Blind Image / Video Quality Assessment Based on DCT-Domain Statistics

Minu Thomas, Saranya Sasidharan[1], Smitha K S[2]

[1, 2]Federal Institute of Science and Technology (FISAT), Department of Computer Science

Abstract: ***As the need of quality assessment increases day by day. Here develops an efficient general-purpose blind image/video quality assessment (I/VQA) algorithm using a natural scene statistics (NSS) model of discrete cosine transform (DCT) coefficients. The image/video quality assessment approach relies on a simple Bayesian inference model to predict image/video quality scores from certain extracted features. The features are based on an NSS model of the image/video DCT coefficients. The estimated parameters of the model are utilized to form features that are indicative of perceptual quality. These features are used in a simple Bayesian inference approach to predict quality scores. The resulting algorithm requires minimal training and adopts a simple probabilistic model for score prediction. Given the extracted features from a test image, the quality score that maximizes the probability of the empirically determined inference model is chosen as the predicted quality score of that image. In case of video, frames are extracted from video and for each frame above process are repeated. When tested on the LIVE database, the system correlates highly with human judgments of quality both in case of image and video.***

Keywords: Discrete cosine transforms (DCT), generalized Gaussian density, natural scene statistics, video quality assessment

1. Introduction

The ubiquity of transmitted digital visual information in daily and professional life, and the broad range of applications that rely on it, such as personal digital assistants, high-definition televisions, internet video streaming, and video on demand, necessitate the means to evaluate the visual quality of this information. The various stages of the pipeline through which an image passes can introduce distortions to the image, beginning with its capture until its consumption by a viewer. The acquisition, digitization, compression, storage, transmission and display processes all introduce modifications to the original image. These modifications, also termed distortions or impairments, may or may not be perceptually visible to human viewers. If visible, they exhibit varying levels of annoyance. Quantifying perceptually annoying distortions is an important process for improving the quality of service in applications such as those listed above. Since human raters are generally unavailable or too expensive in these applications.

Likewise, today's technology permits video content to be ubiquitously created, stored, transmitted, and shared between users on a multitude of devices ranging from hand-held PDAs and tablets, to very large high definition screens. Video content is being transmitted in exponentially increasing volumes via wireless and wired networks. The limited availability of bandwidth and the physical properties of the transmission media and capture and display devices mean that some information from the original source is likely to be lost. It is, however, important that the perceived visual quality at the end-user be maintained at an acceptable level, given rising consumer expectations of the quality of multimedia content delivered to them.

Image and video quality assessment (I/VQA) researchers have been working to understand how distortions introduced throughout the loss path between the source and destination affect the statistics of multimedia signals and how these distortions affect perceived signal quality. The most accurate way to assess the quality of an image or a video is to collect the opinions of a large number of viewers of the image/video in the form of opinion scores that rate the visual quality of the image or video. These opinion scores are then averaged (usually after normalization with respect to each individuals score average). This average is known as the mean-opinion-score (MOS), and the overall process is referred to as subjective I/VQA. While subjective I/VQA is cumbersome, expensive, impractical and for many important applications infeasible (e.g. for real-time monitoring of video or image quality in a network), it is valuable for providing ground truth data for the evaluation of objective I/VQA algorithms.

RR-I/VQA refers to I/VQA models that require partial information about the reference signal in order to predict the quality of a test signal. NR-I/VQA models have potentially much broader applicability that FR and RR models since they can predict a quality score in the absence of a reference image/video or any specific information about it. The problem of blindly assessing the visual quality of images and videos requires dispensing with older ideas of quality such as fidelity, similarity, and metric comparison. Only recently have NR-IQA algorithms been devised that correlate highly with human judgments of quality. Some are distortion specific, i.e. they quantify one or more specific distortions such as blockiness, blur, or ringing and score the image or video accordingly. There are considerably fewer algorithms that work well across multiple classes of distortions.

There are even fewer blind VQA algorithms than blind IQA algorithms. The problem is much more challenging owing to a lack of relevant statistical and perceptual models. Certainly, accurate modeling of motion and temporal change statistics in natural videos would be valuable, since these attributes play an important role in the perception of videos. Indeed, considerable resources in the human visual system (HVS) are devoted to motion perception.

