Chapter 5
An Improved Model for No-Reference Image Quality Assessment and a No-Reference Video Quality Assessment Model based on Frame Analysis

5.1 Introduction

In last decade use of mobile devices to provide AR applications has increased. Multimedia technology has led to the usage of smart mobile phones and other electronic devices, making digital images an important intermediate to obtain information and to establish a more convenient source of interaction. However, various deformities are present in digital images while going through processes like acquisition, compression, transmission, reproduction etc. These deformities mainly occur due to shortage of access devices, storage media, processing technologies and transmission equipment. Image distortions severely influence the ability of humans to excerpt and understand the information contained in images. It also degrades the quality of user interfaces in an Augmented Reality (AR) application. Therefore, it becomes important to identify and measure image distortion in order to ensure, control and enhance image quality. In order to achieve the above objective, various Objective Image Quality Assessment (OIQA) methods have been developed and are of substantial practical significance [Lin and Kuo 2011, Gao et al. 2010].

Availability of an original reference image classifies Image Quality Assessment (IQA) methods into three categories: Full-Reference (FR), Reduced-Reference (RR) and No-Reference (NR). Most of the existing approaches falls under the category of Full-Reference Image Quality Assessment FR-IQA i.e. a complete reference image is assumed to be known. However, in many practical applications, the reference image is not available making No-Reference Image Quality Assessment (NR-IQA) or “blind” quality assessment approach desirable. Similarly, in cases when the reference image is only partially available, for example, set of extracted features used as side information to help evaluate the quality of the distorted image, Reduced-Reference Image Quality Assessment (RR-IQA) approach is adopted for quality analysis.

Present literature on NR-IQA could be divided into two categories namely Distortion-Specific quality assessment and General-Purpose quality assessment. The former quantifies a specific distortion regardless of other factors and scores a distorted image accordingly. A number of NR-IQA methods following such an approach could be listed, for example, Chen and Bovik [Chen and Bovik 2011] developed an NR-IQA method to quantify blur in an image. Similarly, Zhu and Milanfar [Zhu and Milanfar 2009] focused on noise, Sazzad et al. [Sazzad et al. 2008] on JPEG2000 distortion, Sheikh et al. [Sheikh et al. 2005] on JPEG2000 by using Natural Scene Statistics (NSS), Wang et al. [Wang et al. 2000] on JPEG etc.

Besides many advantages, distortion-specific approach limits its applicability with the fact that the type of distortion present in the image should be known in advance. Thus, the later approach i.e. General-Purpose quality assessment based on training and learning is widely adapted for NR-IQA purpose. Examples include a two-step framework designed for distortion classification and distortion-specific quality assessment using several NSS features to implement a simple NR-IQA index named Blind Image Quality Index (BIQI) [Moorthy and Bovik 2010]. This method was later improved by using a series of NSS features in the wavelet domain to predict image quality and is referred as Distortion Identification-based Image Verity and Integrity Evaluation (DIIVINE) [Moorthy and Bovik 2011]. In 2012, Saad et al. [Saad et al. 2012] proposed another efficient NR_IQA method named BLind Image Integrity Notator using Discrete Cosine Transform (DCT) Statistics-II Index (BLIINDS-II) which extracts NSS features in the block DCT domain using a fast single-stage framework.

In order to achieve better predictive performance with low computational complexity, Mittal et al. in 2012 [Mittal et al. 2012] proposed the Blind/Reference-less Image Spatial Quality Evaluator (BRISQUE) NR-IQA metric. In 2013, Mittal et al. [Mittal et al. 2013] proposed the Naturalness Image Quality Evaluator (NIQE) NR-IQA metric based on quality aware collection of statistical features. Later, Liu et al. in 2014 [Liu et al. 2014] proposed an NR-IQA metric based on spatial and spectral entropies of image pixels (Spatial and Spectral entropies based IQA (SSEQ)). The method was examined to match the perceptual quality of an image being statistically superior to many other NR-IQA metrics, like BIQI and DIIVINE.

While research in the field of NR-IQA has progressed over the years by analyzing wide range of features that deteriorate an image quality, No-Reference Video Quality Assessment (NR-VQA) models are rather less and are mostly application specific. However, over the years, research for Distortion-Specific video quality assessment techniques have advanced with an aim to design a universal measure for NR-VQA. The most prominent distortion evaluated in a video is compression and hence in literature, many of the NR-VQA metrics are geared towards compression. Among many quantified artifacts of

compression, blocking is one of the most common artifact [Suthaharan 2003, Muijs and Kirenko 2005]. Jerkiness artifact in videos, also known as strobing, is also used for quantifying video quality [Ou et al. 2011]. A technique for evaluating jerkiness in a video was given by Ong et al. [Ong et al. 2009] by finding the absolute frame differences between adjacent frames in a video. A non-application specific NR-VQA model proposed by Keimel et al. [Keimel et al. 2009] quantifies a multitude of factors to predict the overall video quality. In a study conducted by Saad and Bovik [Saad and Bovik 2012], a NR-VQA model based on the principles of natural video statistics is proposed where motion characteristics in a video are quantified by extracting block motion estimates and DCT coefficient differences between adjacent frames.

In present research, an attempt is made to analyze image and video features that deteriorates the quality of a given image/video. Thereby, a fundamental understanding of spatial and temporal artifacts is acquired for quantifying the distortions present in image/video using a generalized metric. Therefore, in order to identify features that deteriorate an image/video quality, an improved NR-IQA Model is proposed by combining the efficiency of three existing NR-IQA metrics (NIQE, BRISQUE and BLIINDS-II) using Multi-Linear Regression (MLR). Also, a NR-VQA model is presented by exploring the quantification of certain distortions like, ringing, frame difference, blocking, clipping and contrast in video frames.

5.2 Proposed Model for No-Reference Image Quality Assessment

NR-IQA analysis is of significant importance while testing quality of outputs of AR based user interface, newly designed codecs and finding optimal settings for broadcasting digital video etc. In such scenarios, subjective testing becomes an expensive affair and sometimes even impossible, for example, in case of live streaming videos. Therefore, for effectively performing the objective quality analysis of the test image, three pre-existing NR-IQA metrics, namely, NIQE, BRISQUE and BLIINDS-II are combined using MLR model. The three methods are identified from literature on image quality assessment that are significantly different from each other, else, the distortion measurement will be redundant as the pooling of data will become useless [Moorthy and Bovik 2009]. Many such models were identified in the literature, however, NIQE, BRISQUE and BLIINDS-II are chosen because of their diverse footprint over quantifying different distortions.

Multi-Linear regression model is used to describe a single response variable Y which linearly depends upon three predictor variables (NIQE, BRISQUE and BLIINDS-II). The combined model for blind estimation of image quality performs marginally better than BRISQUE (results are discussed in Section 5.5.1), which individually has the best performance among the three metrics. Performance comparison of

the three metrics is done by evaluating the overall correlation with Differential Mean Opinion Score (DMOS).

5.3 Proposed Model for No-Reference Video Quality Assessment

NR-VQA remains to be the most researched field among all media quality estimation fields. There are presently no perceptual models of distortion that may apply to the NR case. However, a promising approach consists of identifying the minimum set of features which influence the quality in most situations and examine their ability to predict the perceived quality. The main problem is that, certain algorithms look out for certain distortions only and with each additional distortion to monitor, the computational complexity goes up. Moreover, some distortions are content dependent and this makes it difficult to come up with a general algorithm for NR-VQA.

Though it is difficult to accurately predict the quality of a video without the availability of reference data, its applications have tremendous importance in the market. The most important application is to design flexible real-time control systems to monitor and deliver high quality steams to consumers. Reliable metrics are important for enabling transparent and competitive ratings of Quality of Service (QoS), which would benefit both the consumers and the producers. Therefore, in this study, to inspect some of the above related issues, quantification of certain distortions is explored and an attempt is made to estimate these distortions more accurately to provide a robust NR-VQA model.

