The thesis committee for Anish Mittal
certifies that this is the approved version of the following thesis.

Natural Scene Statistics-based Blind Image Quality Assessment in the Spatial Domain

APPROVED BY

SUPERVISING COMMITTEE:

Alan C. Bovik, Supervisor

Lawrence K. Cormack, Supervisor

Natural Scene Statistics-based Blind Image Quality Assessment in the Spatial Domain

by

Anish Mittal, B.Tech.

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER of SCIENCE in ENGINEERING

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2011

Dedicated to my mentors:Rajvanth, Vijay, Rita, Rico, Ivan and Anush

Acknowledgments

I want to thank Lord Narayan who engendered great opportunities and materialized himself now and then through superb mentors in my life who channelized me on right track.

I want to thank my parents Daleep and Sudha for their care and support throughout my life, instilling good values, providing me with a good education and encouragement to aim for doctorate degree from abroad even at the expense of their own discomfort. I also want to thank my siblings Esha and Parul for boosting me up when ever I felt down and depressed.

I wish to express deep gratitude towards my advisors Alan and Lawrence as they always instilled the belief of *You can do it. Believe us, it will work* in me. I was always fascinated by how much interest these guys take in their work and still getting charmed like kids when it came to new problem.

I acknowledge all LIVE members for their support and guidance by pointing out resources and clearing my doubts when ever I felt lost. I cherished all moments spent with LIVies and humourous talks with Anush, Joonsoo, Sina, Tony, Ming and Michelle. I want to thank Anush in specific who has been my Guru and role model. I have always resorted to him for all sort of doubts, discussions, questions and he invariably helped me as much as he could. As always he boosts my moral by *We will keep working on it. Don't worry too*

much about it, it will come. I count him as one of the FEW people who have been most influential in my life. Even though we argue over everything, his principles leave quite an impression.

O' Mesiah I treasure all you have given me, may I retort by serving your humans...

Natural Scene Statistics-based Blind Image Quality Assessment in the Spatial Domain

Anish Mittal, M.S.E.
The University of Texas at Austin, 2011

Supervisors: Alan C. Bovik
Lawrence K. Cormack

We propose a natural scene statistic (NSS)-based distortion-agnostic blind/no-reference (NR) image quality assessment (IQA) algorithm – the Referenceless Image Spatial QUality Evaluator (RISQUE) – which operates in the spatial domain. RISQUE does not compute distortion specific features such as ringing, blur or blocking, but instead uses scene statistics of normalized luminance coefficients to quantify the 'naturalness' (or lack thereof) in the image due to presence of distortions thereby leading to a holistic measure of quality.

We detail the algorithm and the statistical features extracted, and demonstrate how each of these features correlate with human perception. We perform a thorough evaluation of the RISQUE index in terms of its correlation with human perception and demonstrate that RISQUE is *statistically better than the full-reference peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM)* as well as *statistically superior to all present-day* *distortion-agnostic NR IQA algorithms.* We demonstrate that RISQUE features may be used for distortion-identification as well. Further, we also show that RISQUE is computationally efficient and its efficiency is superior to other distortion-agnostic approaches to NR IQA, thus making RISQUE an attractive option for practical application.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The mind must see visual achievement of the purpose before action is initiated.

Mack R. Douglas

Interest in the measurement of visual quality can be dated back to times when interest in quality assessment was primarily based on display applications [62]. But as time progressed and so did the prevalance of imaging in multifarious applications, 'Quality' got defined in different ways depending on application for which it was defined. Image acquistion engineers dealing with applications like laser range scanning focused on imaging system aspects when they gauged quality; printer engineers focused on tone, color assessment and fundamental printing attributes, such as area and line quality. In contrast, medical imaging researchers related it with the clarity with which they could detect malfunctions or diseases in body from captured images, for example tumours and cancers from X-Ray images. However, for the scope of our current work, we are interested mainly on digital multimedia applications targeted for entertainment applications.

Advancement in multimedia technologies have brought a host of devices to capture, compress, send and display different kinds of audiovisual stimulations. Great efforts have been devoted by developers working in 2D image and video transmission industry to guarantee end user a satisfactory quality of experience, being most salient for design and deployment of any multimedia service. Perceptual optimization of multimedia services looks promising in current era of bandwidth famine coupled with increased multimedia traffic so as to provide similar quality of service to consumer. In other words, objective function while finding optimum configuration of multimedia framework can incorporate Quality of Experience as an additional term.

A service network codes the produced audiovisual content to transmit it over communication channels to the consumer. Various distortions due to compression, channel noise, packet loss etc are introduced in the signal from this chain of operations from content development till transmission. These distortions in visual stimuli, when perceptible, mar the viewing experience and hence reduction in perceived visual quality. The reduced quality of distorted stimuli can be judged by conducting large-scale human *subjective* studies where human observers are asked to quantify quality of stimulus shown on a fixed scale. However, this kind of human assessment is time, effort and cost expensive; engendering the need to design algorithms capable of duplicating and hence eliminating human involvement altogether. These designed objective quality indices can then be utilized for multifarious applications including but not limited to optimum pre-filtering and bit assignment algorithm design at encoder side; optimal reconstruction, error concealment and post-filtering at decoder side and benchmarking of image and video processing systems.

Quality assessment(QA) algorithms are generally classified as (1) full-reference (FR), (2) reduced-reference (RR), and (3) no-reference (NR) algorithms.

FR QA algorithms assume that apart from distorted stimulus whose quality needs to be judged, a pristine reference stimulus is also available to the algorithm to be compared with.

RR QA algorithms assume that some *auxillary* information about the reference stimulus is available to the algorithm, even though the actual reference stimulus itself in unprocurable.

NR QA algorithms seek to find quality of distorted stimulus with no information about its pristine counterpart. Such an assessment of visual quality, also known with the name of blind image quality assessment, forms the crux of our proposed work.

In absence of any information from pristine image, it may seem that the problem of no reference image quality is insurmountable. Hence researchers first addressed the problem of full reference quality, where distorted images are compared with their pristine counterparts. Mean squared error (MSE) between distorted and pristine images being a standard criterion for signal fidelity determination gave a head start. It had nice properties of simplicity, parameter free estimation, inexpensive computation and memorylessness, it being evaluated at each sample independent of other samples. However, it had a strong assumption neglected by image processing community at first; i.e. no spatial relationships exists between samples of the signal [55]. In other words, if we randomly re-order pristine and distorted signals in the exact same way, mean squared error would remain same. This assumption fails critically for images and videos having a well defined structure which makes neighbouring pixels dependent on each other. Hence the most intuitive way was to reduce the dependencies between neighbouring pixels first in both pristine and distorted images before using MSE. Vision community indeed found that human visual system adopt a similar approach of reducing dependencies between neighbouring coefficients. Second order dependencies are reduced by filtering image using over complete representation of band pass scale and orientative selective Gabor filters in human visual cortex as each band pass image will have equal energy in every octave [35]. However, it is important to note that this filtering operation is linear, hence nonlinear dependencies between neighbouring filtered coefficients still remain. Ruderman sought that dividing the whitened luminance coefficients by local standard deviation reduces non Gaussian image statistics or non linearities present in filtered responses [37]. We can also pyschophysically reason out that linear processing doesnot suffice for modeling neurons responses in primary visual cortex . Even specifying mean firing rate requires represention of saturating nonlinearity. It also partially accounts for contrast masking [24] observed in human visual cortex making it an essential ingredient in quality assessment algorithms.

Following up on this ideology, we propose a no reference natural scene statistic based approach which models 'naturalness'(or lack there of) introduced in images due to presence of distortions. Our approach which we call as **Referenceless Image Spatial QUality Evaluator (RISQUE)**;extracts marginal statistics of locally normalized luminance signals and measures deviation in marginal statistics of given distorted image from natural scenes. Furthermore, we also model natural 'debiasing' of neighbourhood normalized luminance signals which also gets heavily demented in presence of distortions. Modeling of neighbourhood relationships in horizontal, vertical, off and on diagonals provides us with orientation distortion information. We use parameters from the best fit to these empirical distributions as the only features, no tuning is done by using any functional forms to improve the accuracy. This proves the saliency of our model with respect to human perception of quality.

