

A No-Reference Method for Assessing the Quality of Multiply-Distorted Images

by

Pooryaa Cheraaqee

A thesis
presented to Kharazmi University
in fulfillment of the
thesis requirement for the degree of
Masters of Science
in
Computer Engineering-Artificial Intelligence and Robotics

Tehran, Tehran, Iran, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis.

External Examiner: Farah Torkamani Azar
Associate Professor,
Dept. of Telecommunication,
Shahid Beheshti University

Internal Examiner: Manoochehr Kelarestaghi
Assistant Professor,
Dept. of Electrical and Computer Engineering,
Kharazmi University

Supervisors: Azadeh Mansouri
Assistant Professor,
Dept. of Electrical and Computer Engineering,
Kharazmi University

Ahmad Mahmoudi-Aznaveh
Assistant Professor,
Cyberspace Research Institute,
Shahid Beheshti University

This document is prepared using the thesis template available on <https://www.overleaf.com/>.

Abstract

Assessing the visual quality of digital images is of great importance, since humans are the typical users of multimedia applications, and their satisfaction is dependent on the perceived quality. Mathematical models of human-quality perception are desirable, since they can automate the monitoring and optimization of such services. However, the non-linear and complex human visual system is difficult to simulate, and machine learning methods are leveraged to predict human judgements on image quality. The existence of multiple distortions increases the complexity of assessment in comparison to the case of single distortions. To supply academic and practical demands, several datasets are provided for training and evaluation of multiple-distortion quality metrics. In this dissertation, a full-reference and a no-reference method are proposed for assessing multiply-distorted images that incorporate gradient orientations for representing micro structures. Experiments show that the generality of the proposed features is superior to the existing descriptors. The method is also shown to have an acceptable time complexity while being competitively accurate.

Acknowledgements

I would like to thank Dr. Mansouri, Dr. Mahmoudi-Aznaveh, my father, my mother, and my brother, for making this research possible.

I also thank Mohammad Minouei, Hossein Motamednia, Mojtaba Barekati, and Mohammad Ali Arghavan for sharing their creative ideas with me.

Table of Contents

List of Tables	ix
List of Figures	x
1 Introduction	1
1.1 Classes of Objective IQA Problems	2
1.2 Distortions in an Image	2
1.3 Objectives and Motivations	3
2 Literature Review	5
2.1 FR Methods for Single Distortion IQA	6
2.1.1 Structural Similarity	6
2.1.2 Similarity of Gradient Magnitude	9
2.2 No-Reference IQA Methods	11
2.2.1 Feature-Based NR IQA Methods	11
2.2.2 Automatic Feature Extraction	15
2.3 FR methods for Multiple Distortions	15
2.3.1 Combination of Scores	16
2.3.2 Similarity of QWT Subbands	17
2.3.3 Measuring Mutual Information	19
2.4 RR Methods for Multiple Distortions	19

2.4.1	Combining Scores	20
2.4.2	Using the Internal Generative Mechanism	20
2.5	NR methods for Multiple Distortions	20
2.5.1	Five-Steps and Six-Steps Methods	20
2.5.2	Combining Quality Parameters	21
2.5.3	Dictionary-Based Approach	21
2.5.4	Learning to Measure Distortion Combinations	22
2.5.5	Convolutional Neural Networks	24
2.5.6	Using Local Binary Patterns	27
2.5.7	LBP for the MSCN Image	30
2.5.8	LBP for Color Information	33
2.5.9	A New Local Binary Pattern	33
2.5.10	Describing Structural Information with Various Features	33
2.6	Analysis and Conclusion	34
3	The Proposed Method	35
3.1	Gradient Direction Similarity	36
3.2	The NR Method	38
3.2.1	Extracting Direction-Based Features	38
3.2.2	Computing the Quality Score	41
3.3	Summary	42
4	Experimental Results	43
4.1	IQA Datasets	43
4.2	Evaluation Criteria	44
4.3	Evaluating the FR Method	45
4.4	Evaluating the NR Method	46
4.4.1	Performance on Individual Datasets	46

