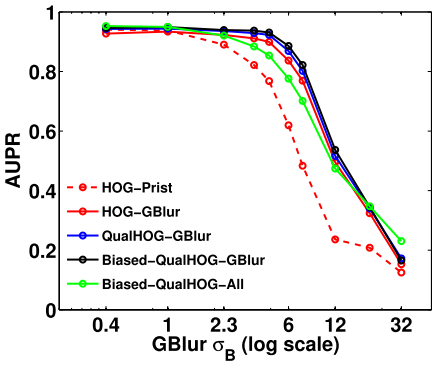


Fig. 12. AUPR vs Gaussian blur kernel σ_B.

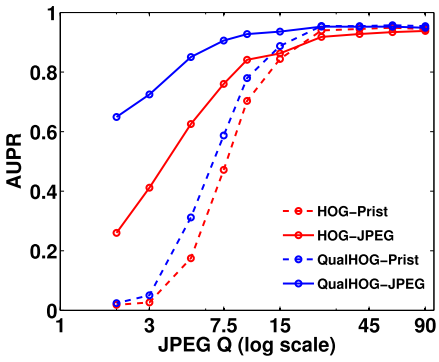


Fig. 10. AUPR vs JPEG Q.

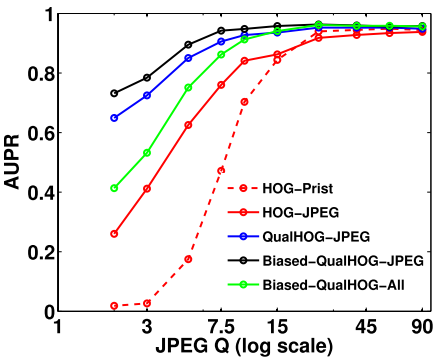


Fig. 13. AUPR vs JPEG Q.

from only the HOG features which are not known to capture quality aspects of the images. This is indeed observed for distortions arising from AWGN and JPEG compression. For these distortions types, the corresponding QualHOG face detector show significant improvement in tolerance to quality degradation as compared to the HOG face detector. However, for distortions arising from gaussian blur, the improvement is marginal.

D. Biased–QualHOG, QualHOG, and HOG

The initial results in Section V-C validate our hypothesis that quality–aware image features can aid in building distortion–robust face detectors. However, as discussed in Section III, in the current SVM formulation, the quality–aware spatial–NSS features of QualHOG are possibly penalized unfairly owing to the smaller number of features compared to HOG. To overcome this, we used the biased SVM formulation in (11). For each QualHOG face detectors described in Section V-C, analogous biased face detectors were trained. We again report results of only the best performing biased and unbiased detectors for each distortion type, along with the distortion–independent detectors, *HOG–Prist*, and *Biased–QualHOG–All*. Note that *Biased–QualHOG–All* (refer Section V-B) is a unified distortion–independent model trained on QualHOG features of samples from all three distortion types. To avoid clutter we did not plot the results for *QualHOG–Prist* and *HOG–All*. These results are compared in Figs. 11–13 for AWGN, GBlur, and JPEG distortions, respectively.

a: AWGN tolerance, HOG–AWGN(left) and Biased–QualHOG–AWGN (right)

b: JPEG tolerance, HOG–JPEG (left) and Biased–QualHOG–JPEG (right)

c: Gaussian blur tolerance, HOG–GBlur (left) and Biased–QualHOG–GBlur (right)

Fig. 14. Qualitative comparison of tolerance of the face detectors. In this illustration, for each distortion type, we show samples of images distorted to the level at which the AUPR of the baseline HOG face detector (left) and the proposed Biased–QualHOG face detector (right) fall below 0.8.

It can be observed that the proposed modification to the traditional SVM formulation significantly improves upon both the unbiased QualHOG based detectors as well as the baseline HOG based detectors, with the exception of Biased–QualHOG–GBlur. For distortions arising from AWGN and JPEG compression, the Biased–QualHOG variants outperform the baseline HOG based detectors by a large margin. To get a qualitative sense of the comparison, we illustrate the improved robustness in Fig. 14. The figure illustrates samples of faces that are distorted to the level at which the performance of the face detectors falls below a reasonably good threshold of AUPR ≥ 0.8. It is clear that the proposed detectors are visibly more tolerant to quality degradation from AWGN and JPEG compression. The resulting improvement is most remarkable for distortions from JPEG compression. For gaussian blur, the improvement is only marginal, and moreover, it can be seen from the illustration that both the baseline HOG–GBlur and the proposed QualHOG–GBlur are highly robust to distortions from gaussian blur compared to other distortion types.