Once we obtain features, we use a two stage framework like DIIVINE[30]. The computed features are first used to find the probability of existence of each distortion in the image. This is followed by evaluating the distortion specific image quality using trained regressor functions for each distortion. Once we have the vector of quality from each regressor, we compute dot product of distortion probability vector and distortion quality vector. The motivation behind this frame work as we will empirically show later is that each distortion induces 'un-naturalness' in features in a different way and different set of features are important for every distortion. Hence the frame work fits quite well with our selection of features. Another advantage of using this kind of regression approach is we can generalize it to as any many distortions as we want. We would also like to emphasize that we have taken a general set of features representing the 'naturalness'(or lack thereof) of images. They do not model any blur, ringing or blocking distortion, yet features are sensitive to diverse distortion shapes. Thus it is possible to train a classifier to equate deviation of parameters.

We prove through our performance evaluation that we are **statiscally better** than other proposed no reference algorithms and full reference indices PSNR and SSIM [57]. Hence our approach holds state of art in no reference image quality assessment with superior performance compared with full reference SSIM. This is by no means a small achievement given that we are not using any information from pristine images. The approach is real time and perception based, hence it can find applications in perceptual based optimization of multimedia networks. Although multiscale, the model is defined in space domain avoiding costly frequency or wavelet transforms.

The paper is organized as follows, we will first discuss previous work done in domain of image quality assessment and then motivate the use of our features to represent 'naturalness' of images and then go on to incite the two stage frame work used for regressing features to quality perception. We will next evaluate performance of our frame work and then conclude thesis with the scope of future work.

Chapter 2

Literature Review

The assumption that seeing is believing makes us susceptible to visual deceptions

Kathleen Hall Jamieson

The problem of no reference image quality seemed insurmountable given that no information is available about the pristine image. Hence researchers first investigated problem of full reference quality. We discuss full reference image quality algorithms proposed in literature. Three kinds of human perception based full reference quality assessment algorithms have been devised - human visual system based, structural similarity based and information theoratic. For human vision modeling based metrics, quality assessment problem is solved using engineering kind of approach where thresholding visibility is measured for signals and noise. These thresholds are then used to normalize the error between the reference and distorted images to get visual quality for an image. Either different stages of visual processing occuring in the HVS or different aspects of the visual system such as average brightness, contrast, spatial frequencies, orientations are modeled to get the visibility thresholds. Quality assessment models in this category include but not limited to Visual signal-to-noise ratio (VSNR) [11], visible differences predictor (VDP)[14], Sarnoff JND vision model [27], Teo and Heeger Model [50] and Modified PSNR based on HVS (PHVS) [15]. Second class of algorithms hypothesize on the fact that human brain is highly conformed to extract structural information from visual scenes and hence they quantify the structural distortion to find visual quality of images. To code for image structure, intuition is derived from optics of image formation perspective. Observed surface luminance is product of illumination and reflectance of the surface but structure of scene is independent of illumination of light falling on surface, hence the goal is to reduce this effect. This can be achieved to an extent by subtracting out local mean and normalizing by local contrast at every pixel in the image. This normalized luminance term in pristine image is compared point by point with normalized luminance term in the distorted image. Universal Quality Index (UQI) [54], Structural Similarity Index (SSIM) [57] and Multi-Scale Structural Similarity Index (MS-SSIM) [60] falls under this category of algorithms. The third school of thought base their approach on information theoratic perspective and hence model statistics of image signals. Their idea is that natural scenes form a tiny sub-space in space of all possible image signals. This can be easily conformed by running any random vector generator of length equal to number of image pixels in consideration and checking tally of natural scene generation from it. Therefore, human brain learnt these statistics of natural scenes with passage of time. Information Fidelity Criterion (IFC) [45], which comes under this category measures shared information between pristine and distorted images whereas Visual Information Fidelity (VIF) [43] quantifies two kinds of information separately: information present in pristine image and information shared between pristine and distorted image. A nice unifying analysis of three schools of thought was done by Seshadharinathan et al. in [42]. Reader is referred to cited reference for details.

Studying full reference quality assessment gave an insight about human perception and image quality in general which motivated researchers to probe further in to reduced reference and no reference quality assessment algorithm design. The aim behind reduced reference algorithms is typifying reference images with as reduced information as possible and still maintain a comparable visual quality estimate as full reference algorithms. More information from pristine image would naturally lead to more accurate estimates but then reduced reference loose valor to distinguish itself from full reference ones. Therfore, when reduced reference are compared across each other for their accuracy of visual quality prediction, both correlation with human scores and percentage of data used from reference should be considered. Again, three schools of thought have been followed here to get reduced information from pristine images. The first thought bases on modeling specific artifacts introduced for particular distortion kind and hence more suited for targeted applications. Authors in [63] [64] have modeled blocking, blurring or ringing artifacts whereas a set of spatial and temporal features suited to quantify video compression and communication environment distortions have been devised in [20] ,[22]. On the other hand, second category of algorithms base their approach on computational models of low level vision [8],[9]. They have an advantage of being generic enough to model any distortion to speak of but then we don't know of any work which applies it to images having generic distortions leaving JPEG and JPEG2000 compression. The third kind of algorithms assume that brain has evolved to learn the statistical regularities found in natural scenes [61]. When distortions are introduced in the image, the statistics of the image gets disturbed and 'un-naturalness' is introduced in the image. A generalized Gaussian density function has been used by authors in [61] for modeling marginal statistics of wavelet coefficients in every sub-band, and fitted model parameters are used as features from reference image. These few parameters can be sent easily as a header information in the channel on receiver side without any information loss. Marginal distributions of pristine images can then be reconstructed given these parameters and compared with distribution of sub-band coefficients from distorted images to quantify the presence of distortion. The framework performs quite well and needs no training. The limitation though is that only marginal distributions of wavelet coefficients have been considered for measuring visual quality but there exist strong dependencies between neighboring wavelet coefficients which reflect the nonlinear mechanisms used by the biological visual systems. These limitations have been addresssed by authors in [25] using divisive normalization which also accounts for contrast masking.

Soundarajan et al. bridged full reference and no reference algorithms by proposing a family of automatic 'reduced reference' algorithms ranging from almost full reference to almost no reference [48]. They differ from each other in amount of data based on which information change is predicted. The difference in entropies between wavelet coefficents of pristine and distorted images is representative of visual quality.

We now follow up towards no reference algorithms which do not use any information whatsoever from pristine images. Researchers commenced with assuming known distortion medium like compression, packet loss induced due to noisy channel etc. Most of the initial work targeted compression artifacts catering to bluriness, blockiness and ringing in images. However recently, remarkable work has been done in distortion agnostic algorithm design which makes no assumptions about distortion medium and model the un-naturalness for the image in general. Our proposed algorithm also falls in the later category.

No reference QA for JPEG distorted images measure quality through strength of edges at block boundaries along with spatial content dependent features to account for masking. A hermite based approach has been used by authors in [29] to model blurred edges whereas thresholding on computed gradients in [13]. The first order differences and spatial activity have been used to quantify perceived distortions in [58] while importance map weighting of spatial blocking scores in [2]. Frequency domain approach is used in [49] where block strengths are measured in fourier domain.

No reference QA for JPEG2000 distorted images quantify ringing artifacts to relate with human opinion of visual quality. Authors in [36], [28], [51] perform some edge detection techniques followed by quantification of edge spead to model ringing. In comparison, an information theoratic natural scene statistics based approach was adopted by researchers in [44] which models the dependencies between wavelets coefficients and its neighbours using threshold and offset approach.