5.3.1 Ringing in a Frame

The ringing effect is most illustrious in edges between high contrast areas with smooth texture. It is noticeable as simmers and ripples extending outwards from the edge up to the blocks which forms the boundary along the edge. The amplitude of the ripple effect is high for high contrast images and vice versa.

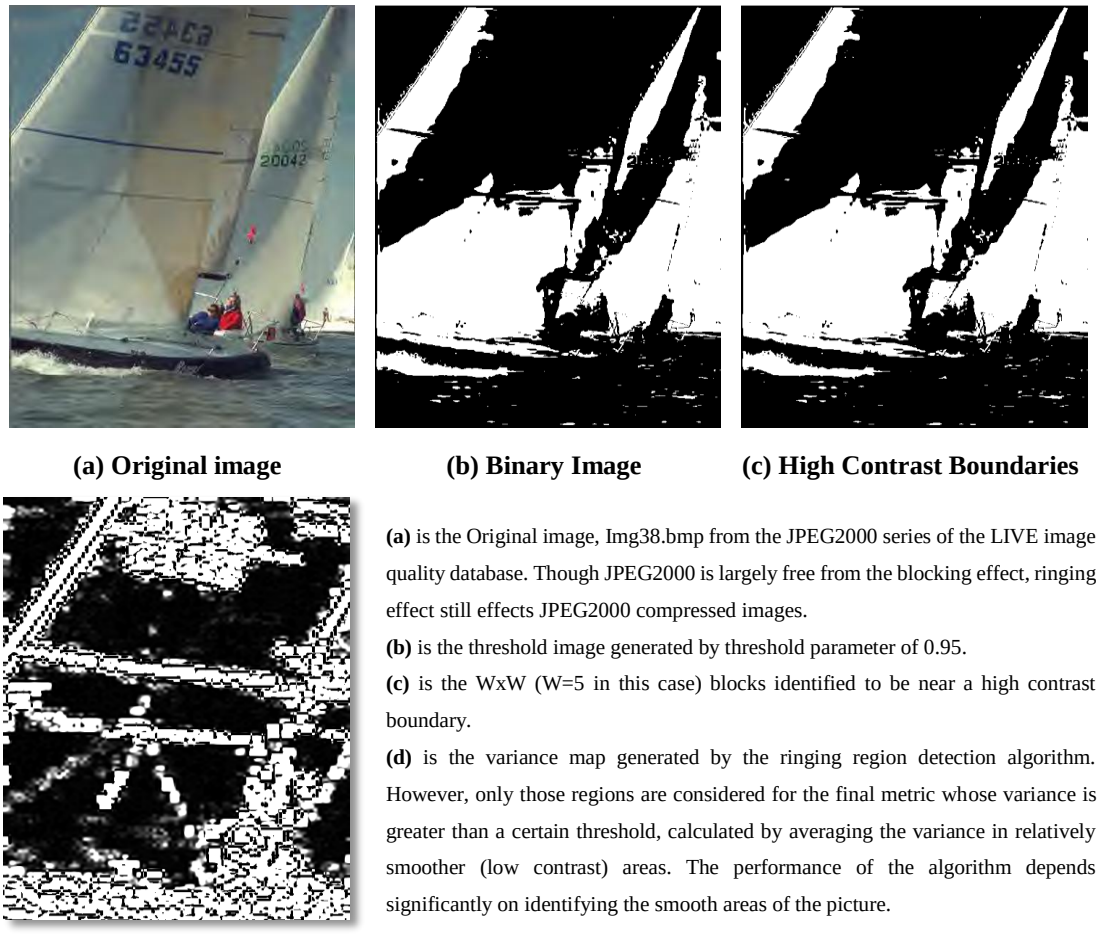
Till date, existing research discusses many methods of measuring ringing in an image, however, all have one thing in common: almost all methods rely on some or the other transformation to transform the domain of the image. Fourier transform is popular among such transformations for transforming an image from spatial domain to the frequency domain. This is usually done to observe which spatial frequencies are observable and which might be masked. Fourier Transform therefore is the bottleneck to the computational complexity of all algorithms that measure ringing in an image. This is acceptable as far as it is limited for quality analysis to calibrate image compression algorithms or an image processing tool, but for large image volume or live video streams this is an impossibility. Keeping this constraint in mind,

a method is designed for estimating ringing in an image without undergoing any actual transformation.

In this method high contrast boundaries are identified. This is done by thresholding the image with a suitable luminance threshold to generate a binary image. It is at these high contrast edges that the need to identify "splatter", i.e. isolated pixels or groups of pixels that widely differ in their luminance to their surrounding pixels luminance (Figure 5.1), becomes essential. For this, variance of luminance is calculated in a window of suitable size in areas near to high contrast boundary.

5.3.2 Frame Difference (Δf)

Frame difference is a (very) rough estimate of the activity in the video. Faster temporal changes in pixel values will lead to greater frame difference. However, it is video dependent and can vary drastically from one video to another. The LIVE Video Quality database (discussed in Section 5.4.2, Section 5.5,



(a) Original image **(b) Binary Image** **(c) High Contrast Boundaries**

(d) Variance map showing ringing area

(a) is the Original image, Img38.bmp from the JPEG2000 series of the LIVE image quality database. Though JPEG2000 is largely free from the blocking effect, ringing effect still effects JPEG2000 compressed images.
(b) is the threshold image generated by threshold parameter of 0.95.
(c) is the WxW (W=5 in this case) blocks identified to be near a high contrast boundary.
(d) is the variance map generated by the ringing region detection algorithm. However, only those regions are considered for the final metric whose variance is greater than a certain threshold, calculated by averaging the variance in relatively smoother (low contrast) areas. The performance of the algorithm depends significantly on identifying the smooth areas of the picture.

Fig. 5.1. Quantification for Ringing Effect in a Frame

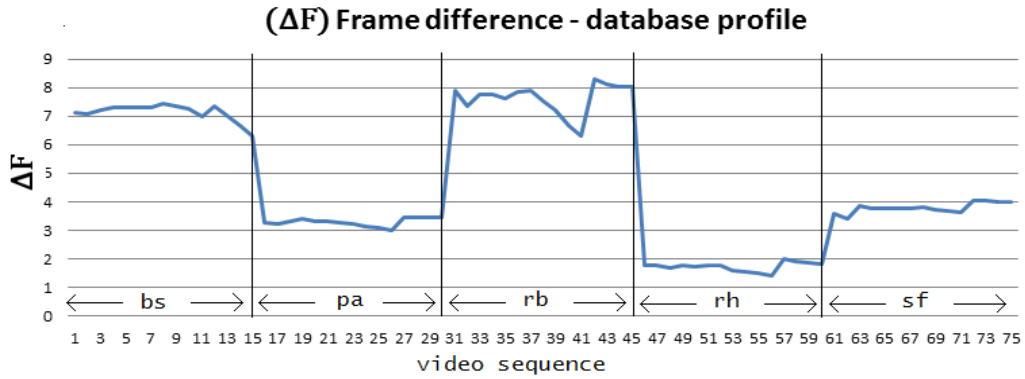


Fig. 5.2. Frame Difference is observed to be highly Video Content Dependent ('bs', 'pa', rb', 'rh', 'sf' are abbreviations for Video Dataset Provided by LIVE (refer section 5.5))

and Appendix A.3) had no videos with scene change. All videos content refers to either panning of a natural scene or observing a smooth moving object. From Figure 5.2 it is clear that using frame difference as-it-is is a bad idea. However, a beneficial result of this experiment is the determination of Δf as a strong metric for estimating scene changes in a larger scale video. Moreover, using psycho visual experiments, higher threshold of scene changes per second/minute beyond which the quality of video deteriorates could also be estimated.