4.4.2	Impact of Training Data	46
4.4.3	Generalization Ability	47
4.4.4	Computational Complexity	47
4.5	Future Works and Conclusion	47
	References	57

List of Tables

4.1	Severity of Blur and JPEG for 15 Distorted Versions of a Reference Image in ‘blurjpeg’ Folder. Six of these 15 versions are singly-distorted (the severity of one distortion is zero) and 9 are multiply-distorted	49
4.2	Severity of Blur and Noise for 15 Distorted Versions of a Reference Image in ‘blurnoise’ Folder.	50
4.3	The Accuracy of the Proposed Method and Other FR Algorithms	51
4.4	The Algorithms’ Run-Time	52
4.5	The Performance of NR Methods on Individual Datasets (MLIVE, MDID2013, and MDID2016).	53
4.6	The Results for Changing the Ratio of Training Data.	54
4.7	SROCC Indexes for Cross-Dataset Evaluation.	55
4.8	Methods’ Time Complexity	56

List of Figures

2.1	<i>MSE</i> and <i>SSIM</i> for different distorted versions of Einstein’s photo.	7
2.2	Different versions of an image with different distortions along with the gradient magnitude map and the map of gradient magnitude similarity with the original image.	10
2.3	Training feature-based methods	12
2.4	Testing feature-based methods (The ‘Regressor’ is the output of the train process.)	12
2.5	Two images from LIVE dataset and their MSCN map and the histogram of values in the MSCN domain	13
2.6	MSCN distributions for distorted versions of two images	14
2.7	The QWT subbands of ‘bikes’ in one scale	18
2.8	The procedure of building a dictionary	23
2.9	Learning the distortion parameters	25
2.10	The architecture of Kang’s network	26
2.11	The architecture of EONSS	27
2.12	An image from MDID13 dataset along with its LBP map variants	27
2.13	The nine possible binary patterns with 2 or less transitions	28
2.14	Images of different quality (left column) with their gradient magnitude map (middle column) and the LBP computed from gradient magnitude (right column)	29
2.15	Images of different quality (left column) with their simple histogram (middle column) and their weighted histogram (right column)	31

2.16 An image with its $LBP_{8,1}^{riu2}$ (c) and $GCSLBP_{8,1}$ (d) maps; extracted from the MSCN values (b)	32
3.1 Magnitude and direction of gradient vectors for an image in (b) and (c), respectively.	36
3.2 Sensitivity of gradient direction to various artifacts	37
3.3 Two alternatives for representing gradient directions	37
3.4 Histogram of direction values after quantization	39
3.5 Weighted histogram of direction values	39
3.6 Gradient magnitude and direction for (a) in (b) and (d), respectively. Derivative of gradient direction in (c)	40
3.7 Difference of $LBP_{P,R}^{riu2}$ and $MLBP_{P,R}^{riu2}$. (a) in $LBP_{P,R}^{riu2}$ the difference of n_c and $P n_i$ s is computed, where n_i s are located in a neighborhood of radius R and arcs with angles $\theta = \frac{360}{P}$ (b) a neighborhood of size 3×3 is shown around each n_i . The ‘median’ of each neighborhood is considered for subtracting from n_c	40
3.8 Proposed features for different qualities of two scenes	41

Chapter 1

Introduction

Digital images are an important means for sharing information in the age of communication. Watching television or viewing images on smartphones are part of our daily lives and video accounts for a large portion data traffic [1]. Digital images are captured, synthesized, and shared for educational, recreational, or business purposes.

It is unpleasant for users to view a low-quality image and it is necessary in many situations to be provided with a measurement of image quality. For example, content providers are willing to be aware of the quality of images viewed by their users and detect the defected pictures. The process of image acquisition can also repeat, until a satisfactory quality is achieved. Compression algorithms try to reduce the size of images, while maintaining its quality. Reduction of size is obtained with loss of information which is followed by image distortion. Therefore, with measuring the quality of the compressed image, the trade off between size and quality can be optimized.