Licensed Under Creative Commons Attribution CC BY

Presently, NR-I/VQA algorithms generally follow one of three trends: 1) distortion- specific approaches. These employ a specific distortion model to drive an objective algorithm to predict a subjective quality score. These algorithms quantify one or more distortions such as blockiness, blur, or ringing and score the image accordingly; 2) training based approaches: these train a model to predict the image/video quality score based on a number of features extracted from the image/video; and 3) natural scene statistics (NSS) approaches: these rely on the hypothesis that images/videos of the natural world (i.e., distortion-free images/videos) occupy a small subspace of the space of all possible images/videos and seek to find a distance between the test image/video and the subspace of natural images/videos.

The rest of the paper is organized as follows. In section 2, related-works of Image/Video Quality Assessment methods are briefly described. Section 3 describes the framework of Blind Image/Video Quality Assessment Based on DCT-Domain Statistics. Experimental results of proposed method are described in section 4. And finally, section 5 summarizes the conclusion of this paper.

2. Literature Survey

Z. Wang [1] proposed a new philosophy in designing image and video quality metrics, which uses structural distortion as an estimate of perceived visual distortion. A computationally efficient approach is developed for full-reference (FR) video quality assessment. One of the most attractive features of the proposed method is perhaps its simplicity. Note that no complicated procedures (such as spatial and temporal filtering, linear transformations, object segmentation, texture classification, blur evaluation, and blockiness estimation) are involved.

Wang [2] propose an RR image quality assessment method based on a natural image statistic model in the wavelet transform domain. Here use the Kullback-Leibler distance between the marginal probability distributions of wavelet coefficients of the reference and distorted images as a measure of image distortion. A generalized Gaussian model is employed to summarize the marginal distribution of wavelet coefficients of the reference image, so that only a relatively small number of RR features are needed for the evaluation of image quality. The proposed method is easy to implement and computationally efficient. In addition, the method find that many well-known types of image distortions lead to significant changes in wavelet coefficient histograms, and thus are readily detectable by our measure. Limitation includes the statistical redundancies insensitive to small geometric distortions such as spatial translation, rotation and scaling.

Qiang [3] propose an RRIQA algorithm based on a divisive normalization image representation. Divisive normalization has been recognized as a successful approach to model the perceptual sensitivity of biological vision. It also provides a useful image representation that significantly improves statistical independence for natural images. By using a Gaussian scale mixture statistical model of image wavelet coefficients, compute a divisive normalization transformation (DNT) for images and evaluate the quality of a distorted image by comparing a set of reduced-reference statistical features extracted from DNT-domain representations of the reference and distorted images, respectively. This leads to a generic or general-purpose RRIQA method, in which no assumption is made about the types of distortions occurring in the image being evaluated. The statistical redundancies between wavelet coefficients can be reduced by using the divisive normalization transform.

R Soundararajan [4] studies the problem of automatic reduced-reference image quality assessment (QA) algorithms from the point of view of image information change. Such changes are measured between the reference- and natural-image approximations of the distorted image. Algorithms that measure differences between the entropies of wavelet coefficients of reference and distorted images, as perceived by humans, are designed. The algorithms differ in the data on which the entropy difference is calculated and on the amount of information from the reference that is required for quality computation, ranging from almost full information to almost no information from the reference. A special case of these is algorithms that require just a single number from the reference for QA. The algorithms are shown to correlate very well with subjective quality scores. Performance degradation, as the amount of information is reduced, is also studied.

M. Masry [5] propose a metric, is based on a model of the human visual system implemented using the wavelet transform and separable filters. The visual model is parameterized using a set of video frames and the associated quality scores. The visual models hierarchical structure as well as the limited impact of fine scale distortions on quality judgments of severely impaired video is exploited to build a framework for scaling the bitrates required to represent the reference signal. Two applications of the metric are also presented. In the first, the metric is used as the distortion measure in a rate-distortion optimized rate control algorithm forMPEG-2 video compression. The resulting compressed video sequences demonstrate significant improvements in visual quality over compressed sequences with allocations determined by the TM5 rate control algorithm operating with MPEG-2 at the same rate. In the second, the metric is used to estimate time series of objective quality scores for distorted video sequences using reference bitrates as low as 10 kbps. The resulting quality scores more accurately model subjective quality recordings than do those estimated using the Mean Squared Error as a distortion metric while requiring a fraction of the bitrates used to represent the reference signal. The reduced-reference metrics performance is comparable to that of the full-reference metrics tested in the first Video Quality Experts Group evaluation.