5.3.3 Blocking Effect Quantization

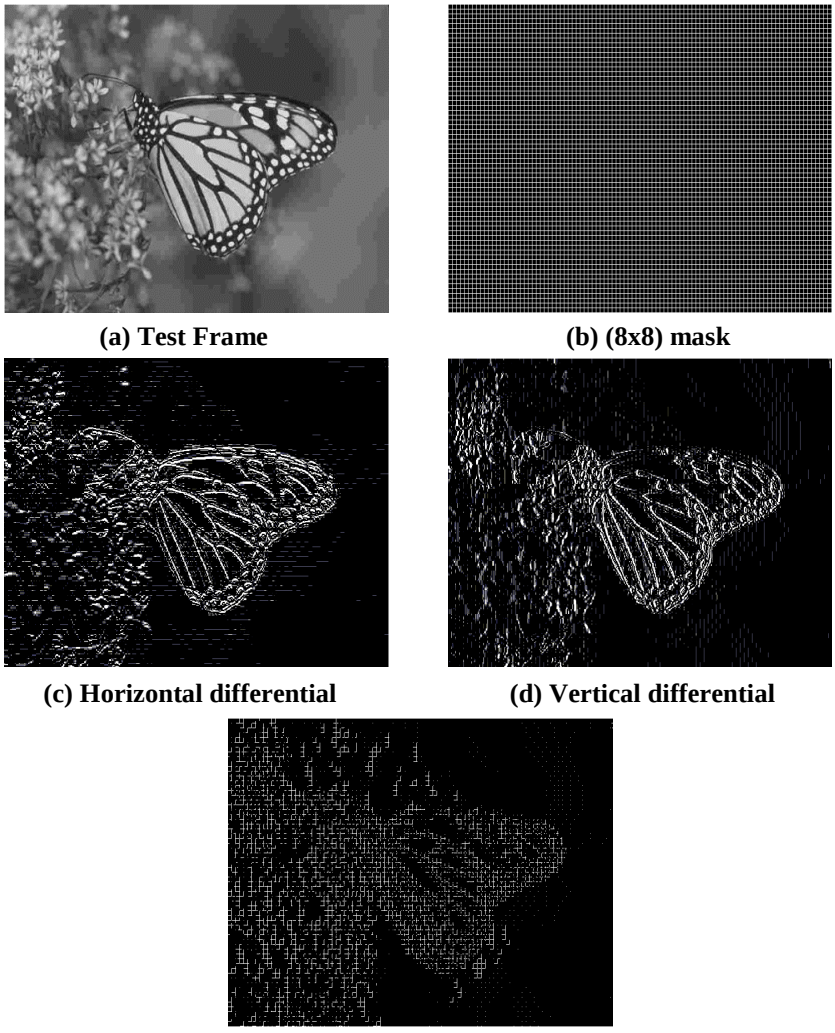
Blocking effect is the most popular video artifact. It is also among the simpler artifacts to be observed as its location is fixed in the spatial domain. Blocking is inherent in all lossy compression algorithms that use DCT. Wavelet Transform is independent of this artifact making JPEG2000 a better format quality wise, however it too suffers from ringing effect, the next most common effect.

In order to quantify the blocking effect in an image or frame of a video, firstly, an arbitrary edge detection algorithm (sobel edge detection) is used to identify the perceptually sensitive areas. A single degree differential is applied on the image in both horizontal and vertical dimensions. Depending on the block dimensions (4 or 8) a mask is created (Figure 5.3) to highlight all differences at block boundaries, making the method computationally less intensive than actually parsing all block edges by looping. The boundary distortions are enhanced by squaring and root of the sums of these horizontal and vertical squared images is taken to obtain the final image. All block boundary pixel luminosities are then added to get an estimate of blocking effect. This value is normalized using the size of the image to keep the metric independent of image size.

Two different block sizes are considered for the experiments, 4x4 and 8x8. The block size 8x8 is understandable because in MPEG-2 compressed videos that is where the blocking effect is located, however, in case of H.264 compressed MPEG-2/AVC videos, it supports multiple block sizes, 4x4 being among the popular ones. Though it is observed that both have a high correlation among them (>90%), both block sizes are considered in this study for completeness sake and to bring generality to the proposed model.

5.3.4 Clipping

Clipping is defined as the truncation in the number of bits of the luminance or chrominance components of the image values (clipping histogram for a sample video is shown in Figure 5.4). It results



(a) Test Frame **(b) (8x8) mask**

(c) Horizontal differential **(d) Vertical differential**

(e) The final image is the combined effect of horizontal and vertical squared differentials

Fig. 5.3. Blocking Effect Quantization using (8x8) mask

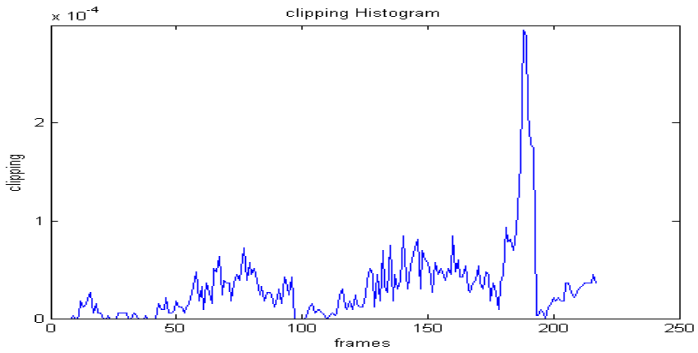


Fig. 5.4. Clipping for a sample video (temporal profile)

in abrupt cutting of peak values at the top and bottom of the dynamic range, which leads to aliasing artifacts caused by the high frequencies created at those discontinuities. The sharpness enhancing technique known as peaking can lead to clipping. In peaking, edges are enhanced by adding positive and negative overshoots to it, but if these values are beyond the limits of the dynamic range ([0,255] for 8 bit precision), then saturation occurs and pixels are clipped.

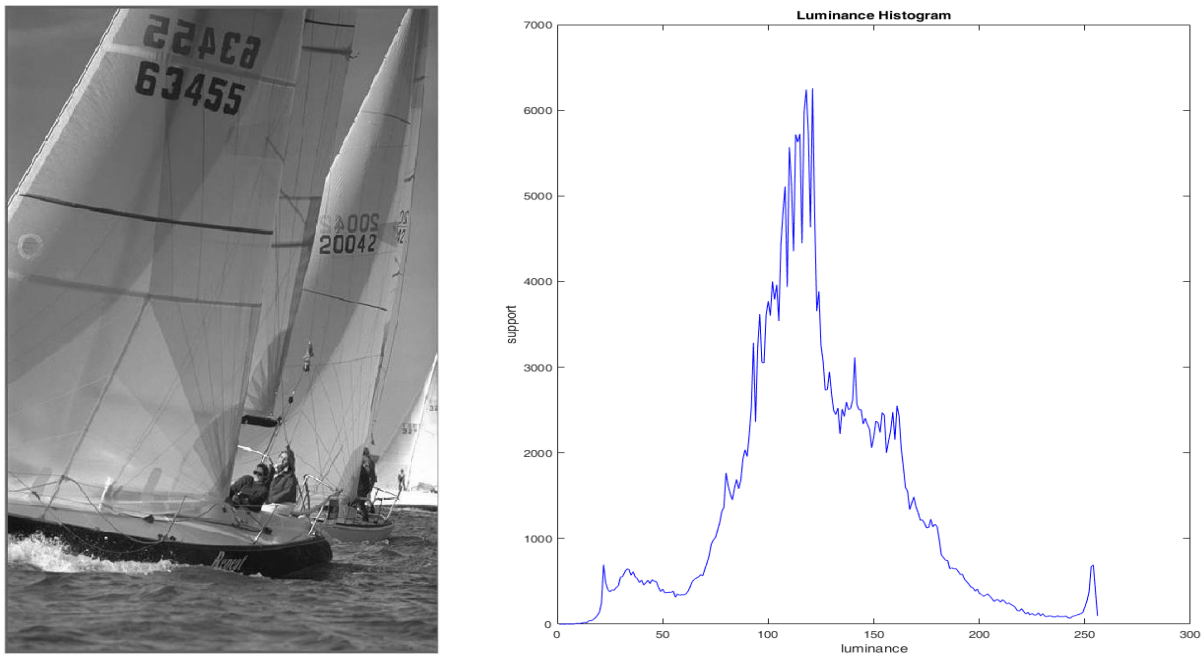
Clipping can be represented mathematically as the percentage of pixels having boundary values i.e. either 0 or 255 for 8 bit precision. However, care must be taken at the margins where in some videos a blank line is introduced due to coding error. To take care of this, pixels till a particular distance from the margin are ignored.