TABLE I
NIQE SCORES UPTO WHICH THE FACE DETECTORS HAVE PERFORMANCE, $AUPR \geq 0.8$

Distortion D	HOG–D	Biased–QualHOG–D
AWGN	3.4	5.1
GBlur	5.5	6.6
JPEG	4.3	6.7

Finally, we note that for AWGN and JPEG, even the distortion–unaware QualHOG detector, *Biased–QualHOG–All*, outperforms the distortion–aware baselines of *HOG–AWGN* and *HOG–JPEG*, respectively. This provides further evidence for the claim that spatial–NSS features capture a distortion–agnostic measure of quality. However, gaussian blur is again an exception. It is possible that the HOG features are inherently robust to distortions from blur. This would also explain the observation in Fig. 4, that the baseline detector is more tolerant to quality degradation from blur as compared to quality degradation from other distortions considered.

E. Performance Measured Against NIQE Scores

To complete the analysis we also evaluate the face detectors against an effective distortion–agnostic image quality measure, NIQE. We created test datasets at various levels of perceptual quality by binning the NIQE scores of the test images at various distortion levels into 10 bins.

The results follow a trend similar to that observed against ground truth distortion levels, and thus the plots omitted to avoid redundancy. For a reasonable required level of performance AUPR ≥ 0.8, the tolerance of Biased–QualHOG face detectors as compared to the baseline are tabulated in Table I. The results corroborate our conclusions so far.

F. Performance on Natural Images

To study the performance of proposed face detectors on natural images encountered in real–life, we evaluate the face detectors on a subset of images in a face annotated database "FDDB: Face Detection Data Set and Benchmark" which consists of annotated faces images collected from news photographs [42]. We choose a single fold of the database consisting of 290 images. These images were pre–processed to discard non–frontal faces and faces with large occlusions as detecting such faces is outside the scope of this paper. A total of 405 face patches and a comprehensive set of $\sim 26K$ non–face patches were extracted. Distortion agnostic face detectors were evaluated on this dataset and the resulting Precision–Recall curves are shown in Fig. 15.

Here again we observe that the QualHOG based face detectors perform better than their HOG based counterparts. The Biased–QualHOG–All detector, however, provides only a marginal improvement. A possible explanation for this behavior could be that the images in the database were only mildly distorted (based on a visual examination). The Biased–QualHOG–All detector, on the other hand, was trained to operate under harsher distortions often arising in transmission and storage. We propose to consider more extensive experimentation by curating a dataset of images with various degrees of natural distortions from real–life applications as a part of future work.

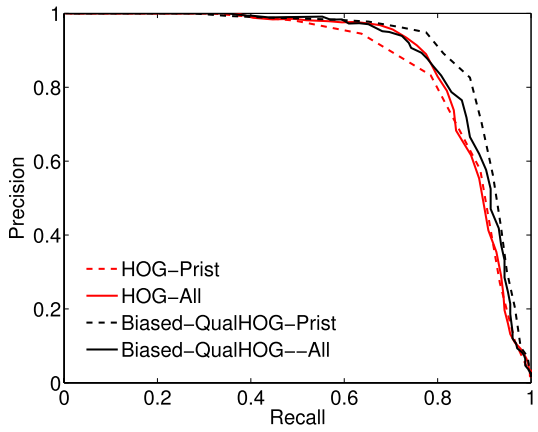


Fig. 15. Precision–Recall curves for the performance of distortion agnostic face detectors on a subset of FDDB benchmark data set.

G. Computation

In training and testing the face detectors, computation involved depends primarily on two tasks: (a) computation of the features (Spatial–NSS and HOG), and (b) learning the SVM for classification. In comparison to the $2268D$ HOG features, computing and learning from 36 additional Spatial–NSS features does not cause significant overhead in computation time.

VI. CONCLUSIONS

In this paper we first established that the easily computable NR image quality score, NIQE is effective as a proxy for actual distortion levels when evaluating the trade–off between face detection performance against image impairments arising from three common distortion types, AWGN, gaussian blur, and JPEG. The performance of generic HOG–based face detectors was found to degrade rapidly for NIQE scores greater than 4. It was also observed that for NIQE scores in the 5–8 range, a modest improvement in perceived image quality measures drastically improves face detection performance. This region can be fruitfully targeted when allocating resources in constrained settings. Another interesting observation was that, face detector performances are consistently more tolerant of quality impairments due to gaussian blur than those due to other distortions considered.