R. Soundararajan [6] adopted a hybrid approach of combining statistical models and perceptual principles to design QA algorithms. A Gaussian scale mixture model for the wavelet coefficients of frames and frame differences is used to measure the amount of spatial and temporal information differences between the reference and distorted videos, respectively. The spatial and temporal information

Licensed Under Creative Commons Attribution CC BY

differences are combined to obtain the spatio-temporal-reduced reference entropic differences. The algorithms are flexible in terms of the amount of side information required from the reference that can range between a single scalar per frame and the entire reference information. The spatio-temporal entropic differences are shown to correlate quite well with human judgments of quality.

Wang [7] propose a new approach that can blindly measure blocking artifacts in images without reference to the originals. The key idea is to model the blocky image as a non-blocky image interfered with a pure blocky signal. The task of the blocking effect measurement algorithm is then to detect and evaluate the power of the blocky signal. The proposed approach has the flexibility to integrate human visual system features such as the luminance and the texture masking effects. The main applications Encoder - optimize parameter selection and bit allocation Decoder - design post-processing algorithm. A modified version of the measurement system has also been developed, which combines human visual masking effects. The new measurement systems can be applied blindly, while most of the other image quality measures need the reference images. The new algorithms employ higher order statistics (HOS) features. It is a new application of HOS technique in the field of image processing. The new measurement systems can be applied blindly, while most of the other image quality measures need the reference images. The new algorithms employ higher order statistics (HOS) features. It is a new application of HOS technique in the field of image processing.

X. Zhu [8] propose a no-reference objective sharpness metric detecting both blur and noise is proposed in this paper. This metric is based on the local gradients of the image and does not require any edge detection. Its value drops either when the test image becomes blurred or corrupted by random noise. It can be thought of as an indicator of the signal to noise ratio of the image. Experiments using synthetic, natural, and compressed images are presented to demonstrate the effectiveness and robustness of this metric. Its statistical properties are also provided. Local gradients affect is taken and then decomposed to singular value and characterize its behavior in the presence of blur and noise. The sharpness metric is defined using the noise variance and is the fixed small positive constants.

A. K. Moorthy [9] proposes a new two-step framework for no-reference image quality assessment based on natural scene statistics (NSS). Once trained, the framework does not require any knowledge of the distorting process and the framework is modular in that it can be extended to any number of distortions. Here describe the framework for blind image quality assessment and a version of this framework the blind image quality index (BIQI) is evaluated on the LIVE image quality assessment database. In fact, the use of an algorithm that correlates well with human perception in one category will probably help improve the overall performance of BIQI. This is a unique advantage, since overall better performance may be obtained as we find or create NR algorithms that correlate better with human perception.

A. C. Bovik [10] proposed The Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) index that assesses the quality of a distorted image without need for a reference image. DIIVINE is based on a 2-stage framework involving distortion identification followed by distortion-specific quality assessment. DIIVINE is capable of assessing the quality of a distorted image across multiple distortion categories, as against most NR IQA algorithms that are distortion-specific in nature. DIIVINE is based on natural scene statistics which govern the behavior of natural images. The principles underlying DIIVINE, the statistical features extracted and their relevance to perception and thoroughly evaluate the algorithm on the popular LIVE IQA database. DIIVINE is statistically superior to the often used measure of peak signal-to-noise ratio (PSNR) and statistically equivalent to the popular structural similarity index (SSIM). DIIVINE does not compute specific distortion features (such as blocking), but instead extracts statistical features which lend themselves to a broad range of distortion measurements. Future work will involve increasing the subset of distortions beyond those considered here, in an effort to further relax any distortion dependence. Distortions in the image reduce the quality of the image and by using the no reference DIIVINE algorithm the quality of the image can be predicted across the multiple distortion categories. In this the oriented band pass responses are obtained by decomposing the distorted image in to wavelet transform and then extract the statistical features and stacked to form a vector and it describes the distortions in the image. The proportion of distortion is identified from the subset and then estimates the quality of the image. On linear SVM training for the classification and non linear SRV training for regression is used within each class.