The straight-forward way to evaluate the quality of an image, is to survey human opinions [23]. A set of subjects view the image and each of them gives a score to the quality of the image. The average of these scores, called *mean opinion score*-MOS, is a reliable measurement of the visual quality of the test picture.

The process described above, is called subjective assessment and it is infeasible for online, large-scale or real-time applications. Hence, automatic methods are desirable that can predict human's judgement on the quality of an image. However, due to its complexity, human visual system (HVS) is difficult to model and the prediction of an algorithm may not be well correlated with the opinion of human.

Devising computational models that can correctly predict human opinion on the quality of a digital picture, is called objective image quality assessment (IQA). The accuracy and

speed of this prediction are the performance criteria of a proposed algorithm. The accuracy of a method is measured by the correlation between its scores and the scores obtained from subjective assessments.

In this thesis, the problem of objective IQA is studied. In the following sections a classification of the problem and an introduction on multiple distortions is presented.

1.1 Classes of Objective IQA Problems

According to the availability of a pristine image, there can be three cases when evaluating the quality of a distorted picture. In some scenarios, like image compression, a distorted version of the original image is produced by the compression algorithm. So when assessing the quality of the compressed image, the original can be used as a reference. This case is called *full reference-FR* IQA. If instead of pixel-by-pixel values, there are some information available from the reference image, like an extracted feature vector, the assessment is called *reduced reference-RR* IQA. This case can happen in image communications when it is expensive and unreliable to transmit bulky images.

An example for the case that there are no reference images, is a hand-held camera that captures multiple shots at a time and automatically selects the one with highest visual quality. There are no references available from the recorded scene and the evaluation must be done in a *no reference-NR* manner. It is obvious that correct prediction in NR IQA is more difficult [39].

Although IQA methods can be classified in different ways, but we will have the least overlap when classifying according to the availability of a reference picture. However, there are methods that can be adapted to a desirable amount of information from a reference signal [6, 82]. In this research, a FR and a NR objective IQA methods are proposed and evaluated.

1.2 Distortions in an Image

Common distortions that cause deviation in the perceived quality of images, are limited in practice [8]. Blocking effect, out-of-focus blur, and white Gaussian noise are examples of these artifacts. If we are already aware of the distortion type that contaminated the image and we want to measure its severity, we are performing a *distortion-specific* assessment.

General-purpose or *non-distortion-specific* assessment, addresses the case that the image can be distorted with an arbitrary artifact and the algorithm is not aware of its type.

There has been numerous studies for distortion-specific and general-purpose IQA [47, 15, 94, 56, 5] and many datasets of distorted images are available along with their corresponding subjective scores [75, 7, 69, 70, 35]. Each distorted image in these datasets is contaminated with one type of distortions of arbitrary severity, i.e., the image is *singly-distorted*.

In real-world problems, an image can have multiple distortions. An example is an image that is blurred because of environmental conditions at the time of acquisition, and then quantized by compression, and noised after transmission. In 2012, Jayaraman et al. [37], provided a dataset of subject-rated *multiply-distorted* images and demonstrated that the accuracy of methods devised for single distortions, drops significantly when applied to multiply-distorted images. Therefore, different approaches are required for multiple distortions [46].

With the increase in academic and commercial demands, two other datasets are constructed for training and testing the algorithms on multiply-distorted images [79, 29] and many current IQA studies are devoted to multiple distortions [46, 29, 17, 14, 13, 43, 55, 54, 28, 60, 87, 109, 51]. IQA of multiply-distorted images is the problem that is tackled in this thesis.

1.3 Objectives and Motivations

Many image processing systems lack a reliable and fast metric of image quality. Parameters related to contrast and luminance of hand-held digital cameras can be automatically set for any arbitrary scene according to a measure of quality [2]. Compression algorithms can also optimize the quantization using the metric [100, 101]. Another example is video streaming services in which, streaming resources can be allocated in a smarter manner according to the quality of the videos [3]. Image recommendation systems can also sort the images based on a quality index [22].