Secondly, we showed that QualHOG features, which combine face indicative HOG features with quality–aware spatial NSS features are more effective at learning a face detector that is robust to common and important image distortions. The QualHOG based face detectors show significant improvement over their HOG based analogues when trained on distorted images. In a practical distortion–unaware setting, the QualHOG–All face detector typically produced reliable results ($AUPR \geq 0.8$) for test datasets with NIQE scores of up to 6.5, while HOG–All provided equivalent performance on images with NIQE score up to 5.

Initial comparison of the proposed QualHOG and the baseline HOG face detectors in both distortion–aware and distortion–agnostic settings validate our hypothesis that quality–aware image features can aid in building distortion–robust face detectors. The biased variants the QualHOG face detectors further improve the robustness of the proposed face detectors. For distortions arising for AWGN and JPEG compression, the Biased–QualHOG face detectors show visibly higher tolerance to quality impairments. However, for distortions arising from gaussian blur, the improvement is marginal.

Interestingly, for AWGN and JPEG, in spite of being distortion–independent, Biased-QualHOG-All also provides better performance compared to distortion–aware HOG–AWGN and HOG–JPEG models when tested on individual distortion types. Thus, the QualHOG based face detectors are able to achieve acceptable face detection performance at much higher levels of visual impairments than what is currently possible.

Going forward, we anticipate the development of quality–aware face recognition models, where quality–predictive features in combination with anthropometric facial features [43] could yield recognition engines with significantly improved distortion resilience.

Further, in real–life applications, the distortions observed are sometimes more complex than the primitive distortion types considered in this paper. Going forward, we are planning extensive experimentation where we will create datasets of facial images affected by (a) multiple distortions, and (b) authentic (non–synthetic) distortions drawn from real–life photographic facial imaging applications. Given such a resource, we will conduct extensive studies on the efficacy of QualHOG features for face detection on images impaired by complex mixtures of distortions. While such a study is far beyond the scope of the work reported here, we are greatly motivated by the results we have obtained.

ACKNOWLEDGEMENT

The authors would like to thank the reviewers for useful comments and suggestions.

REFERENCES

[1] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for image artifacts based on human visual sensitivity," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 1994, pp. V/117–V/120.

[2] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. Conf. Rec. 37th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2003, pp. 1398–1402.

[3] P. Ye and D. Doermann, "No-reference image quality assessment using visual codebooks," *IEEE Trans. Image Process.*, vol. 21, no. 7, pp. 3129–3138, Jul. 2012.

[4] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: From natural scene statistics to perceptual quality," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3350–3364, Dec. 2011.

[5] M. A. Saad, A. C. Bovik, and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3339–3352, Aug. 2012.

[6] A. Mittal, A. K. Moorthy, and A. C. Bovik, "No-reference image quality assessment in the spatial domain," *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012.

[7] P. Campisi, M. Carli, G. Giunta, and A. Neri, "Blind quality assessment system for multimedia communications using tracing watermarking," *IEEE Trans. Signal Process.*, vol. 51, no. 4, pp. 996–1002, Apr. 2003.

[8] Q. Li and Z. Wang, "Reduced-reference image quality assessment using divisive normalization-based image representation," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 2, pp. 202–211, Apr. 2009.

[9] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a 'completely blind' image quality analyzer," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 209–212, Mar. 2013.

[10] M. Abdel-Mottaleb and M. H. Mahoor, "Application notes—Algorithms for assessing the quality of facial images," *IEEE Comput. Intell. Mag.*, vol. 2, no. 2, pp. 10–17, May 2007.

[11] R.-L. V. Hsu, J. Shah, and B. Martin, "Quality assessment of facial images," in *Proc. Biometrics Symp., Special Session Res. Biometric Consortium Conf.*, Sep./Aug. 2006, pp. 1–6.

[12] Y. Chen, S. C. Dass, and A. K. Jain, "Localized iris image quality using 2-D wavelets," in *Proc. Int. Conf. Adv. Biometrics*, 2006, pp. 373–381.

[13] N. D. Kalka, V. Dorairaj, Y. N. Shah, N. A. Schmid, and B. Cukic, "Image quality assessment for iris biometric," *Proc. SPIE Conf. Biometric Technol. Human Identificat. III*, vol. 6202, pp. 61020D-1–62020D-11, 2006.

[14] *Information Technology—Biometric Data Interchange Formats—Part 5: Face Image Data*, document ISO/IEC 19794-5, 2005.

[15] C. Zhang and Z. Zhang, "A survey of recent advances in face detection," Microsoft Research, Redmond, WA, USA, Tech. Rep. MSR-TR-2010-66, 2010.