T. Brandao [11] proposed a method, a DCT-based video quality prediction model (DVQPM) is proposed to blindly predict the quality of compressed natural videos. The model is frame-based and composed of three steps. First, each decoded frame of the video sequence is decomposed into six feature maps based on the DCT coefficients. Then efficient frame-level features (kurtosis, smoothness, sharpness, mean Jensen Shannon divergence, and blockiness) are extracted to quantify the distortion of natural scenes due to lossy compression. In the last step, each frame-level feature is averaged across all frames (temporal pooling); a trained multilayer neural network takes the features as inputs and outputs a single number as the predicted video quality. The DVQPM model was trained and tested on the H.264 videos in the LIVE Video Database. Results show that the objective assessment of the proposed model has a strong correlation with the subjective assessment. The method is limited to H.264 coded videos. In our future work, to extract more efficient features, we will further study the properties of natural scenes and the influence of various compressions on these properties. It can be extended to measure the blockiness in JPEG and JPEG 2000 compressed images or MPEG-2 compressed videos.

H. Boujut [12] proposed a No-Reference video quality assessment of broadcasted HD video over IP networks and DVB. This work has enhanced our bottom-up spatio-temporal saliency map model by considering semantics of the

Licensed Under Creative Commons Attribution CC BY

visual scene. Thus, propose a new saliency map model based on face detection that is called semantic saliency map. A new fusion method has been proposed to merge the bottom-up saliency maps with the semantic saliency map. Tests are performed on two H.264/AVC video databases for video quality assessment over lossy networks.

3. Proposed Method

The system works both for image quality assessment and video quality assessment. The method for doing image quality assessment is explained below. In case of VQA first frames are computed from the video, then for each frame method used in image quality assessment is done. In the final phase average of the quality scores is considered as the quality score of video.

3.1 Overview of the method

The framework of the proposed approach is summarized in Figure 1. An image entering the IQA pipeline is first subjected to local 2-D DCT coefficient computation. This stage of the pipeline consists of partitioning the image into equally sized n*n blocks, henceforth referred to as local image patches, then computing a local 2-D DCT on each of the blocks. The coefficient extraction is performed locally in the spatial domain in accordance with the HVSs property of local spatial visual processing (i.e., in accordance with the fact that the HVS processes the visual space locally). This DCT decomposition is accomplished across spatial scales. The second stage of the pipeline applies a generalized Gaussian density model to each block of DCT coefficients, as well as for specific partitions within each DCT block.

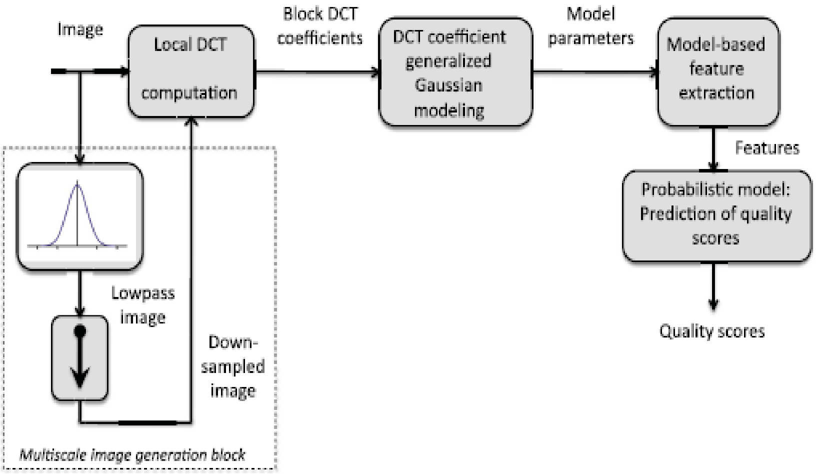
Image
Local DCT computation
Block DCT coefficients
DCT coefficient generalized Gaussian modeling
Model parameters
Model-based feature extraction
Features
Probabilistic model: Prediction of quality scores
Quality scores
Lowpass image
Down-sampled image
Multiscale image generation block

Figure 1: Overview of the BLIINDS-II framework

Next briefly describe the DCT block partitions that are used. In order to capture directional information from the local image patches, the DCT block is partitioned directionally into three oriented sub regions. A generalized Gaussian _t is obtained for each of the oriented DCT coefficient sub regions. The partition reflects three radial frequency sub bands in the DCT block. The upper, middle, and lower partitions correspond to the low-frequency, mid frequency, and high-frequency DCT sub bands, respectively.