Subjective assessments must be proceeded according to certain standards and comply certain conditions [86]. For example, the number of subjects must satisfy the requirements for statistical reliability. Also, the display condition and the rating format affect the opinion of the participants. Regardless of the tediousness of subjective experiments, it is not possible to survey human judgements in on-line and real-time applications and

automation is the only option in these cases. So, a fast and reliable image quality metric is of great practical interest.

Occurrence of multiple distortions in real-world problems demands methods that can handle the mutual and joint effect of the artifacts. On the other hand, traditional methods are greatly challenged in assessing the quality of multiply-distorted images.

These theoretical and practical necessities motivates us to study the objective assessment of multiple distortions. Since the eventual observer in the majority of image processing applications is the human [23], the intelligence of an objective IQA system is measured by its ability to mimic human rating and we aim to propose a metric with an acceptable performance on subject-rated multiply-distorted IQA datasets.

In this chapter, we introduced and classified the problem of objective IQA, mentioned its applications, and distinguished the cases of single and multiple distortions. In the rest of the thesis we review the literature in Chapter 2 and propose a FR and a NR method in Chapter 3. In Chapter 4 the method is tested and the thesis is concluded with suggestions for future studies.

Chapter 2

Literature Review

For a long time (until 2002), mean squared error (MSE) and peak signal to noise ratio (PSNR) were the common criteria for comparing the quality of a distorted image to a reference [91]. Although these metrics are intuitively explainable and computationally efficient, they are not well-correlated with human opinions [89, 90]. Structural similarity was proposed to cover this gap [88]. With the introduction of this criterion, many methods were proposed to measure the structural similarity, which a large number of them are based on image edges [97].

For referenceless assessment, natural scene statistics (NSS) were the criterion of choice for many metrics. Ruderman [72] showed that natural images¹ obey statistical regularities that are stable, regardless of the image content. So deviation from these regularities were exploited as a measure for distortion.

With the introduction of multiple distortion datasets, methods were proposed that tried to consider the simultaneous existence of multiple distortions in an image and extract features that describe their severity.

In this chapter we introduce the structural similarity criterion and two methods for its measurement. Then we review the approaches taken for NR IQA. The FR, RR, and NR algorithms proposed for IQA of multiply-distorted images are explained and analyzed in subsequent sections.

¹“An image recorded from the physical environment with a camera that is sensitive to the visible spectrum of electromagnetic wave” [49]

2.1 FR Methods for Single Distortion IQA

Numerous researches in IQA were conducted for full reference assessment [47]. Assuming a reference image, r , and its distorted version, d ; computing the difference between r and d maybe the first approach that comes to our minds for comparing them. This is what MSE, PSNR, or RMSE do. For two grayscale images, r and d :

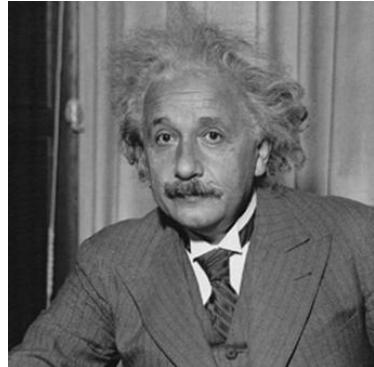
$$MSE(r, d) = \frac{1}{M \times N} \sum_{x=1}^M \sum_{y=1}^N (r(x, y) - d(x, y))^2 \quad (2.1)$$

Where M and N are image dimensions. For two identical images, MSE will be zero and this is in accordance with our opinion: “*the quality perceived from two identical images is the same*”, but there are many cases that MSE is independent of image distortions; Fig. 2.1 shows an example (figure from [91]). We see different versions of the reference image in Fig. 2.1a with four artifacts (the image in Fig. 2.1b is the reference, only magnified). MSE for the identical images equals zero (Fig. 2.1b), but in spite of the different visual quality of images in Fig. 2.1c- 2.1f, MSE is the same for all them, while a human observer prefers the image in Fig. 2.1c to the image in Fig. 2.1f.