5.3.5 Contrast

Contrast sensitivity, which largely depends upon on the dynamic range of the luminance signal, is the ability to distinguish objects from the background. The perception of contrast is subjective because it depends upon other factors including the mental reference image of the object and sometimes color. Following procedure is used for contrast quantification:

1. Luminance histogram for the given image/video-frame is computed (an example shown in Figure 5.5).
2. The Luminance histogram is then divided in half (vertically) so that luminosities less than half of the maximum luminosity lie on one side of the histogram.
3. Difference between the cumulative luminance of each part is calculated and normalized by dividing the value with the average luminance of the image.

Significance of the proposed NR-VQA model: The features identified by this work are limited but diverse enough to provide an overall quality assessment to the video. Most generic algorithms quantify blocking effect to get an accuracy of around 80% (For example Generalized Block-edge Impairment Metric (GBIM) [Wu and Yuen 1997] which doesn't scale too well with videos). However, the proposed



(a) Image **(b) Luminance Histogram**

Fig. 5.5. In the above Luminance Histogram (b) for image (a), it can be easily seen that majority of the luminance present are in the upper half of the histogram

model involves the quantification of more features, namely clipping and contrast, while keeping simplicity intact.

In this study, four features are dedicated for estimating the quality of the video in spatial domain. Metrics to calculate the ringing effect in an image and the frame difference are explored but not considered as they did not scale too well in case of videos. Therefore, the four features selected are: Blocking4 (Blocking with block size 4x4), Blocking8 (Blocking with block size 8x8), Clipping and Contrast. Additionally two features are dedicated for estimating the quality of video in the temporal profile. Earlier in this chapter [Section 5.3.2], an argument is presented regarding the unsuitability of inter-frame difference (Δf) quantification for NR-VQA as it is heavily dependent upon video content, however, has its uses in other places like detecting frame freeze, scene change etc. and as LIVE video quality database (discussed in Section 5.4.2, Section 5.5, and Appendix A.3) restricts the quality estimation of only smooth motion videos, the two features used for quantifying temporal distortion are: Clipping Standard Deviation (observed as glaring/anti-aliasing in a video) and Contrast Standard Deviation (observed as flickering in video).

If these features are treated independent of interactions among themselves, and an MLR model is used

to predict collective DMOS, the results are rather dismal (because LIVE Video Quality Database restricts the quality estimation of only smooth motion videos i.e., there is not much difference between current and next frames of video) as shown in Table 5.1. The MLR analysis results are ineffectual and this is intuitive as knowledge of feature interactions cannot be known. Moreover, presently, according to the present data, suggesting a non-linear model that these features are expected to follow is difficult. Therefore, for fitting such data, a Neural Network (NN) model with one middle hidden layer of 15 units is used. The NN model is represented in Figure 5.6.

5.4 Methodology & Experimental Setup

5.4.1 Methodology

Methodology and how to compare the performance: For NR-IQA, three diverse metrics are selected and MLR is performed on their results to come up with a combined and a more efficient estimation of quality. The combined model is tested with individual performance of three metrics in terms of DMOS correlation with five distortions. Finally, average DMOS correlation coefficient is calculated considering all five distortions for the three existing metrics and the combined model.

Table 5.1. MLR Model for Video Quality Assessment

Training ratio	Maximum Correlation	Minimum Correlation	Average Correlation
0.5	0.4847	0.1733	0.26906
0.66	0.6299	0.2369	0.27458
0.75	0.6299	0.2369	0.28884
0.8	0.6707	0.2369	0.29423

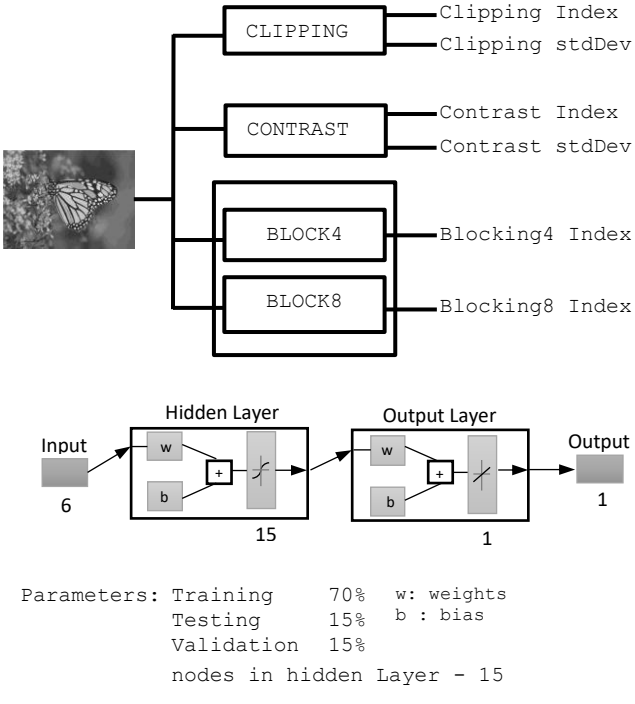


Fig. 5.6. NR-VQA using NN Model

For NR-VQA, certain feature quantification metrics are proposed, namely, ringing effect quantization metric, frame difference metric, a multi-scaling blocking metric, metric for clipping and metric for contrast quantification. A NN model with one middle hidden layer of 15 units is used for fitting appropriate metrics that quantify temporal and spatial distortions in a video.

Which statistics/metrics to use and how: Table 5.2 describes the corresponding tables and figures listing in the chapter with respect to NR-IQA and NR-VQA.

5.4.2 Experimental Setup

Language, Software and Tools used for implementation and system specification: The experiments are carried out using single threaded code on a computer with 16GB RAM and Intel® Core ™ i5-3470 CPU@3.20ghz × 4 processor with cache size of 6144 KB.

Implementation details: The experiments are conducted using LIVE Image/Video Quality database. For Image Quality Database, LIVE, in collaboration with The Department of Psychology at the University of Texas, Austin, performed an extensive experiment to obtain scores from human subjects for a number of images distorted with different distortion types. This data is made available by Sheikh et al. [Sheikh et al. 2006, Appendix A.2] and the images are acquired in support of a research project on generic shape matching and recognition. LIVE Video Quality Database supplements the LIVE Image Quality Database to provide researchers with a much-needed tool to advance the state-of-the-art in objective video quality assessment [Seshadrinathan et al. 2010, Appendix A.3]. The three NR-IQA metrics, NIQE, BRISQUE and BLIINDS-II are implemented using MATLAB implementations made available by their

Table 5.2. Tables & Figures Representing Respective Performance Evaluation

	Table / Figure	Comments
No-Reference Image Quality Assessment	Table 5.4 Table 5.5 Table 5.6 Table 5.7 Figure 5.7 Figure 5.8 Figure 5.9 Figure 5.10	Tables (Table 5.4, Table 5.5 and Table 5.6) and Figures (Figure 5.7, Figure 5.8 and Figure 5.9) shows statistics for DMOS correlation with five distortions for NIQE, BRISQUE and BLIINDS-II respectively. Table 5.7 and Figure 5.10 shows sample results for one iteration of the proposed model.
No-Reference Video Quality Assessment	Figure 5.11	Figure 5.11 represents the goodness of fit results for a NN trial for training, validation and testing stage of the proposed NR-VQA model.

respective developers [Appendix B.1]. For other implementation details with respect to the proposed NR-IQA and NR-VQA model, refer Appendix C.