The third step of the pipeline computes functions of the derived generalized Gaussian model parameters. These are the features used to predict image quality scores. In the following sections, define and analyze each model-based feature, demonstrate how it changes with visual quality, and examine how well it correlates with human subjective judgments of quality.

The fourth and final stage of the pipeline is a simple Bayesian model that predicts a quality score for the image. The Bayesian approach maximizes the probability that the image has a certain quality score given the model-based features extracted from the image. The posterior probability that the image has a certain quality score given the extracted features is modeled as a multidimensional GGD.

3.1.1 Generalized Probabilistic Model

The Laplacian model has often been used to approximate the distribution of DCT image coefficients. This model is characterized by a large concentration of values around zero and heavy tails. In the prior work , sample statistics were used(kurtosis, entropy, etc.,), without image modeling, to create a reasonably successful but preliminary blind IQA algorithm. Here in this work refined approach by modeling image features using a generalized Gaussian family of distributions which encompasses a wide range of observed behavior of distorted DCT coefficients. The generalized Gaussian model has recently been used as a feature in a NSS-based RR-IQA algorithm and in a simple two-stage NR-IQA algorithm. The univariate generalized Gaussian density is given by

$$f(x|\alpha,\beta,\gamma) = \alpha e^{-(\beta|x-\mu|)^{\gamma}} \quad (1)$$

Where μ is the mean, γ is the shape parameter, and α and β are the normalizing and scale parameters where σ is the standard deviation.

This family of distributions includes the Gaussian distribution (β= 2) and the Laplacian distribution (β= 1). As β to infinity, the distribution converges to a uniform distribution. Figure 2 shows the GGD at varying levels of the shape parameter.

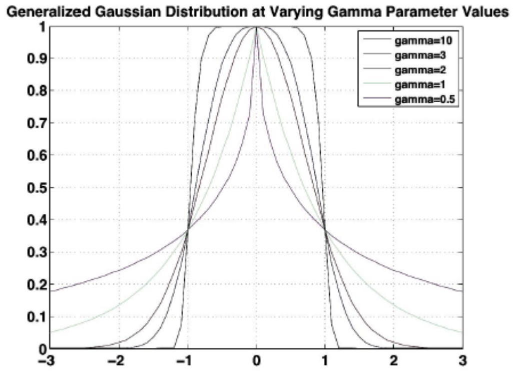

Figure 2: Generalized Gaussian density for varying levels of the shape parameter.

A variety of parameter estimation methods have been proposed for this model. The multivariate version of the generalized Gaussian density is given by

$$f(x|\alpha,\beta,\gamma) = \alpha\, e^{-(\beta(x-\mu)^T \Sigma^{-1}(x-\mu))^{\gamma}} \qquad (2)$$

Where Σ is the covariance matrix of the multivariate random variable x, and the remaining parameters are as defined in the univariate case. Here use (2) to form a probabilistic prediction model.

3.2 Model-Based DCT Domain NSS Features

The method proposes a parametric model to model the extracted local DCT coefficients. The parameters of the model are then utilized to extract features for perceptual quality score prediction. Extract a small number of model-based features. By using this system additionally explain the importance of multiscale feature extraction.

3.2.1 Generalized Gaussian Model Shape Parameter

Deploy a generalized Gaussian model of the non-DC DCT coefficients from n x n blocks. The DC coefficient neither increases nor decreases performance. The generalized Gaussian density in (1) is parameterized by mean μ, scale parameter β, and shape parameter γ. The shape parameter γ is a model-based feature that is computed over all blocks in the image.