No reference QA for blurred images adopt a similar line of thought as used for modeling JPEG2000 artifacts to quantify the edge spread or blurring artifacts. Edge strengths have been quantified in different ways, as block kurtosis of DCT coefficients in [10], iterative thresholding of gradient images in [52], probability computation of blur detection in [33] or modeling the just noticable blur in the blurred images [16]. Also, saliency weighted foveal pooling strategy was explored for quality assessment in [40] while gradient based approach along with singular value decomposition is used by authors in [65].

Distortion Agnostic Algorithms as the name suggests do not model any specific distortion characteristics but rather model 'naturalness' of images in general. However, some devised algorithms in this area only rely on a host of heuristic measures. For instance, Li came up with a set of three heuristics to charactertize the visual quality namely edge sharpness, random noise and structural noise [26] where edge sharpness was measured using edge detection, random noise level using local smoothness to model implusive noise and PDE-based model to capture artifacts introduced due to gaussian noise. To account for compression artifacts manifested as JPEG and JPEG2000, Li came up with a structural noise model. All the same, he did not analyse the performance of the proposed measure neither did he talk about the way to blend different heuristics to a single quality score.

All algorithms we have discussed till now only characterize nature of distortions. However, as each distortion manifests differently, it doesnot seem assuring to model generic behaviour holding true across distortions. Hence researchers started to explore if natural scenes obey some statistical regularities which get demented in presence of distortions. This approach was followed by authors in [18] where they modeled anisotropy in images using Renyi entropy along different orientations with the presumption that natural images have strong directional properties and introduction of distortions destroy it. Mean, standard deviation and range along four orientations were extracted as features to represent visual quality. An exhaustive evaluation of it's correlation with human opinion score lacks here too.

Moorthy et al.[30] proposed DIIVINE, a natural scene statistics based approach where 'un-naturalness' in an image is observed by quantifying deviation in marginal and joint sub-band wavelet statistics of natural scenes. Once features are computed, a 2 stage frame work where distortion identification followed by distortion specific quality assessment is done to compute quality score. Also, Saad et al. proposed BLINDS-II, again a generic no-reference image quality assessment algorithm based on statistics of discrete cosine transform coefficients. Image is first divided in to blocks and generalized gaussian distribution is fitted to each block of DCT coefficients. Model parameters retrieved from each block are pooled across space to retrieve features representative of visual quality. Probabilistic model based on multivariate generalized gaussian density is then applied for mapping computed features to image quality. DIIVINE is shown to perform **statistically better** than PSNR and comparable with full reference SSIM [57]. Both of the models have their associated pitfalls. We now address the limitations of the two discussed models and how our approach is able to overcome it.

As we model coefficients of the whole image together in comparison with block based approach of BLIINDS-II, we glean out three advantages. First, it makes the approach much faster with parameter estimation just done once in comparison with it being repeated for every block in BLIINDS-II. Also, the moment based method of finding parameters of generalized gaussian used by BLINDS-II assumes first and second order sample statistics being representative of population statistics but that's not really true if you consider as few coefficients as present in a block. Alternately, if whole image is considered then number of coefficients are large enough to correctly estimate population moments from sample moments. Thirdly, a pooling strategy needs to be devised if block based parameter estimation is used again based on heuristics.

Both DIIVINE and BLIINDS-II deliver top NR IQA performance (to date), but it is of interest to develop space domain only equivalents towards efficiency of performance computation due to reasons explained above.

Chapter 3

Natural Scene Statistics based Quality Assessment Model

Statistics are like bikinis. What they reveal is suggestive, but what they conceal is vital.

Aaron Levenstein

Satan delights equally in statistics and in quoting scripture....

H.G. Wells, The Undying Fire

The science of producing unreliable facts from reliable figures

Evin Esar

3.1 Natural Scene Statistics Model

Natural Images follow certain regularities and structure causing significant statistical redundancies. Vision scientists Atteneave and Barlow argued that reduction of significant redundancy is achieved in early vision through efficient coding hypothesis which decorrelates the input so that it matches the expected input of high levels of vision [1] [3]. The principle behind the argument was statistical independence needs to be achieved for probabilistic inference to be performed at higher levels of vision.

Vision scientists therefore started finding a linear decomposition that could model linear receptive fields but it required additional constraints like symmetry, spatial locality etc to find the unique principal components[34]. Hence research pitched towards finding higher-order statistical measurements such as independent component analysis (ICA) to uniquely constrain the choice of linear decomposition [7]. Physiological measurements also proved that the basis functions derived from ICA are localized in spatial position, orientation and scale are similar to cortical receptive fields [4]. The filtered coefficients using this basis were found to be decorrelated in second order, highly kurtotic, and more independent than principal components in general. Hence second order spatial correlation represented through autocorrelation or power spectrum [17] gets removed by filtering image using over complete representation of band pass filters as each band pass image will have equal energy in every octave. This can also be achieved by subtracting local mean luminance from every pixel luminance in the image as proposed by Ruderman [37]. But it is important to note that as this filtering operation is linear, the nonlinear dependencies between the neighbouring filter coefficients still remain.

Ruderman sought that dividing the whitened luminance coefficients by local standard deviation removes the non Gaussian image statistics or non linearities present in responses but it was more of a serendipity at that point of time [37]. He visualized this variance normalization as a means of increasing entropy. Authors in [53] then psychophysically reasoned out that this control mechanism explains a wide variety of the nonlinear behaviors of these neurons. Linear processes are not sufficient for modeling neurons responses in primary visual cortex as even specifying the mean firing rate requires represention of saturating nonlinearity. This statistically-derived divisive normalization reduces higher order statistical dependencies between adjacent responses, which further supports the efficient coding hypothesis. In this proposed work, we follow up the spatial domain approach taken by Ruderman involving the subtraction of local mean from luminance to whiten the luminance followed by normalization with local standard deviation to remove the non linear dependencies[37]. We will refer to these coefficients as mean subtracted contrast normalized coefficients from further on (MSCN).

The **R**erferenceless **I**mage **S**patial **Q**uality **E**valuator (RISQUE) that we have designed along these lines is described as follows.To compute MSCN coefficients, we first transform the color image into the perceptually uniform, color opponent CIE-Lab color space [21]. We just work in luminance domain, color is out of scope for the present work. We extract 7×7 patches around each pixel and compute local mean and variance by centering a circularly-symmetric Gaussian filter $w = \{w_{k,l} | k = -K/2, -2...K/2; l = -L/2, -2...L/2\}$ at that pixel. We sample the Gaussian out to 3 standard deviations and normalize the filter to unit sum. Changing the patch size from 5 to 9 did not change the signature of the coefficient distribution. We shall use the images shown in figure 3.1 to show the signature distributions and various kinds of distortion behaviours. Mean, whitened, standard deviation and MSCN version of the bikes image is shown in figure 3.2. Standard deviation image highlights the object boundaries and suppresses the object textures. In contrast, MSCN image looks much more uniform than the original image as small fluctuations have been expanded but large fluctations being attenuated. The image appears to be more like a noise left with a few residual object boundaries which can be thought of objects stripped from their textures. This falls in line with the image formation perspective given by Wang in [57] as we discussed in full reference quality algorithm review. The local mean reduces the effect of illumination from the local regions and division by variance acts as texture stripping operation from the object surfaces accounting for some sort of contrast masking. Figure 3.4 shows how correlations between neighbouring coefficents are reduced by this operation.

$$\forall i \in 1, 2...M \quad and \quad \forall j \in 1, 2...N$$

$$\mu(i,j) = \sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} w_{k,l} Y_{k,l}(i,j) \tag{3.1}$$

$$\sigma(i,j) = \sqrt{\sum_{k=-K/2}^{K/2} \sum_{l=-L/2}^{L/2} w_{k,l} (Y_{k,l}(i,j) - \mu(i,j))^2} \tag{3.2}$$

$$MSCN(i,j) = \frac{Y(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{3.3}$$

where M denotes the height and N denotes the width of image. Y denotes the luminance and K,L denotes the window size in vertical and horizontal direction respectively. For our analysis, K=7 and L = 7. C is a small constant added to