[16] E. Hjelmas and B. K. Low, "Face detection: A survey," *Comput. Vis. Image Understand.*, vol. 83, no. 3, pp. 236–274, 2001.

[17] D. M. Rouse, R. Pépion, S. S. Hemami, and P. Le Callet, "Image utility assessment and a relationship with image quality assessment," *Proc. SPIE*, vol. 7240, p. 724010, Feb. 2009.

[18] D. M. Rouse and S. S. Hemami, "Quantifying the use of structure in cognitive tasks," *Proc. SPIE*, vol. 6492, p. 64921O, Feb. 2007.

[19] D. M. Rouse and S. S. Hemami, "Analyzing the role of visual structure in the recognition of natural image content with multi-scale SSIM," *Proc. SPIE*, vol. 6806, p. 680615, Feb. 2008.

[20] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.

[21] R. V. Babu, S. Suresh, and A. Perkis, "No-reference JPEG-image quality assessment using GAP-RBF," *Signal Process.*, vol. 87, no. 6, pp. 1493–1503, Jun. 2007.

[22] X. Zhu and P. Milanfar, "A no-reference sharpness metric sensitive to blur and noise," in *Proc. Int. Workshop QoMEx*, Jul. 2009, pp. 64–69.

[23] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[24] R. Lienhart, E. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Proc. 25th DAGM Symp.*, 2003, pp. 297–304.

[25] S. C. Brubaker, J. Wu, J. Sun, M. D. Mullin, and J. M. Rehg, "On the design of cascades of boosted ensembles for face detection," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 65–86, 2008.

[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. CVPR*, Dec. 2001, pp. I-511–I-518.

[27] R. Lienhart and J. Maydt, "An extended set of Haar-like features for rapid object detection," in *Proc. Int. Conf. Image Process.*, 2002, pp. I-900–I-903.

[28] A. Albiol, D. Monzo, A. Martin, J. Sastre, and A. Albiol, "Face recognition using HOG–EBGM," *Pattern Recognit. Lett.*, vol. 29, no. 10, pp. 1537–1543, 2008.

[29] A. Gala and S. Shah, "Joint modeling of algorithm behavior and image quality for algorithm performance prediction," in *Proc. Brit. Mach. Vis. Conf.*, 2010, pp. 31.1–31.11.

[30] J. E. Dowling, *The Retina: An Approachable Part of the Brain*. Cambridge, MA, USA: Harvard Univ. Press, 1987.

[31] D. J. Field, "Relations between the statistics of natural images and the response properties of cortical cells," *J. Opt. Soc. Amer. A*, vol. 4, no. 12, pp. 2379–2394, 1987.

[32] D. L. Ruderman, "The statistics of natural images," *Netw., Comput. Neural Syst.*, vol. 5, no. 4, pp. 517–548, 1994.

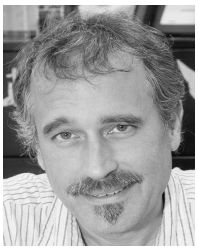
[33] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: A fresh view of inhibition in the retina," *Proc. Roy. Soc. London B, Biol. Sci.*, vol. 216, no. 1205, pp. 427–459, 1982.

[34] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Vis. Neurosci.*, vol. 9, no. 2, pp. 181–197, 1992.
[35] Z. Wang and A. C. Bovik, "Reduced- and no-reference image quality assessment," *IEEE Signal Process. Mag.*, vol. 28, no. 6, pp. 29–40, Nov. 2011.
[36] A. C. Bovik, "Automatic prediction of perceptual image and video quality," *Proc. IEEE*, vol. 101, no. 9, pp. 2008–2024, Sep. 2013.
[37] K. Sharifi and A. Leon-Garcia, "Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 5, no. 1, pp. 52–56, Feb. 1995.
[38] N.-E. Lasmar, Y. Stitou, and Y. Berthoumieu, "Multiscale skewed heavy tailed model for texture analysis," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 2281–2284.
[39] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1627–1645, Sep. 2010.
[40] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
[41] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.
[42] V. Jain and E. Learned-Miller, "FDDB: A benchmark for face detection in unconstrained settings," Dept. Comput. Sci., Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. UM-CS-2010-009, 2010.
[43] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3D face recognition," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.

Suriya Gunasekar received the B.Tech. degree in electronics and communications engineering from the National Institute of Technology at Warangal, Warangal, India, in 2010, and the M.S. degree from the University of Texas at Austin, Austin, TX, USA, in 2012, where she is currently pursuing the Ph.D. degree. Her field of study is machine learning, and her focus includes modeling dyadic matrix data and matrix estimation, rank aggregation, interaction of perceived image quality and computer vision task, and distributed estimation.