The System demonstrates the distortion prediction efficacy of the shape feature γ on a large database of distorted images. The LIVE IQA Database consists of five subset datasets, each of which consists of images distorted by five types of representative realistic distortions [JPEG2000 compression, JPEG compression, white noise, Gaussian blur, and fast fading channel distortions (simulated by JPEG2000 distortion followed by bit errors)].Observe that the correlations are consistently higher when the lowest 10th percentile pooling strategy is adopted. This may be interpreted as further evidence that human sensitivity to image distortions is not a linear function of the distortion.

3.2.2 Coefficient of Frequency Variation

Let X be a random variable representing the histogrammed DCT coefficients. The next feature is the coefficient of frequency variation feature. If X has probability density function (1) and μ|X|=0, then

$$\mu|X| = \int_{-\infty}^{+\infty} |x|\, \alpha\, e^{-(\beta|x|)^{\gamma}} dx = \frac{2\alpha}{\beta^2 \gamma}\Gamma\left(\frac{2}{\gamma}\right) \qquad (3)$$

Substituting for α and β yields

$$\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} = \frac{\sigma^2}{\mu^2_{|x|}} \qquad (4)$$

Further

$$\sigma^2_{|X|} = \sigma^2_X - \mu^2_{|x|} \qquad (5)$$

So that

$$£ = \frac{\sigma|X|}{\mu|X|} = \sqrt{\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} - 1} \qquad (6)$$

The feature £ is computed for all blocks in the image. The feature is pooled by averaging over the highest 10th percentile and overall (100th percentile) of the local block scores across the image. The motivation behind the percentile pooling strategy is similar to that for pooling of the shape parameter feature γ. In the coefficient of frequency variation £, the denominator $\mu_{|X|}$ measures the center of the DCT coefficient magnitude distribution, while $\sigma_{|X|}$ measures the spread or energy of the DCT coefficient magnitudes. The ratio £ correlates well with visual impressions of quality. The high correlation between £ and subjective judgments of perceptual quality is an indication of the monotonicity between £ and subjective DMOS. Since £ is the ratio of the variance $\sigma_{|X|}$ to the mean $\mu_{|X|}$, the effect of an increase (or decrease) of $\sigma_{|X|}$ in the numerator is mediated by the decrease (or increase) of $\mu_{|X|}$ in the denominator of £. Indeed, two images may have similar perceptual quality even if their respective DCT coefficient magnitude energy is very different, depending on where the distribution of the coefficient magnitude energy is centered.

3.2.3 Orientation Model-Based Feature

Image distortions often modify local orientation energy in an unnatural manner. The HVS, which is highly sensitive to local orientation energy, is likely to respond to these changes. To capture directional information in the image that may correlate with changes in human subjective impressions of quality, this model the block DCT coefficients along three orientations. The three differently shaded areas represent the DCT coefficients along three orientation bands. A generalized Gaussian model is fitted to the coefficients within each shaded region in the block, and £ is obtained from the model histogram fits for each orientation. The variance of £ is computed along each of the three orientations. The variance of £ across the three orientations from all the blocks in the image is then pooled (highest 10th percentile and 100th percentile averages) to obtain two numbers per image.

Licensed Under Creative Commons Attribution CC BY

3.3 Prediction Model

A simple probabilistic predictive model is adequate for training the features used in BLIINDS-II. The prediction model is the only element of BLIINDS-II that carries over from BLIINDS-I. The efficacy of this simple predictor demonstrates the effectiveness of the NSS-based features used by BLIINDS-II to predict image quality. Let $X_i = [x_1; x_2;\ldots X_m]$ be the vector of features extracted from the image, where i is the index of the image being assessed, and m be the number of pooled features that are extracted. Additionally, let DMOSi be the subjective DMOS associated with the image i. Here model's the distribution of the pair is (X_i,$DMOS_i$).

The probabilistic model is trained on a subset of the LIVE IQA database, which includes DMOS scores, to determine the parameters of the probabilistic model by distribution fitting. The multivariate GGD model in (5) is used to model the data. The probabilistic model P(X, DMOS) is applied by fitting (2) to the empirical data of the training set. Specifically, once the quantity $(x-\mu)^T \Sigma^{-1} (x-\mu)$ is estimated from the sample data, parameter estimation of the GGD model in (5) is performed using the fast method. The distribution fitting (P(X, DMOS)) on the training data is only a fast intermediate step toward DMOS prediction. The end goal is not to fit the sample data of the training set as accurately as possible to the prediction model. The training and test sets are completely content-independent, in the sense that no two images of the same scene are present in both sets. The probabilistic model is then used to perform prediction by maximizing the quantity P ($DMOS_{i|}X_i$). This is equivalent to maximizing the joint distribution P(X, DMOS) of X and DMOS since P(X, DMOS) =P (DMOS|X)p(X).