(a) Bikes

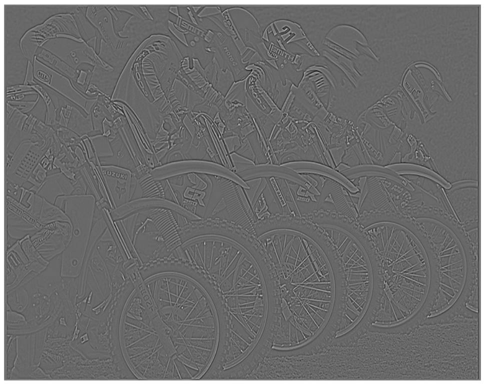
(b) Rapids

(c) House

(d) Building

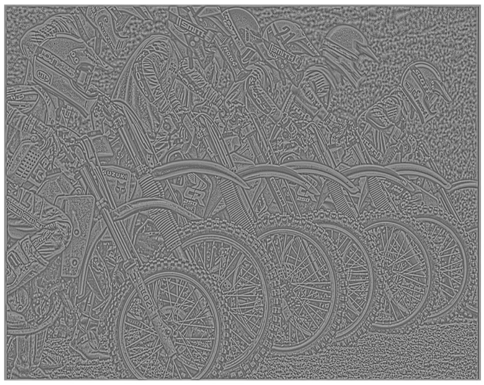
Figure 3.1: Several of images used in study from the LIVE database

(a) Local Mean Luminance



(b) Local mean subtracted from pixel luminance

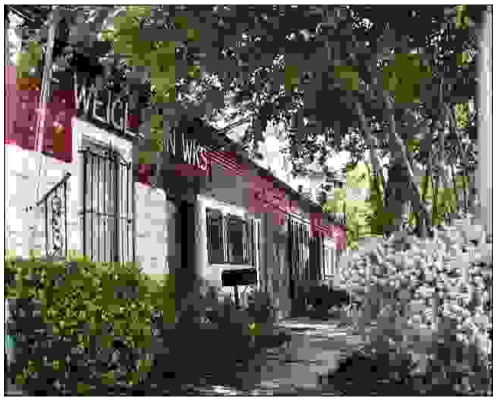
(c) Local Standard deviation



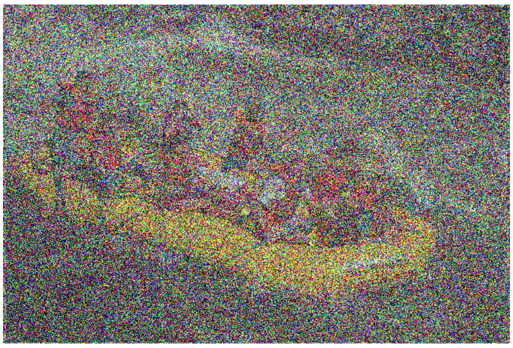
(d)MSCN

Figure 3.2: Mean subtraction and contrast normalization operation on the 'bikes' image

(a) JP2K



(b) JPEG



(c) WN

(d) Gblur

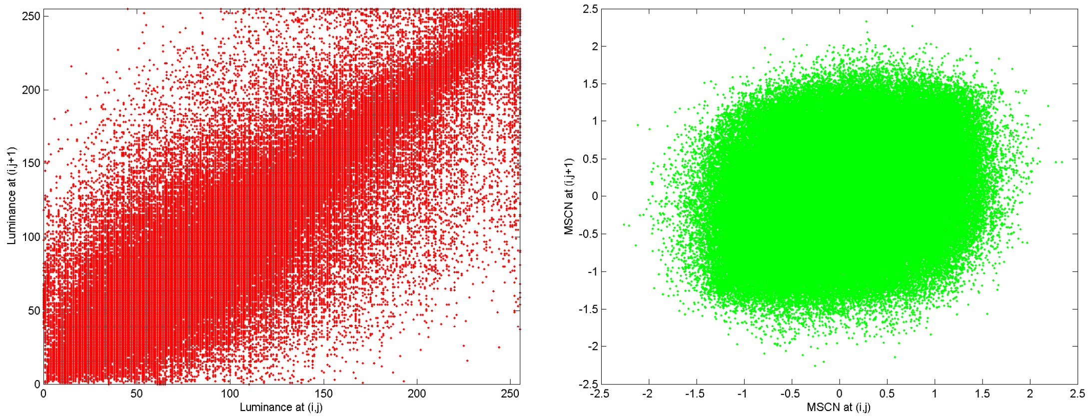
(e) FF

Figure 3.3: Several images depicting different kinds of distortions

prevent numerical instability when denominator of equation 3.3 goes to zero. This can happen in smooth regions of the image where standard deviation of the region is zero, for example: sky, smooth object surfaces etc.

Ruderman [37] found that MSCN coefficients follow gaussian like signatures for natural scenes but not exactly gaussian. Nonetheless, we are more interested in finding how distortions affect this pristine image signature and if unique characteristic signature exists for each kind of distortion which will aid the first stage of our framework. Figure 3.3 shows 5 kinds of different distortions which span the set of distortions considered in LIVE database [46] - JPEG2000 (JP2K) compression,JPEG compression, additive white noise (WN), Gaussian Blur (blur) and a Rayleigh fading channel labeled fast fading (FF). Authors in [32] indeed found that sub-band wavelet coefficients have a unique signature for each distortion. Deriving motivation from there, we check if the same is true for MSCN coefficients irrespective of content considered and we found that is indeed the case. Figure 3.5 shows MSCN empirical distribution for pristine and five distortions considered from database. We can observe that white noise increases the variance of the distibution but other distortions make it more peaky and kurtotic to different extents.

In order to capture orientation distortion information, we model directional relationships between neighbouring MSCN coefficients. In other words, we take the pairwise products of neighouring MSCN coefficients. Neighbours are chosen in 4 directions- $\{0°, 90°, 45°, -45°\}$ yielding horizontal pair wise products (HZ), vertical pair wise products(VC), on diagonal correlation (DC1)



(a) (b)

Figure 3.4: Comparison between (a) and (b) shows how correlation between neighbours get reduced with mean subtraction and variance normalization

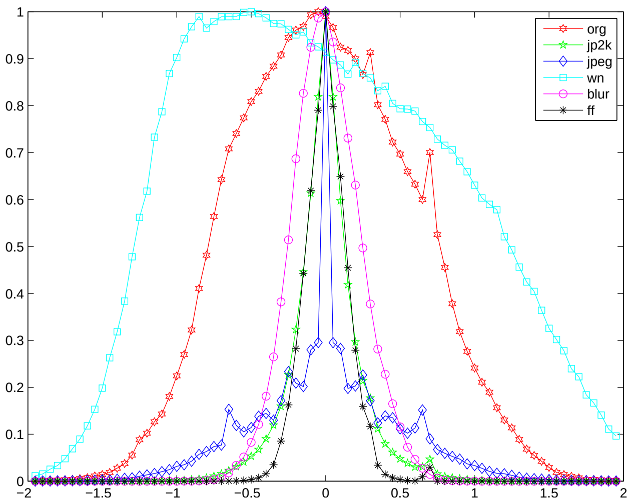


Figure 3.5: Signatures of MSCN coefficients for different distortions

and off diagonal correlation (DC2) respectively. Both negative as well as positive pair wise products are present, but as the coefficients are de-correlated, probability of positive and negative products should be about equal. Distortions cause de-biasing operations to not decorrelate the images in the same way. The empirical distributions get more kurtotic as local directional dependencies are modified. It is interesting to observe WN has more uniform like distribution though.