Joydeep Ghosh is currently the Schlumberger Centennial Chair Professor of Electrical and Computer Engineering with the University of Texas at Austin (UT Austin), Austin, TX, USA, where he joined the faculty in 1988. He received the B.Tech. degree from UT Austin in 1983 and the Ph.D. degree from the University of Southern California, Los Angeles, CA, USA, in 1988. He is the Founder and Director of the Intelligent Data Exploration and Analysis Laboratory. He has taught graduate courses on data mining and Web analytics every year to UT Austin students and to the industry, for over a decade. He was voted as the Best Professor in the Software Engineering Executive Education Program at UT Austin. His research interests lie primarily in data mining and Web mining, predictive modeling/predictive analytics, and machine learning approaches such as adaptive multilearner systems, and their applications to a wide variety of complex real-world problems. He has authored over 400 refereed papers and 50 book chapters, and coedited over 20 books. His research has been supported by the NSF, Yahoo!, Google, ONR, ARO, AFOSR, Intel, IBM, and several others. He has received 14 Best Paper Awards over the years, including the Best Research Paper Award from UT Austin in 2005 and the Darlington Award from the IEEE Circuits and Systems Society for the overall Best Paper in CAS/CAD in 1992. He has been a Plenary/Keynote Speaker on several occasions, such as ICDM'13, (Health Informatics workshops at) KDD'14, ICML'13, ICHI'13, MICAI'12, KDIR'10, and ISIT'08, and has widely lectured on intelligent analysis of large-scale data. He served as the Conference Cochair or Program Cochair for several top data mining-oriented conferences, including SDM'13, SDM'12, KDD'11, CIDM'07, ICPR'08 (Pattern Recognition Track), and SDM'06. He was the Conference Cochair for Artificial Neural Networks in Engineering from 1993 to 1996 and 1999 to 2003, and the Founding Chair of the Data Mining Technical Committee of the IEEE Computational Intelligence Society. He has also co-organized workshops on health informatics, high-dimensional clustering, Web analytics, Web mining, and parallel/distributed knowledge discovery.

Dr. Ghosh has served as a cofounder, consultant, or advisor to successful startups (Accordion Health, Stadia Marketing, Neonyoyo, and Knowledge Discovery One), and a consultant to large corporations, such as IBM, Motorola, and Vinson & Elkins.



Alan C. Bovik is currently the E. J. Cockrell Endowed Chair in Engineering with the University of Texas at Austin (UT Austin), Austin, TX, USA, where he is also the Director of the Laboratory for Image and Video Engineering. He is a faculty member with the Department of Electrical and Computer Engineering and the Institute for Neuroscience. His research interests include image and video processing, computational vision, and visual perception. He has authored over 700 technical articles in these areas and holds several U.S. patents. His publications have been cited over 35 000 times in the literature and his current H-index is over 70. He is listed as a Highly Cited Researcher by Thompson Reuters. His several books include the companion volumes entitled *The Essential Guides to Image* and *Video Processing* (Academic Press, 2009). He was a recipient of a number of major awards from the IEEE Signal Processing Society, including the Society Award (2013), the Technical Achievement Award (2005), the Best Paper Award (2009), the *IEEE Signal Processing Magazine* Best Paper Award (2013), the Education Award (2007), the Meritorious Service Award (1998), and (coauthor) the Young Author Best Paper Award (2013). He was also a recipient of the Honorary Member Award of the Society for Imaging Science and Technology (2013), the Society of Photo-Optical and Instrumentation Engineers (SPIE) Technology Achievement Award (2012), the IS&T/SPIE Imaging Scientist of the Year Award (2011), the Hocott Award for Distinguished Engineering Research from the Cockrell School of Engineering at UT Austin (2008), and the Distinguished Alumni Award from the University of Illinois at Urbana-Champaign (2008). He is a fellow of the Optical Society of America and SPIE. He cofounded and was the longest serving Editor-in-Chief of the IEEE TRANSACTIONS ON IMAGE PROCESSING (1996–2002), and created and served as the General Chairman of the first IEEE International Conference on Image Processing in Austin (1994), along with numerous other professional society activities, including the Board of Governors, the IEEE Signal Processing Society (1996–1998), an Editorial Board of PROCEEDINGS OF THE IEEE (1998–2004), and has been a Series Editor of the *Image, Video, and Multimedia Processing* (Morgan and Claypool Publishers, 2003–present).

Dr. Bovik is a registered Professional Engineer in the State of Texas and is a frequent consultant to legal, industrial, and academic institutions.