4. Experimental Results

4.1 IQA Results

BLIINDS-II was rigorously tested on the LIVE IQA database which contains 29 reference images, each impaired by many levels of five distortion types: JPEG2000, JPEG, white noise, Gaussian blur, and fast-fading channel distortions (simulated by JPEG2000 compression followed by channel bit errors.). The total number of distorted images (excluding the 29 reference images) is 779.

The DCT computation was applied to 5*5 blocks with a 2-pixel overlap between the blocks. Multiple train test sequences were run. In each, the image database was subdivided into distinct training and test sets (completely content separate). In each train test sequence, 80% of the LIVE IQA database content was chosen for training, and the remaining 20% for testing. Specifically, each training set contained images derived from 23 reference images, while each test set contained the images derived from the remaining 6 reference images.

The project report quality score prediction results for features extracted at one scale only (8 features), over two scales (16 features, 8 features per scale), and over three scales (24 features, 8 per scale). Linear correlation coefficient (LCC) scores (on a logistic fitted function of the predicted DMOS using BLIINDS-II and subjective DMOS scores) as well as SROCC scores between the predicted DMOS scores and the subjective DMOS scores of the LIVE IQA database are computed for each of the 1000 iterations. The comparison of prediction results for 1 scale, 2 scale, and 3 scale feature extraction is shown in Tables 1 and 2. We found that no significant gain in performance was obtained beyond the third scale of feature extraction.

Table 1: median SROCC correlations for 1000 iterations of randomly chosen train and test sets). Comparison for multiple scales of feature extraction

LIVE subset	*One scale*	*Two scales*	*Three scales*
JPEG2000	0.9313	0.9533	0.9506
JPEG	0.9294	0.9403	0.9419
White noise	0.9753	0.9772	0.9783
GBlur	0.9417	0.9509	0.9435
Fast fading	0.88555	0.8657	0.8622
ALL	0.8973	0.8980	0.9202

Table 2: median LCC correlations for 1000 iterations of randomly chosen train and test sets. Comparison for multiple scales of feature extraction

LIVE subset	*One scale*	*Two scales*	*Three scales*
JPEG2000	0.9550	0.9571	0.9630
JPEG	0.9664	0.9781	0.9793
White noise	0.9804	0.9833	0.9854
GBlur	0.9300	0.9450	0.9481
Fast fading	0.8500	0.8701	0.8636
ALL	0.8919	0.9091	0.9232

The approach is not heavily dependent on the training set, here performed the following analysis. The results are shown in Figure 3. Notice that an SROCC of 0.85 is obtained when using only 30% of the content for training, and that the knee of the curve occurs at roughly 20%. This shows that our reported results are not tainted by overtraining or over fitting to the training data.

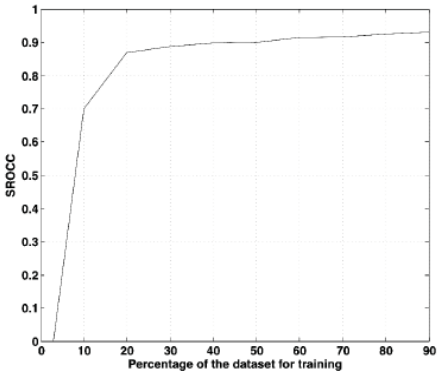

Figure 3: Plot of median SROCC between predicted and subjective DMOS scores (on all distortions)

4.2 VQA Results

The algorithm was evaluated on the publicly available LIVE VQA database [72]. The LIVE VQA database has a total of 160 videos derived from 10 reference videos of highly diverse spatial and temporal content. The patch size for the

DCT computation that was used is 5*5. This is similar to the feature extraction block size chosen in BLIINDS-2.