$\forall i \in 1, 2 ... M \quad and \quad \forall j \in 1, 2 ... N$

$$HC(i,j) = MSCN(i,j)MSCN((i+1)modM, j) \tag{3.4}$$

$$VC(i,j) = MSCN(i,j)MSCN(i,(j+1)modN) \tag{3.5}$$

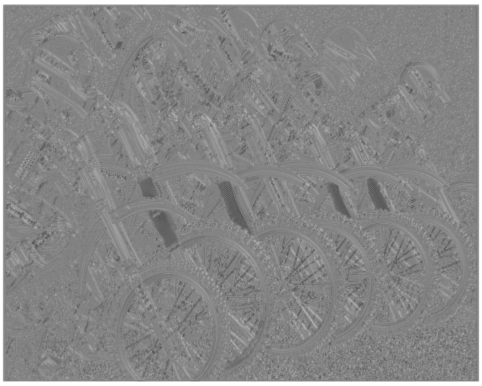
$$DC1(i,j) = MSCN(i,j)MSCN((i+1)modM,(j+1)modN) \tag{3.6}$$

$$DC2(i,j) = MSCN(i,j)MSCN((i+1)modM,(j-1)modN) \tag{3.7}$$

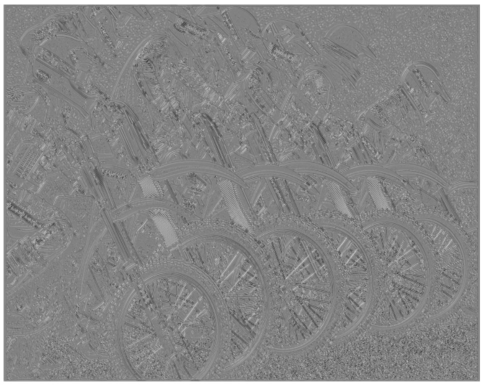
With the inference that distortion affects the empirical density of MSCN coefficients and directional pair wise products in a characteristic manner, we now proceed towards parameterizing it. The MSCN coefficients are distributed with gaussian like signatures which in the presence of distortions becomes highly kurtotic, hence generalized gaussian distribution seems a prudent choice for modeling. In contrast, the variance pair wise products have skewed kurtotic distributions, hence we need assymetric density functions to model that. *Asymmetric generalized gaussian density (AGGD)* functions cater well to our requirements being specific enough to model gaussian signatures as well as

(a) Horizontal Correlation

(b) Vertical Correlation



(c) Diagonal 1 Correlation



(d) Diagonal 2 Correlation

Figure 3.6: Directional pair wise products of the bikes image

symmetric generalized gaussian for different shape parameters but with equal left and right variance. But also, general enough to model skewed distributions. They have been widely been used to model skewed heavy tailed distributions in texture analysis [23].

Assymetric generalized gaussian pdf with zero mode is given by:

$$f_X(x;\alpha,\sigma_l^2,\sigma_r^2) = \begin{cases} \frac{\alpha}{(\beta_l+\beta_r)\gamma(\frac{1}{\alpha})} \exp(-(\frac{-x}{\beta_l})^\alpha) & x < 0 \\ \frac{\alpha}{(\beta_l+\beta_r)\gamma(\frac{1}{\alpha})} \exp(-(\frac{x}{\beta_r})^\alpha) & x \geq 0 \end{cases} \tag{3.8}$$

where α is the shape parameter and

$$\sigma_l = \sqrt{\frac{\sum_{k=1,x_k<0}^{N_l} x_k^2}{N_l - 1}} \tag{3.9}$$

$$\sigma_r = \sqrt{\frac{\sum_{k=1,x_k<0}^{N_r} x_k^2}{N_r - 1}} \tag{3.10}$$

$$\beta_l = \sigma_l \sqrt{\frac{\gamma(\frac{1}{\alpha})}{\gamma(\frac{3}{\alpha})}} \tag{3.11}$$

$$\beta_r = \sigma_r \sqrt{\frac{\gamma(\frac{1}{\alpha})}{\gamma(\frac{3}{\alpha})}} \tag{3.12}$$

and $\gamma(.)$ is the gamma function:

$$\gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt \quad a > 0 \tag{3.13}$$

The shape parameter α controls the 'shape' of the distribution whereas σ_l^2 and σ_r^2 controls the left and right variance where their relative magnitude controls the skew in the distribution. The parameters are calculated using the estimation procedure described in [23].

3.2 Marginal Statistics from MSCN coefficients:

The MSCN coefficients are zero mean being band pass in nature. Being symmetric, they have same left and right variance. Hence we take average of left and right variance as a feature. The shape parameter which models the peakiness of the distribution is picked as the second feature.

3.3 Marginal Statistics from pointwise local pair wise products:

We fit a AGGD to HC,VC,DC1,DC2 separately to model directional neighbourhood relationships and get the parameters leftvariance, rightvariance and shape as parameters. We also take mean of distribution as a parameter as it informs about where distribution is centered. The mean of AGGD is given by:

$$\mu = (\beta_r - \beta_l)\frac{\gamma(\frac{2}{\alpha})}{\gamma(\frac{1}{\alpha})} \tag{3.14}$$

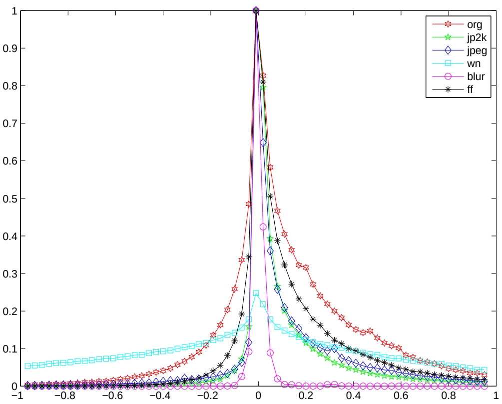
Hence we take 4 parameters:{mean, shape, left variance, right variance} as four parameters from each pointwise direction correlation. We get 16 features, 4 features from each direction.

The perception of image quality also depends on distance from which it is viewed. Further, images are naturally multiscale and distortions affect structure across scales. We therefore analyse the image at 2 scales. We downsample the image using bicubic interpolation and compute same features at downscaled image too. This provides us with 36 features with 18 features at each scale.

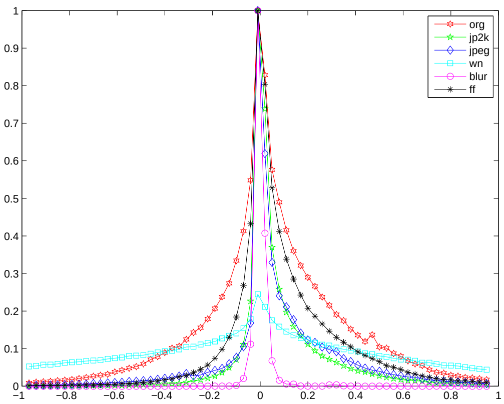
3.4 Quality Evaluation Framework

To gauge how different features correlate with human perception of image quality with different type of distortions, we plot correlation of each feature with subjective scores. In figure 3.8, all features seems to correlate well with human perception for one kind of distortion or the other. This justifies our choice of selected features. We can clearly see that some features correlate more than others for a given distortion. However, as crests and troughs are located at different positions in curves for different distortions, it intuits that important features for each distortion are different depending on nature of distortion. This prompts us to take up a approach which first finds probability of an image being distorted by a particular distortion for every distortion and then learn separate regressors for each distortion.