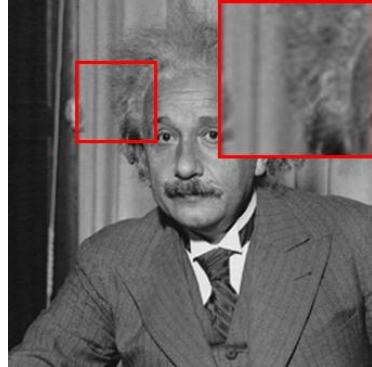
MSE and similar approaches are based on “error visibility”. Wang and Bovik [88] showed that the semantic information which we grasp from images are based on image structures. With a degradation in image structures we do not understand what we expect from a scene and perceive this as a low quality. They proposed a method, called “SSIM [91]”, for measuring the structural similarity between two images. With the popularity of this method, researchers proposed other metrics for structural similarity and, as mentioned before, many of them are based on image edges. In what follows SSIM and one of the gradient based methods are explained.

2.1.1 Structural Similarity

SSIM assumes that the structural similarity of two images is dependent on three factors: the correlation between the images, the luminance similarity, and the contrast similarity of the two image. That is, the more similar the two images are according to these criteria, the better is the quality of the distorted image. The abstract definition of SSIM for two



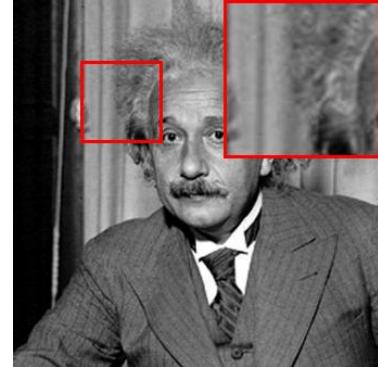
(a) Einstein's photo



(b) original image

$MSE = 0$

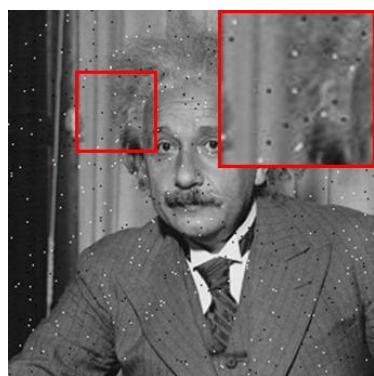
$SSIM = 1$



(c) contrast change

$MSE = 144$

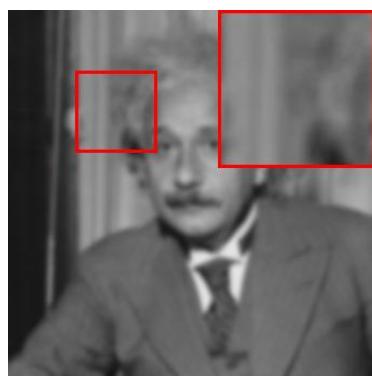
$SSIM = 0.913$



(d) salt & pepper noise

$MSE = 144$

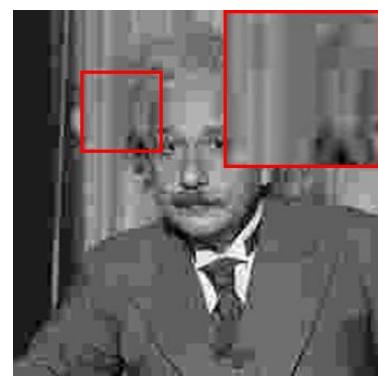
$SSIM = 0.840$



(e) blur

$MSE = 144$

$SSIM = 0.694$



(f) Jpeg quantization

$MSE = 144$

$SSIM = 0.662$

Figure 2.1: MSE and $SSIM$ for different distorted versions of Einstein's photo.