5.5 Data Reporting

Dataset used for experiments: LIVE Image/Video Quality Assessment database is used for the experiments. The image database consists of primarily five types of distortions, each distortion type provided separately [Sheikh et al. 2006]:

1. Bit errors in JPEG200 bit-streams (fast fading distortion) (FF) – 145 images
2. JPEG2000 compressed images (J2)– 175 images
3. White Noise distortion (WN)– 145 images
4. Gaussian Blur distortion (GB)– 145 images
5. JPEG compressed images (J)– 169 images

DMOS value for each of the 779 images is provided in MATLAB compatible .mat files.

The LIVE Video Quality Database uses ten uncompressed high-quality videos, detailed in Table 5.3, with a wide variety of content as reference videos. These reference videos are down sampled using various techniques to obtain the distorted videos in this database. A set of 150 distorted videos are created from these reference videos (15 distorted videos per reference video) using four different distortion types: MPEG-2 compression, H.264 compression, Simulated Transmission of H.264 compressed bit-streams through error-prone IP networks and Simulated Transmission of H.264 compressed bit-streams through error-prone wireless networks.

The mean and variance of the DMOS obtained from the subjective evaluations, along with the reference and distorted videos, is made available as part of the database.

Table 5.3. LIVE Video Quality Database Reference Videos

Video	Abbreviation	FPS (Frames per second)	Number of Frames
Blue sky	bs	25	217
Mobile and Calendar	mc	50	500
Pedestrian Area	pa	25	250
Park run	pr	50	500
Riverbed	rb	25	250
Rush hour	rh	25	250
Sunflower	sf	25	250
Shields	sh	50	500

5.5.1 No-Reference Image Quality Assessment

Through experiments, it is found that NIQE performs well with most type of distortions but performs poorly for JPEG compression artifacts and white noise, which destroys its overall accuracy. Table 5.4 represents the results of its accuracy over five distortions of LIVE image database and Figure 5.7 represents six scatter plots against NIQE and five types of distortions respectively with one additional plot for all distortion types. Figure 5.7(a), (b) and (c) exhibits good performance of NIQE with FF, GB and J2 distortions, while from Figure 5.7(d), (e) and (f) it is observed that the poor performance of NIQE with J and WN distortion destroys its overall estimation effort, leading to low overall accuracy of NIQE.

Table 5.5 represents the results for BRISQUE accuracy for five respective distortion types. Figure 5.8 shows six scatter plots, one each for representing the performance of BRISQUE with five distortion types and one for representing the performance of BRISQUE with all distortion types. From the scatter plots (Figure 5.8) it is observed that, in contrast to NIQE's poor performance on WN distortion, BRISQUE does exceptionally well (Figure 5.8(e)). For all other distortion types too, i.e. FF, GB, J2 and J, BRISQUE is observed to outperform NIQE.

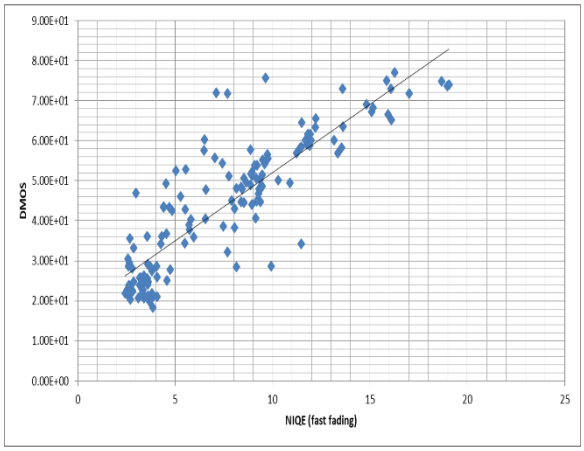
It is also observed that compared to the other two blind models, BLINDS-II is significantly slower, on the account of DCT transformations done by the algorithm. It is therefore not suitable for real-time streaming images. However, unlike the other two blind models, it provides information about DCT coefficients, which is ignored by faster algorithms. Table 5.6 and Figure 5.9 signifies overall good performance of BLIINDS-II for all distortion types.

Since iteration is done over 30 times, the scatter plot in Figure 5.10 is a sample taken from one of those iterations (for arbitrarily chosen training ratio of 0.66), representing good correlation between the predicted DMOS and actual DMOS values. The most fruitful result of this study is a guaranteed lowest performance of no less than 89.5% (for high training ratio, Table 5.7, Minimum Correlation).

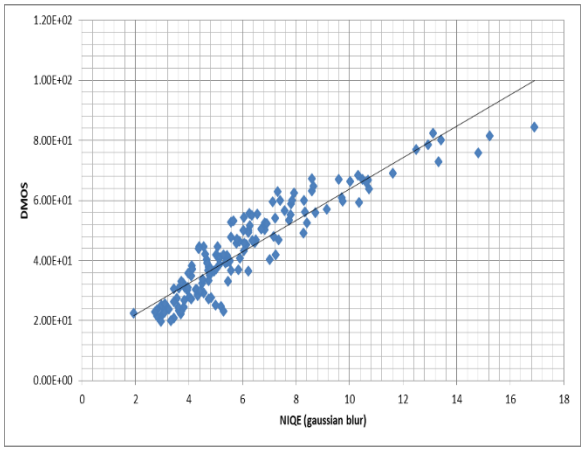
The average correlation coefficient of 91.6% is marginally better, 0.6%, than the best performing metric (BRISQUE) (Table 5.5), 5% better than BLIINDS-II (Table 5.6) and 37.6% better than NIQE (Table 5.4) metrics. What makes the combined model better than the other models is its requirement of very low fraction of samples for training to provide a consistent accuracy over many different training to testing ratios. Another advantage of the proposed NR-IQA model is that its performance is more or less same irrespective of the training ratio as long as the training ratio is 0.5 or above.

Table 5.4. NIQE

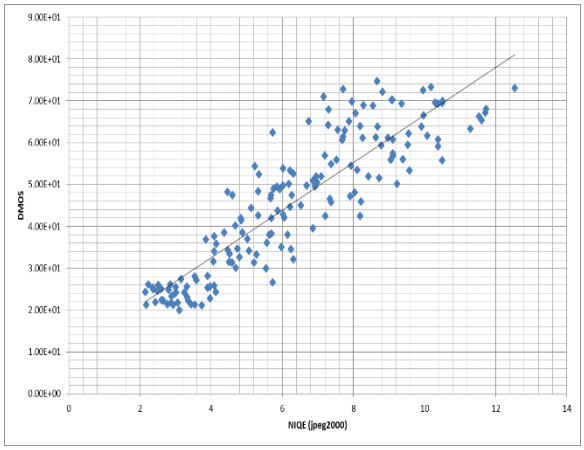
Distortion	FF	GB	J2	J	WN	ALL
Average DMOS Correlation	0.868	0.923	0.890	0.725	0.858	0.540