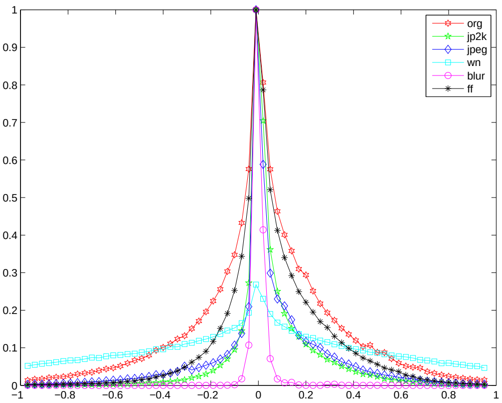
To carry forward this methodology, we adopt a two stage approach



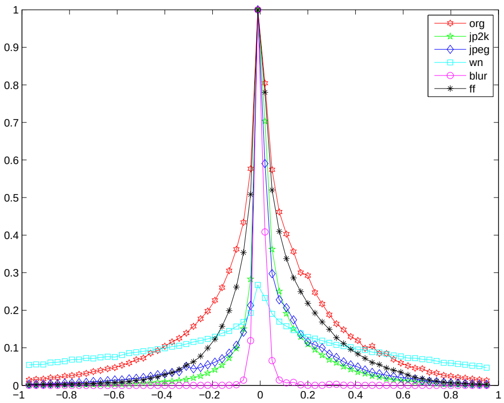
(a) Horizontal



(b) Vertical



(c) Diagonal at 45°



(d) Diagonal at −45°

Figure 3.7: Signatures of pairwise neighbouring MSCN coefficients for different distortions at different orientations

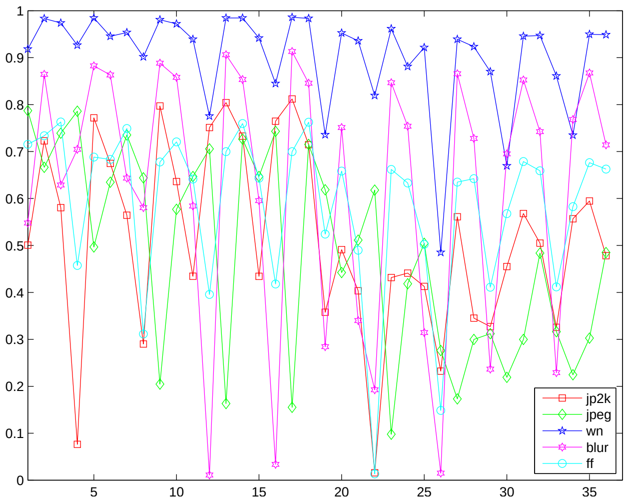


Figure 3.8: Correlation of features with human perception of quality for different distortions

taken by authors in [30] where same set of features are first utilized for distortion probability computing and then used for distortion specific quality assessment. Both stages require training with a set of distorted images with each of them associated with ground truth distortion class and human scores. We would like to emphasize that pristine images have not been utilized for any training for calibration of our framework.

In stage 1, a mapping between the feature vector and true distortion class of the image is learnt from the training set. When a new test image is fed to this classification framework, it provides probability of test image belonging to every distortion class rather than doing a hard decision of class assignment. The intuition behind making this soft decision is images have a mixture of distortions, for example, Blocking is always accompanied by blur. Hence we should assign the probabilities of it belonging to every distortion class.Lets denote this probability vector as $\overline{p}$.

Now that we have distortion probabilites for a given test image, we need the quality score from each regression model trained for every distortion. To accomplish that, a regression function needs to be learnt separately for each distortion which takes set of computed features as an input and regress to the objective quality score. To learn the feature weights for the regression function for a particular distortion, distorted training images are taken with their associated subjective scores for that distortion. When a new test image comes, regressor for each distortion gives out an objective score. Lets denote this probability vector as $\overline{s}$. RISQUE score is the dot product between the two vectors computed above.

The frame work is general to allow use of any classifier and regressor. For our purpose here, we made use of SVM classification and regression [41]. They have been used successfully for quality assessment by authors in [30] and high dimensional data in general [5]. We have utilized the LIBSVM package [12] for implementation purpose. Radial bias kernel has been used as a kernel for both classification and regression.

Chapter 4

Performance Evaluation

Let your performance do the thinking

H. Jackson Brown, Jr.

4.1 LIVE IQA database

We make use of LIVE IQA database [46] for testing the performance of our algorithm. It consists of 29 reference images with 779 distorted images which span five different distortion categories- JPEG and JPEG 2000 compression, white noise, gaussian blur and a Rayleigh fast fading channel distortion. Each of the distorted image has an associated difference mean opinion score (DMOS) which represents how an average human would rate it.

4.2 Performance Comparison with existing algorithms

Since our algorithm involves training at first to enable the classifier to learn how to find distortion probabilities associated with a particular image. Also, the regressor function associated with each distortion needs to be learnt to predict quality. Hence we divide the dataset in to 90% training and 10% testing ensuring that no overlap between train and test content occurs. To make sure that the reported results do not depend on selection of spatial content used to train the framework, we randomly pick 90% of sptial content for training and rest 10% to test and repeat this procedure 1000 times on LIVE database. We then report the median of the performance indices namely spearman rank ordered correlation coefficient (SROCC), pearson linear correlation coefficient(PLCC), root mean squared error (RMSE) between the predicted score from algorithm and the DMOS. Before performing RMSE and PLCC, we need to map it to the scale of DMOS scores by using a logistic non linearity as described in [46]. The performance index close to 1 denotes superior performance for SROCC and PLCC but close to zero denotes the same for RMSE. The results of SROCC, PLCC and RMSE are tabulated in Tables 4.1, 4.2 and 4.3 respectively.

To observe how we perform compared as compared with other algorithms, we have used peak-signal-to-noise ratio (PSNR), structural similarity index(SSIM)[57], multi-scale structural simimlarity index(MS-SSIM) [60] from set of full reference algorithms. As we discussed before, PSNR is not perceptually relevant but due to its widespread use we include it for comparison purposes. The latter indices are gaining popularity owing to their state of art performance and simplicity. We have also included no reference algorithms dicussed in introduction section for comparison purposes. These include Anisotropy based NR IQA [18], BLind Image Integrity Notator using DCT Statistics (BLIINDS)[39] and Distortion Identification-based Image Verity and INtegrity Evaluation (DIIVINE) [30]. The implementation of full

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.8947	0.9191	0.9536	0.8357	0.8964	0.8873
SSIM	0.9461	0.9547	0.9679	0.9321	0.9393	0.9201
MSSSIM	0.9632	0.9773	0.9786	0.9607	0.9464	0.9550
Anisotropy	0.2931	0.1519	0.7286	0.6929	0.5929	0.3614
BLIINDS	0.8287	0.6172	0.8964	0.8714	0.7250	0.6802
DIIVINE	0.9174	0.9337	0.9821	0.9500	0.8714	0.9303
RISQUE	**0.9150**	**0.9563**	**0.9857**	**0.9536**	**0.8821**	**0.9363**

Table 4.1: Median spearman rank ordered correlation coefficient (SROCC) across 1000 train-test combination on LIVE database

	JP2k	JPEG	WN	Blur	FF	All
PSNR	0.9044	0.9375	0.8981	0.8490	0.9119	0.8866
SSIM	0.9575	0.9630	0.9887	0.9395	0.9644	0.9110
MSSSIM	0.9795	0.9357	0.9919	0.9762	0.9689	0.9529
Anisotropy	0.2235	0.1351	0.5855	0.5631	0.5290	0.2332
BLIINDS	0.8393	0.6690	0.9271	0.9004	0.8008	0.6975
DIIVINE	0.9325	0.9517	0.9903	0.9526	0.9068	0.9327
RISQUE	**0.9284**	**0.9743**	**0.9932**	**0.9606**	**0.9259**	**0.9431**

Table 4.2: Median linear correlation coefficient across 1000 train-test combination on LIVE database

reference indices SSIM and MS-SSIM is available online at [56] [59]. Also, we used available online code for anisotropy measure [19] and DIVINE [31]. BLINDS was implemented as described in [39].

As seen from Tables 4.1, 4.2 and 4.3; our approach triumphs over all existing no reference algorithms and better than SSIM and PSNR. Hence we perform better with a simpler, faster and perception based framework. Though we are still worse than MS-SSIM where there is scope of future work.