4.2.1 Algorithm Prediction Performance

There are no existing blind VQA approaches that are non-distortion specific, which makes it difficult to compare our algorithm against other methods. Full-reference and reduced reference approaches have the enormous advantage of access to the reference video or information about it. Blind algorithms generally require that the algorithm be trained on a portion of the database. However, compare against the naturalness index NIQE in, which is a blind IQA approach applied on a frame-by-frame basis to the video, and also against top performing full-reference and reduced reference algorithms.

Table 3: no-reference median SROCC and LCC correlations on train/test set splits

	SROCC		LCC	
Distortion	NIQE	Video BLINDS	NIQE	Video BLINDS
MPEG-2	0.523	0.869	0.490	0.924
H.264	0.541	0.839	0.579	0.893
Wireless	0.280	0.815	0.387	0.951
IP	0.276	0.779	0.443	0.946
ALL	0.151	0.759	0.317	0.881

Video BLIINDS clearly outperforms the blind NIQE index and the full-reference PSNR and SSIM measures. Video BLIINDS does not quite attain the performance level of state-of-the-art full-reference VQA measures, (MOVIE and ST-MAD), but its performance is nearly as good and with much less computational cost. Of course, Video BLIINDS does not rely on any information from the pristine version of the video to make quality predictions. It does, however, rely on being trained a priori on a set of videos with associated human quality judgments.

5. Conclusion

Natural scene statistic model-based approach is described for the blind I/VQA problem. The new NR-I/VQA model uses a small number of computationally convenient DCT-domain features. The algorithm can be easily trained to achieve excellent predictive performance using a simple probabilistic prediction model. The method correlates highly with human visual judgments of quality both in case of image and video. Experiments are done on LIVE Database. The system is compared with other quality assessment algorithms and obtained good results. Quality assessment takes only less time. Thus efficient blind I/VQA System can be used instead of full-reference and reduced reference methods.

References

[1] Z. Wang, L. Lu, and A. C. Bovik, Video quality assessment based on structural distortion measurement, Signal Process., Image Commun., vol. 19, no. 2, pp. 121132, Feb.2004.

[2] Z. Wang and E. P. Simoncelli, Reduced-reference image quality assessment using a wavelet-domain natural image statistic model, Proc. SPIE, vol. 5666, pp. 149159, Jan. 2005.

[3] L. Qiang 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. 202211, Apr. 2009.

[4] R. Soundararajan and A. C. Bovik, RRED indices: Reduced reference entropic differencing for image quality assessment, IEEE Trans. Image Process., vol. 21, no. 2, pp.517526, Feb. 2012.

[5] M. Masry, S. S. Hemami, and Y. Sermadevi, A scalable wavelet-based video distortion metric and applications, IEEE Trans. Circ. Syst. Video Technol., vol. 16, no. 2, pp.260273, Feb. 2006.

[6] R. Soundararajan and A. C. Bovik, Video quality assessment by reduced reference spatio-temporal entropic differencing, IEEE Trans. Circuits Syst. Video Technol., vol.23, no. 4, pp. 684694, Apr. 2013.

[7] Z. Wang, A. C. Bovik, and B. L. Evans, Blind measurement of blocking artifacts in images, in Proc. IEEE Int. Conf. Image Process., vol. 3. Sep. 2000, pp. 981984.

[8] X. Zhu and P. Milanfar, A no-reference sharpness metric sensitive to blur and noise,in Proc. Int. Workshop Qual. Multimedia Exper., Jul. 2009, pp. 6469.

[9] A. K. Moorthy and A. C. Bovik, A two-step framework for constructing blind image quality indices, IEEE Signal Processing Letters, vol. 17,no. 2, pp. 587599, May 2010

[10] 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. 33503364, Dec. 2011.

[11] T. Brandao and M. P. Queluz, No-reference quality assessment of H.264/AVC encoded video, IEEE Trans. Circuits Syst. Video Technol., vol. 20, no. 11, pp. 14371447, Nov.2010.

[12] H. Boujut, J. Benois-Pineau, T. A. O. Hadar, and P. Bonnet, No-reference video quality assessment of H.264 video streams based on semantic saliency maps, Proc.SPIE, vol. 8293, pp. 82930T-182930T- 9, Jan. 2012.