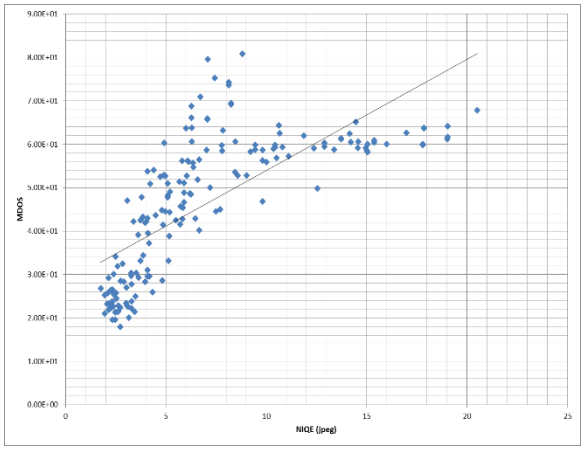
(a) NIQE for FF distortion



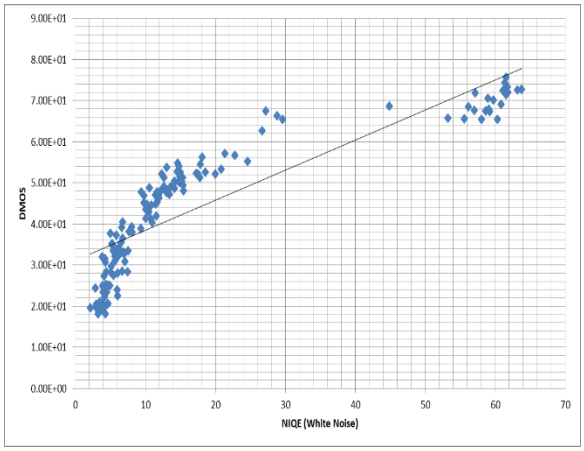
(b) NIQE for GB distortion



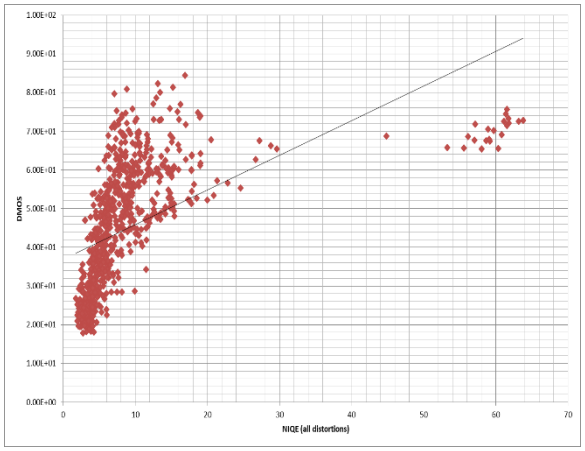
(c) NIQE for J2 distortion



(d) NIQE for J distortion



(e) NIQE for WN distortion

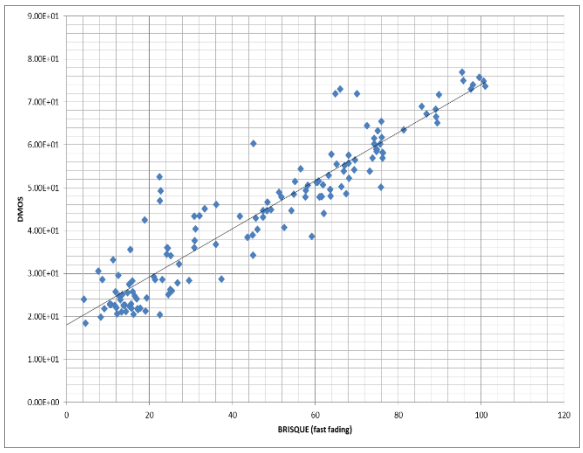


(f) NIQE for ALL distortions

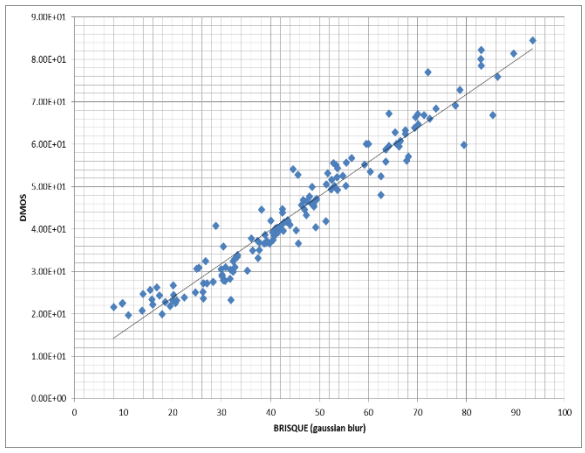
Fig. 5.7. The scatter plot for NIQE and DMOS for different distortions

Table 5.5. BRISQUE

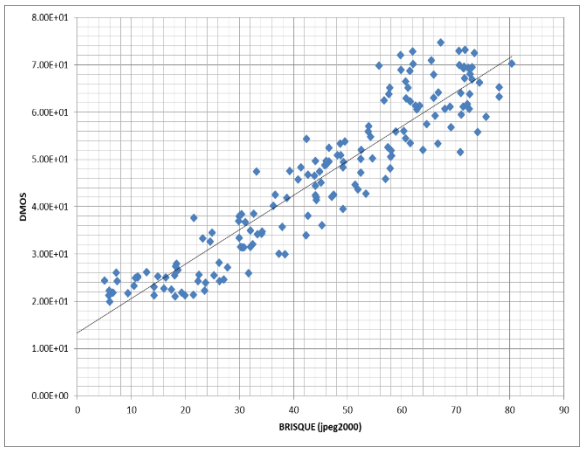
Distortion	FF	GB	J2	J	WN	ALL
Average DMOS Correlation	0.935	0.968	0.931	0.897	0.991	0.910



