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.

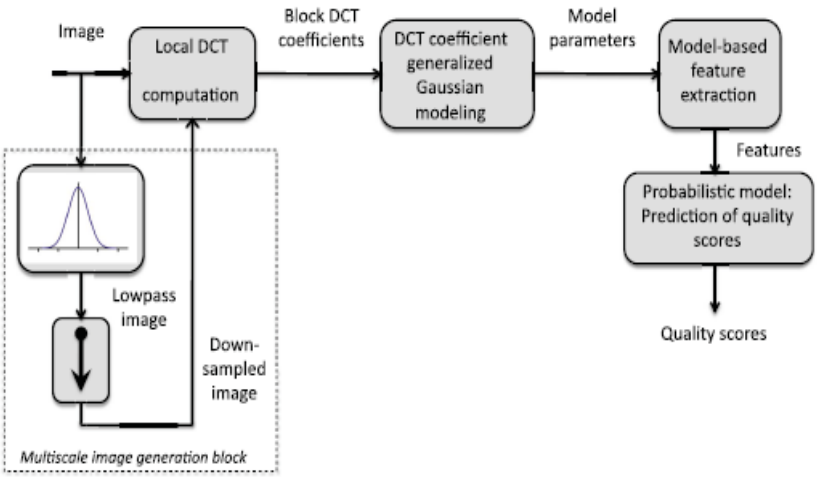

Figure 1: Overview of the BLIINDS-II framework

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.

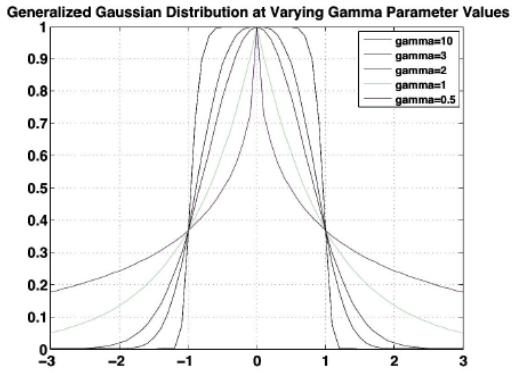

Figure 2: Generalized Gaussian density for varying levels of the shape parameter.

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

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

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

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

Substituting for α and β yields

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

Further

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

So that

$$£ = \frac{\sigma_{|X|}}{\mu_{|X|}} = \sqrt{\frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} - 1} \quad (6)$$

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.

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 $X_i = [x_1; x_2;....X_m]$ 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.

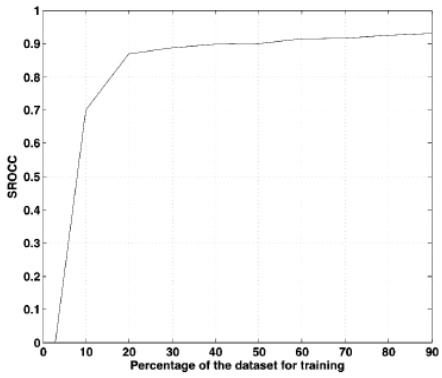

Figure 3: Plot of median SROCC between predicted and subjective DMOS scores (on all distortions)

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.