	JP2k	JPEG	WN	Blur	FF	All
PSNR	10.5570	10.7319	12.1428	9.6039	11.1865	12.5726
SSIM	7.2509	8.4891	4.1872	6.0804	7.2300	11.2531
MSSSIM	4.9428	10.7845	3.5554	3.8664	6.9918	8.2280
Anisotropy	23.8035	30.7586	22.6060	14.9969	22.7926	26.2066
BLIINDS	13.4780	23.0312	10.4778	8.0047	15.3423	19.4712
DIIVINE	8.8453	9.7194	3.8461	5.4923	11.4130	9.7412
RISQUE	**9.2419**	**7.0282**	**3.2235**	**4.9951**	**9.7636**	**9.0662**

Table 4.3: Median RMSE across 1000 train-test combination on LIVE database

4.3 Classification Accuracy

While we are using a probabilistic framework for distortion classification where we use the probability of an image being distorted with a particular distortion, but just as a proof of how good the featues used in the framework act as distortion identifiers and also which distortions are misclassified with which ones, we are reporting the confusion matrix for first stage classification in Table 4.4. We would like to point out that each entry in the confusion matrix is the mean of confusions across 1000 trials. We can see from table 4.4 that fast fading and JPEG2000 are confused with each other. Also, JPEG 2000 and JPEG are also confused sometimes. WN and Blur are comparatively more robust in detection and not confused usually with other distortions.

4.4 Statistical Signifiance Testing

Next, we would like to point out that even though differences in median correlation exist between different algorthms, we need to conform if correlation

	JPEG2000	JPEG	WN	Blur	FF
JPEG200	83.72	7.02	0.05	0.52	8.68
JPEG	8.34	88.46	0.57	1.59	1.04
WN	0	2.01	97.99	0	0
Blur	2.29	2.27	0.61	91.15	3.68
FF	9.11	5.07	0	2.40	83.42

Table 4.4: Table shows the median percentage of confusion of distortions with each other. We would like to point out that each entry in the confusion matrix is the median of confusions across 1000 trials. Hence each row may not sum up to 1.

differences are statisically significant. We use spearman rank ordered correlation coeffficients for checking statistical significance of performance differences between algorithms. The statistical significance is obtained by performing a t-test between scores obtained across 1000 trials for any algorithm in a row with algorithm in column [47]. The null hypopthesis set for t-test is that *mean correlation for row algorithm is equal to mean correlation for column algorithm at confidence of 95%*. The alternate hypothesis, on the other hand states that mean correlation of row is greater than or lesser than the mean correlation of the column. A value of '1' indicates that row algorithm is statiscally better than column algorithm whereas '-1' shows it being worse.

From table 4.5, we can clearly see that RISQUE does a better job than all existing no reference algorithms statistically too. It is also (statistically) better than full reference algorithms PSNR and SSIM. Given the fact that these measures need information from pristine too, this is by no means a small achievement. We can replace full reference algorithms with the current proposed RISQUE frame work without any loss of correlation with human

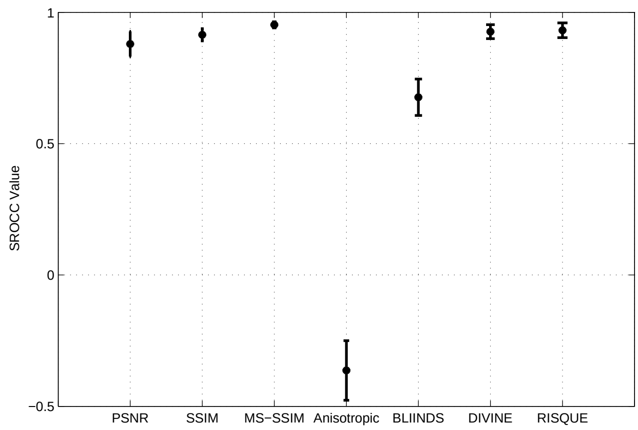


Figure 4.1: Mean SROCC and error bars one standard deviation wide for the algorithms across 1000 train-test trials on LIVE IQA database.

perception and loss of computational speed. This can be applied in gauging quality of real time networks.

As we computed have correlations for each algorithm over 1000 train-test trials, we find mean SROCC value and the standard error associated with these correlation values. In figure 4.1, we plot the same across the dataset along with error bars one standard deviation wide for each of the evaluated algorithms.

	PSNR	SSIM	MSSSIM	Anisotropic IQA	BLIINDS	DIVINE	RISQUE
PSNR	0	-1	-1	1	1	-1	-1
SSIM	1	0	-1	1	1	-1	-1
MSSSIM	1	1	0	1	1	1	1
Anisotropic IQA	-1	-1	-1	0	-1	-1	-1
BLIINDS	-1	-1	-1	1	0	-1	-1
DIIVINE	1	1	-1	1	1	0	-1
RISQUE	**1**	**1**	**-1**	**1**	**1**	**1**	**0**

Table 4.5: Statistical Significance results of one sided t-test performed between SROCC values. A value of '1' indicates that row algorithm is statiscally better than column algorithm whereas '-1' shows it being worse. Value of '0' gives an indication of equivalence between 2 algorithms in consideration.

Chapter 5

Conclusion and Future Work

A conclusion is the place where you get tired of thinking

Arthur Bloch

We proposed a natural scene statistic based quality assessment model **Referenceless Image Spatial QUality Evaluator (RISQUE)** which performs quality assessment of an image with out any information from pristine 'reference' image. No distortion specific features such as ringing, blur or blocking has been modeled in the algorithm in specific. The algorithm only quantifies the 'naturalness'(or lack thereof) in the image due to presence of distortions. The designed framework is spatial domain, human perception based, simpler and faster which makes it superior to other no reference algorithms.

The index is been shown to perform well across different distortions verifying its distortion agnostic nature. An exhaustive analysis of performance is done using LIVE IQA database on five kinds of distortions through pearson linear correlation coefficient, spearman rank ordered correlation coefficent and mean squared error. The frame work is found to perform **statistically better** than other proposed no reference algorithms and full reference structural similarity index(SSIM). Performing superior to full reference algorithms is a great achivement from perspective of doing better with lesser amount of information which validates the feature relevance to human perception. The use of two stage framework similar to DIIVINE makes it extensible to any number of distortions.

Future work will involve improving RISQUE performance to the state of the art full reference quality metric Multi Scale Structural Similarity Index (MS-SSIM) and testing the robustness of framework with more distortions.

Bibliography

[1] F. Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3):183–193, 1954.

[2] R. Barland and A. Saadane. Reference free quality metric using a region-based attention model for JPEG-2000 compressed images. In *Proceedings of SPIE*, volume 6059, pages 605905–1–605905–10, 2006.

[3] H.B. Barlow. Possible principles underlying the transformation of sensory messages. *Sensory Communication, W.A. Rosenblith, Ed.*, 1:217–234, 1961.

[4] A. Bell and T.J. Sejnowski. The 'independent components' of natural scenes are edge filters. *Vision Research*, 37:3327–3338, 1997.

[5] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[6] M. Carandini, D.J. Heeger, and J.A. Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21):8621, 1997.

[7] J.F. Cardoso. Source separation using higher order moments. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 2109–2112. IEEE, 1989.

[8] M. Carnec, P. Le Callet, and D. Barba. An image quality assessment method based on perception of structural information. In *International Conference on Image Processing*, volume 2, pages 185–188. IEEE, 2003.

[9] M. Carnec, P. Le Callet, and D. Barba. Visual features for image quality assessment with reduced reference. In *International Conference on Image Processing*, volume 1, pages 421–424, 2005.

[10] J. Caviedes and S. Gurbuz. No-reference sharpness metric based on local edge kurtosis. In *International Conference on Image Processing*, volume 3, pages 53–56. IEEE, 2002.

[11] D.M. Chandler and S.S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 16(9):2284–2298, 2007.

[12] C.C. Chang and C.J. Lin. LIBSVM: a library for support vector machines. 2001.

[13] J. Chen, Y. Zhang, L. Liang, S. Ma, R. Wang, and W. Gao. A no-reference blocking artifacts metric using selective gradient and plainness measures. pages 894–897, 2008.

[14] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. *Digital images and human vision*, 4, 1993.

[15] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. New full-reference quality metrics based on HVS. In *International Workshop on Video Processing and Quality Metrics*, 2006.

[16] R. Ferzli and L.J. Karam. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB). *IEEE Transactions on Image Processing*, 18(4):717–728, 2009.

[17] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *Journal of Optical Society of America*, 4(12):2379–2394, 1987.