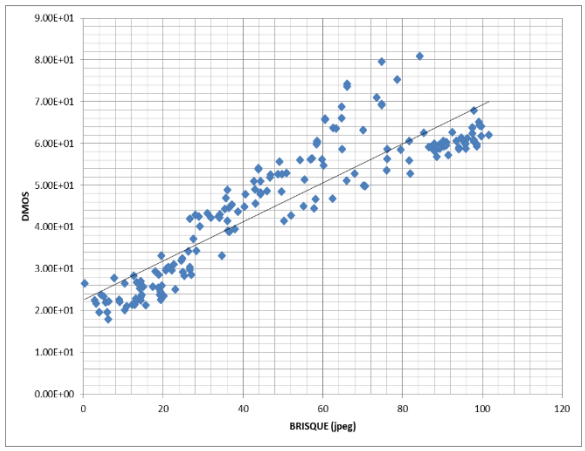
(a) BRISQUE for FF distortion



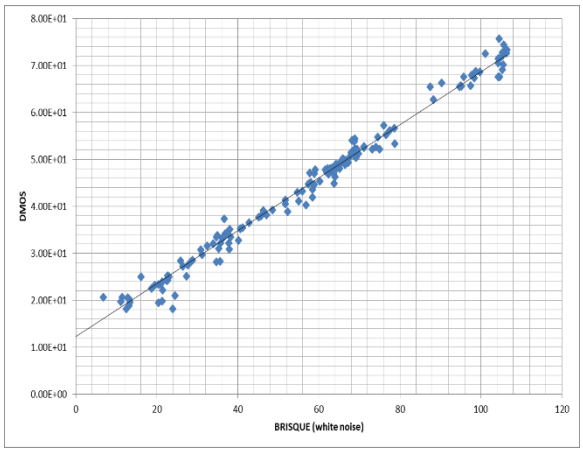
(b) BRISQUE for GB distortion



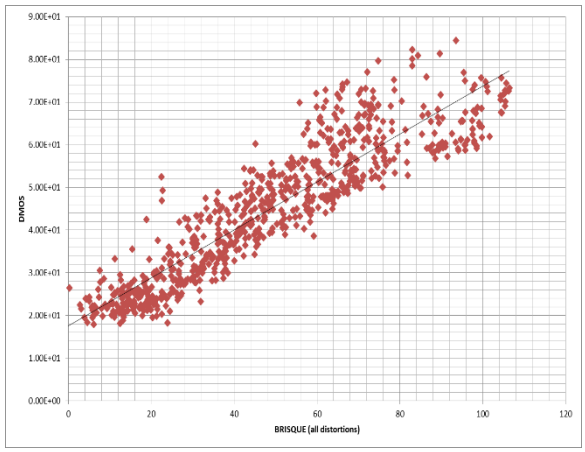
(c) BRISQUE for J2 distortion



(d) BRISQUE for J distortion



(e) BRISQUE for WN distortion

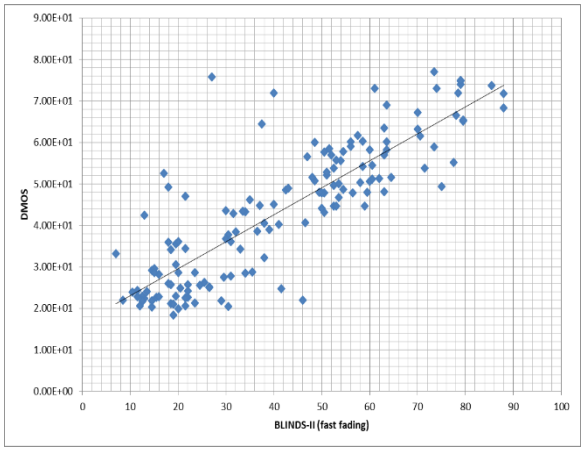


(f) BRISQUE for ALL distortions

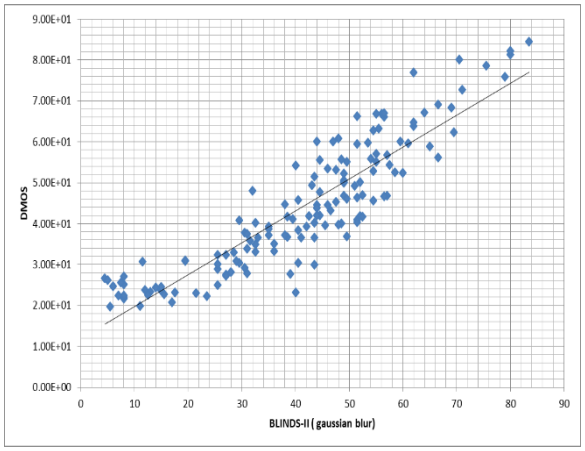
Fig. 5.8. The scatter plot for BRISQUE and DMOS for different distortions

Table 5.6. BLIINDS-II

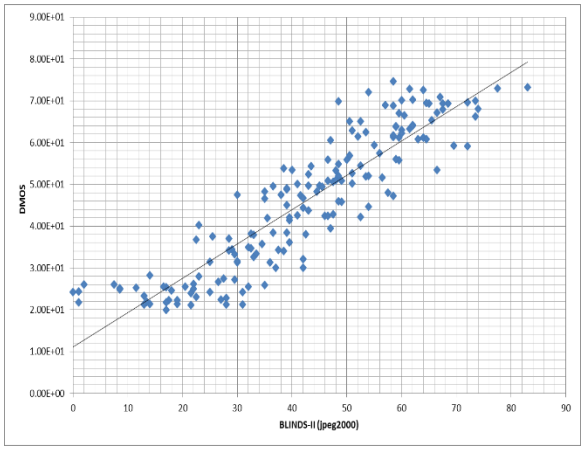
Distortion	FF	GB	J2	J	WN	ALL
Average DMOS Correlation	0.844	0.899	0.902	0.898	0.965	0.866



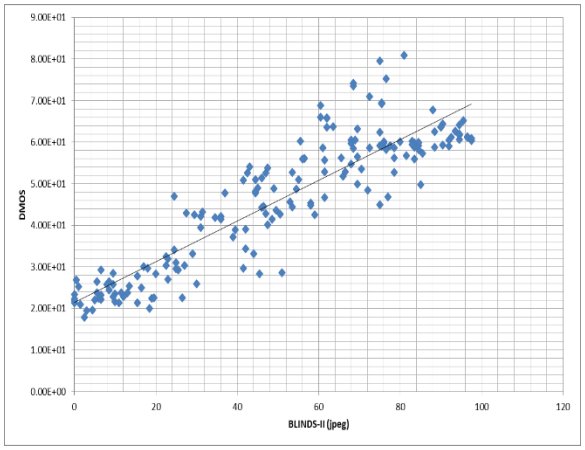
(a) BLIINDS-II for FF distortion



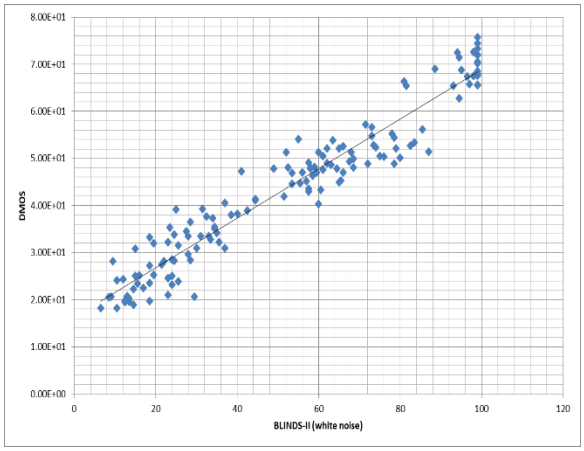
(b) BLIINDS-II for GB distortion



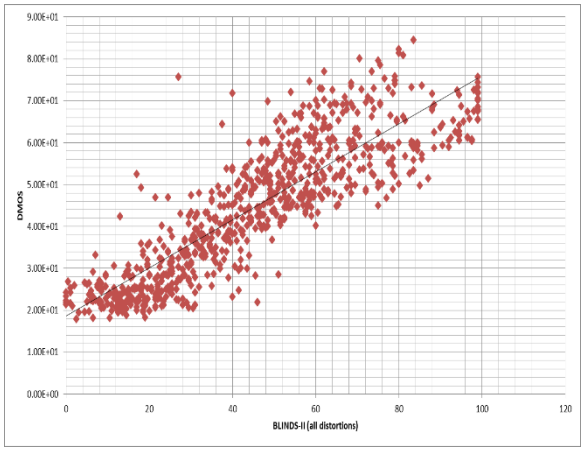
(c) BLIINDS-II for J2 distortion



(d) BLIINDS-II for J distortion



(e) BLIINDS-II for WN distortion



(f) BLIINDS-II for ALL distortions

Fig. 5.9. The scatter plot for BLIINDS-II and DMOS for different distortions

Table 5.7. Proposed NR-IQA Model (For ALL distortion types)

Training ratio	0.5	0.66	0.75	0.8
Maximum Correlation	0.9261	0.9291	0.9307	0.9351
Minimum Correlation	0.9052	0.9018	0.9007	0.8956
Average Correlation	0.9160	0.9164	0.9165	0.9160
Standard Deviation	0.0048	0.0055	0.0063	0.0075

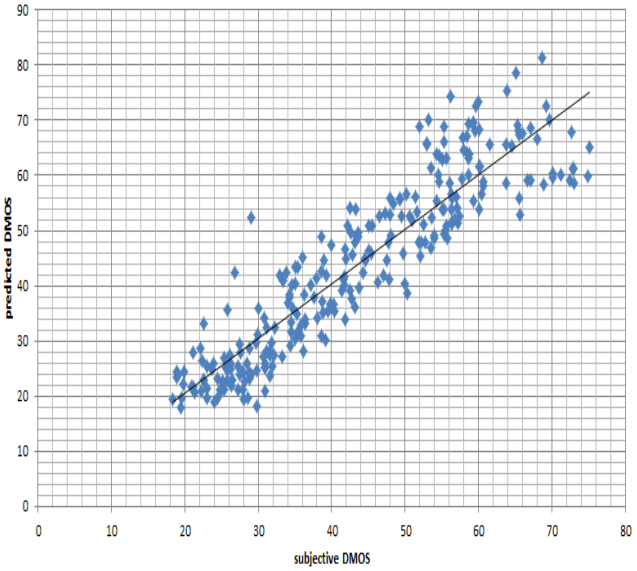


Fig. 5.10. Scatter plot for Subjective and Predicted DMOS for one iteration

The marginal improvement of the proposed model with respect to BRISQUE NR-IQA metric and very high improvement for NIQE is attributed to nonlinear nature of the scatter plots given in Figure 5.7(a), 5.7(d), 5.7(e) and 5.7(f) and the wide dispersion of the scatter plot shown in Figure 7.8(b). In all of these plots, a straight line has been fit while there would have been a piecewise linear curve fit for these scatter plots.

5.5.2 No-Reference Video Quality Assessment

Six metrics (Blocking4, Blocking8, Clipping, Clipping Standard Deviation, Contrast and Contrast Standard Deviation) are fed to the NN model with six nodes in the input layer and 15 nodes in the hidden layer. 70% data is used for training and 15% data is used for validation and testing each. Figure 5.11 represents the corresponding graphs for a trial got from running the NN model for the proposed NR-VQA model. The regression lots represent the relationship between Output *(Y)* (NN output) and the Target *(T)* value (desired output). The dashed line in the plots symbolize the desired outcome (target value) and the solid line symbolizes the best fit linear regression line between Output *(Y)* and Target *(T)*. For goodness of fit value R, if R=1, an exact linear relationship exists between Output *(Y)* and Target *(T)*. If R=0, no linear relationship exists between Output *(Y)* and Target *(T)*. The plots are shown for goodness of fit for training, validation and testing stage. It reports R=0.9639 for Training, R=0.7672 for validation, R=0.3383 for testing, and R=0.8785 for all samples of dataset.

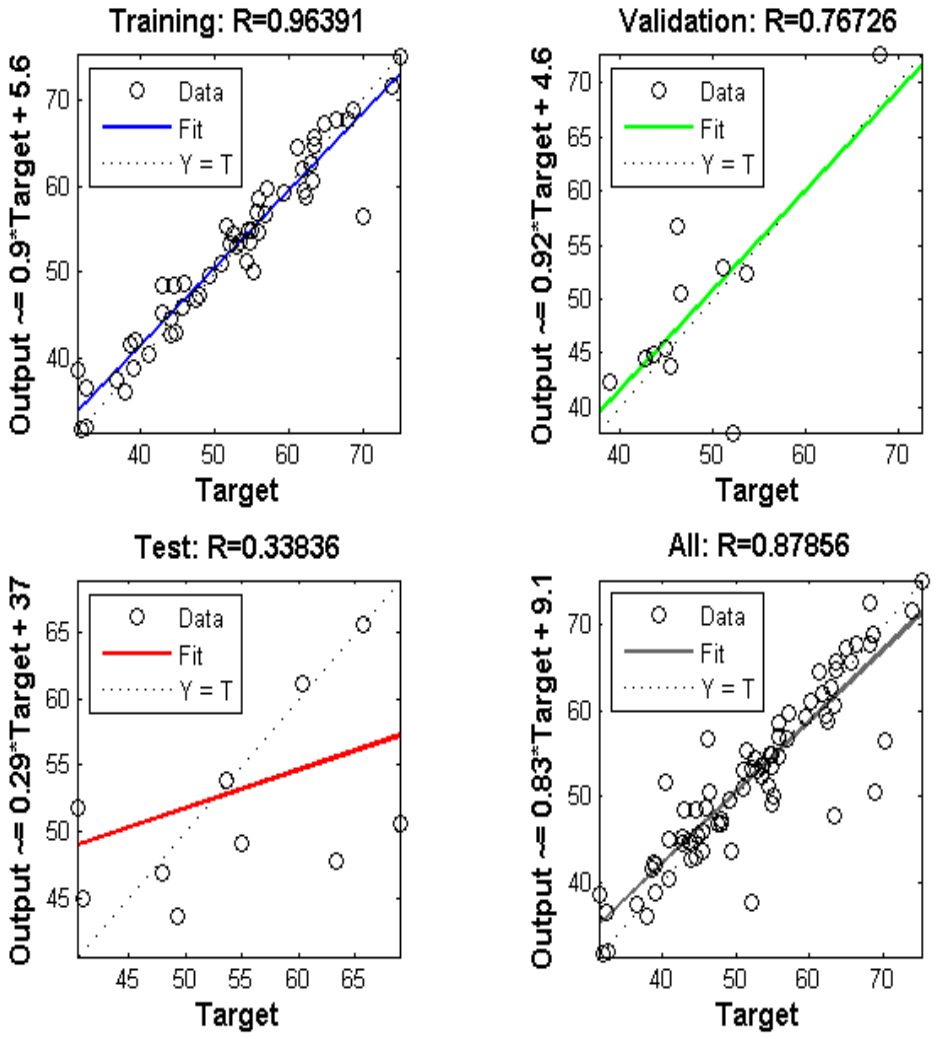


Fig. 5.11. Results for a single training test case with 15 nodes in the hidden layer of the NN model. 70% data is used for training and 15% for testing and validation each

The NN goodness of fit R=0.8785 is surprisingly a good accuracy taking into account the limited number of features available to the model. An overall R of 0.8785 is a very promising figure. Various other metrics for video assessment struggle to perform uniformly for all types of distortions and probably the database too suffers from loss of generality.

5.6 Result Analysis and Interpretation

No-Reference quality assessment technique requires only the test image to calculate various quality metrics to assess the image quality. Also, as video essentially comprises of image frames with additional temporal dimension, video quality assessment requires a thorough understanding of image quality assessment metrics and models. Therefore, in order to identify features that deteriorate an image or video

quality, a fundamental analysis of spatial and temporal artifacts has to be done. In this chapter, an NR-IQA model is proposed by combining the efficiency of three existing NR-IQA metrics (NIQE, BRISQUE and BLIINDS-II) using MLR. The efficiency of the proposed NR-IQA model is compared with the three NR-IQA metrics in terms of DMOS correlation for five distortion types (FF, GB, J2, J, WN) individually and for all distortions in combination (Table 5.4, Table 5.5, Table 5.6, scatter plot for the same are represented in Figure 5.7, Figure 5.8 and Figure 5.9 respectively). The results show marginally better performance, i.e. 0.6%, of the proposed NR-IQA model against BRISQUE NR-IQA metric, the best performing metric among the three metrics (Table 5.5). Also, the proposed model for image quality assessment requires very low fraction of samples for training to provide consistent precision over different training to test ratios.

The chapter also discusses the proposed NR-VQA model designed by exploring the quantification of certain distortions like, ringing, frame difference, blocking, clipping and contrast in video frames. The performance of NR-VQA model is examined using a simple NN Model to attain high value of goodness of fit. Experiment is concluded with high accuracy in terms of goodness of fit, but requires a wider selection of database videos to check its generality. Improvements done to design the NR-VQA model includes addition of features to detect and quantify clipping, contrast etc. in video frames. Also, low fraction of samples required for training to provide a consistent accuracy over many different training to testing ratios makes the proposed model to execute better. Even though accuracy in attained in terms of goodness of fit R=0.8785, the generality of the model, however remains a drawback due to lack of availability of features that could fairly determine effects like ringing, freeze frame, etc. in a video frame.

5.7 Summary

In this chapter, an NR-IQA Model is designed by combining NIQE, BRISQUE and BLIINDS-II NR-IQA metrics using Multi-Linear Regression. Also, five feature quantification metrics, i.e., ringing effect quantization metric, frame difference metric, a multi-scaling blocking metric, metric for clipping and metric for contrast quantification are composed for video quality estimation. Next chapter presents an improved implementation of MSER feature detector to address some of the limitations among extensive computation, accurate view alignment and real time performance for designing an AR application.