[18] S. Gabarda and G. Cristóbal. Blind image quality assessment through anisotropy. *Journal of Optical Society of America*, 24(12):42–51, 2007.

[19] S. Gabarda and S. Cristobal. Image Anisotropy Code [Online]Available:. http://www.iv.optica.cisc.es/page49/page51/page51.html.

[20] I.P. Gunawan and M. Ghanbari. Reduced-reference picture quality estimation by using local harmonic amplitude information. 2003, 2003.

[21] R.S. Hunter. Photoelectric color difference meter. *Journal of Optical Society of America*, 48(12):985–993, 1958.

[22] T.M. Kusuma and H.J. Zepernick. A reduced-reference perceptual quality metric for in-service image quality assessment. In *Joint First Workshop on Mobile Future and Symposium on Trends in Communications*, pages 71–74. IEEE, 2003.

[23] N.E. Lasmar, Y. Stitou, and Y. Berthoumieu. Multiscale skewed heavy tailed model for texture analysis. In *International Conference on Image Processing*, pages 2281–2284. IEEE, 2009.

[24] G.E. Legge and J.M. Foley. Contrast masking in human vision. *Journal of Optical Society of America*, 70(12):1458–1471, 1980.

[25] Q. Li and Z. Wang. Reduced-reference image quality assessment using divisive normalization-based image representation. *IEEE Journal of Selected Topics in Signal Processing*, 3(2):202–211, 2009.

[26] X. Li. Blind image quality assessment. In *International Conference on Image Processing*, volume 1, pages 449–452. IEEE, 2002.

[27] J. Lubin. The use of psychophysical data and models in the analysis of display system performance. pages 163–178, 1993.

[28] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi. Perceptual blur and ringing metrics: application to JPEG2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

[29] L. Meesters and J.B. Martens. A single-ended blockiness measure for JPEG-coded images. *Journal of Signal Processing*, 82(3):369–387, 2002.

[30] A.K. Moorthy and A.C. Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*.

[31] A.K. Moorthy and A.C. Bovik. DIVINE Code implementation [Online]Available:. http://live.ece.utexas.edu/research/quality/DIVINE_release.zip.

[32] A.K. Moorthy and A.C. Bovik. Statistics of natural image distortions. In *International Conference on Acoustics Speech and Signal Processing*, pages 962–965. IEEE.

[33] N.D. Narvekar and L.J. Karam. A no-reference perceptual image sharpness metric based on a cumulative probability of blur detection. In *International Workshop on Quality of Multimedia Experience*, pages 87–91. IEEE, 2009.

[34] E. Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15(3):267–273, 1982.

[35] B.A. Olshausen and D.J. Field. How close are we to understanding V1? *Neural Computation*, 17(8):1665–1699, 2005.

[36] E.P Ong, W. Lin, Z. Lu, S. Yao, X. Yang, and L. Jiang. No-reference jpeg-2000 image quality metric. In *International Conference on Multimedia and Expo*, pages 545–548. IEEE Computer Society, 2003.

[37] D.L. Ruderman. The statistics of natural images. *Network computation in neural systems*, 5(4):517–548, 1994.

[38] M. Saad, A.C. Bovik, and C. Charrier. Model-Based Blind Image Quality Assessment Using Natural DCT Statistics . *IEEE Transactions on Image Processing*, 2011.

[39] M.A. Saad, A.C. Bovik, and C. Charrier. A DCT statistics-based blind image quality index. *Signal Processing Letters, IEEE*, 17(6):583–586, 2010.

[40] N.G. Sadaka, L.J. Karam, R. Ferzli, and G.P. Abousleman. A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling. In *International Conference on Image Processing*, pages 369–372. IEEE, 2008.

[41] B. Schölkopf, A.J. Smola, R.C. Williamson, and P.L. Bartlett. New support vector algorithms. *Neural computation*, 12(5):1207–1245, 2000.

[42] K. Seshadrinathan and A.C. Bovik. Unifying analysis of full reference image quality assessment. In *International Conference on Image Processing*, pages 1200–1203. IEEE, 2008.

[43] H.R. Sheikh and A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.

[44] H.R. Sheikh, A.C. Bovik, and L. Cormack. No-reference quality assessment using natural scene statistics: JPEG2000. *IEEE Transactions on Image Processing*, 14(11):1918–1927, 2005.

[45] H.R. Sheikh, A.C. Bovik, and G. De Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14(12):2117–2128, 2005.

[46] H.R. Sheikh, M.F. Sabir, and A.C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, 2006.

[47] D. Sheskin. *Handbook of parametric and nonparametric statistical procedures*. 2004.

[48] R. Soundararajan and A.C. Bovik. RRED Indices: Reduced Reference Entropic Differencing Framework for Image Quality Assessment. In *International Conference on Acoustics, Speech and Signal Processing*, 2011.

[49] S. Suthaharan. No-reference visually significant blocking artifact metric for natural scene images. *Journal of Signal Processing*, 89(8):1647 – 1652, 2009.

[50] P.C. Teo and D.J. Heeger. Perceptual image distortion. In *International Conference on Image Processing*, volume 2, pages 982–986. IEEE, 1994.

[51] H. Tong, M. Li, H.J. Zhang, and C. Zhang. No-reference quality assessment for JPEG2000 compressed images. In *International Conference on Image Processing*, volume 5, pages 3539–3542. IEEE, 2004.

[52] S. Varadarajan and L.J. Karam. An improved perception-based no-reference objective image sharpness metric using iterative edge refinement. In *International Conference on Image Processing*, pages 401–404. IEEE, 2008.

[53] M.J. Wainwright, O. Schwartz, and E.P. Simoncelli. Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons. *Statistical theories of the brain*, pages 203–222, 2002.

[54] Z. Wang and A.C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, 2002.

[55] Z. Wang and A.C. Bovik. Mean squared error: love it or leave it? A new look at signal fidelity measures. *IEEE Signal Processing Magazine*, 26(1):98–117, 2009.

[56] Z. Wang, A.C. Bovik, and E.P. Sheikh, H.R.and Simoncelli. SSIM Code [Online]Available:. http://live.ece.utexas.edu/research/Quality/index.htm.

[57] Z. Wang, A.C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[58] Z. Wang, H.R. Sheikh, and A.C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *International Conference on Image Processing*, volume 1, pages 477–480. IEEE, 2002.

[59] Z. Wang, E.P. Simoncelli, and A.C. Bovik. MS-SSIM Code [Online]Available:. http://live.ece.utexas.edu/research/Quality/index.htm.

[60] Z. Wang, E.P. Simoncelli, and A.C. Bovik. Multiscale structural similarity for image quality assessment. In *Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402. IEEE, 2003.

[61] Z. Wang, G. Wu, H.R. Sheikh, E.P. Simoncelli, E.H. Yang, and A.C. Bovik. Quality-aware images. *IEEE Transactions on Image Processing*, 15(6):1680–1689, 2006.

[62] A.B. Watson. *Digital images and human vision.* MIT Press, 1993.

[63] S. Wolf and M.H. Pinson. Spatio-temporal distortion metrics for in-service quality monitoring of any digital video system. 3845:266277, 1999.

[64] S. Wolf and M.H. Pinson. Low bandwidth reduced reference video quality monitoring system. In *International Workshop on Video Processing and Quality Metrics*, 2005.

[65] X. Zhu and P. Milanfar. A no-reference sharpness metric sensitive to blur and noise. In *International Workshop on Quality of Multimedia Experience*, pages 64–69. IEEE, 2009.

Vita

Anish Mittal received the Bachelor of Technology from Indian Institue of Technology,Roorkee with a CGPA of 8.7 on the scale of 10 in 2009. He joined the University of Texas at Austin in the Fall of 2009 and joined the Laboratory of Image and Video Engineering (LIVE) under Dr. Alan C. Bovik, co-advised by Dr Lawrence Cormack and is currently working towards his doctorate in the area of stereo image and video quality assessment at LIVE.

Email address: mittal.anish@gmail.com

This thesis was typed by the author.