grayscale images, r and d , will be as follows:

$$\begin{aligned} SSIM(r, d) = \\ pixel_wise_correlation \times \\ luminance_similarity \times \\ contrast_similarity \end{aligned}$$

In SSIM, the contrast of an image is modeled with the variation in intensity values of pixels and luminance is modeled with the average of the values. If I is an image with N pixels, its luminance and contrast are represented as μ_I and σ_I , respectively, and obtained via the statistical definitions:

$$\mu_I = \frac{1}{N} \sum_{x=1}^N I(x) \quad (2.2)$$

$$\sigma_I = \sqrt{\frac{1}{N-1} \sum_{x=1}^N (I(x) - \mu_I)^2} \quad (2.3)$$

Where, $I(x)$ is the value of the x^{th} pixel of I . For obtaining the similarity of two numbers, authors used a similarity relation. The similarity of two numbers, a and b , are represented as $sim(a, b, c)$, and defined as:

$$sim(a, b, c) = \frac{2 \times a \times b + c}{a^2 + b^2 + c} \quad (2.4)$$

Where c is a constant for avoiding the instability of the fraction. Therefore, contrast and luminance similarity for r and d is obtained from 2.5 and 2.6.

$$contrast_similarity(r, d) = sim(\sigma_r, \sigma_d, c) \in [0, 1] \quad (2.5)$$

$$luminance_similarity(r, d) = sim(\mu_r, \mu_d, c) \in [0, 1] \quad (2.6)$$

The correlation between two images are defined as the statistical correlation of the intensity values of the pictures. If $r(x)$ and $d(x)$ are the x^{th} pixel from the N pixels in r and d , the correlation between r and d is obtained from the following equation:

$$pixel_wise_correlation(r, d) = \frac{\frac{1}{N-1} \sum_{x=1}^N (r(x) - \mu_r)(d(x) - \mu_d)}{\sigma_r \times \sigma_d} \in [-1, 1] \quad (2.7)$$

For two identical images, SSIM will be equal to one. It is expected that it decreases with the deviation of the image. In Fig. 2.1 we see that this occurs and SSIM distinguishes different perceived qualities. It must be mentioned that SSIM is computed for each corresponding 8×8 window in the images and for pooling the scores of the windows into one score for the entire image, the SSIM values are averaged.

2.1.2 Similarity of Gradient Magnitude

Apart from SSIM, there are other metrics proposed for measuring the structural similarity. One of the approaches, is considering the edges of the image as a representative of image structures. Image gradients are known as good estimations of the edges [24]. In location (x, y) of image I , the gradient vector $\vec{G}_I(x, y)$ has a horizontal and a vertical component, g_h and g_v , respectively. The magnitude and direction of \vec{G}_I in (x, y) are defined in relations 2.8 and 2.9 with respect to these components.

$$GM_I(x, y) = \sqrt{g_v^2 + g_h^2} \quad (2.8)$$

$$GD_I(x, y) = \tan^{-1}\left(\frac{g_v}{g_h}\right) \quad (2.9)$$

The method GMSD [97] used GM for measuring image distortions. In Fig. 2.2 a sample is illustrated with its distorted versions and their gradient magnitudes. It is observable that image edges are sensitive to distortions. Gradient magnitude similarity map for r and d is called $GMS(r, d)$ and defined as:

$$GMS(r, d)_{(x,y)} = sim(GM_r(x, y), GM_d(x, y), c) \quad (2.10)$$

It is seen in Fig. 2.2c that GMS of an image and itself, equals to one for all pixels and when the image is degraded, this value will decrease (images in the right column).

GMSD computes the standard deviation for pooling the values of GMS. If $GMS(r, d)$ has N pixels and $GMS(r, d)_x$ is the x^{th} pixel, $GMSD(r, d)$ is obtained from the following equation:

$$GMSD(r, d) = \sqrt{\frac{1}{N-1} \sum_{x=1}^N \left(GMS(r, d)_x - \frac{1}{N} \sum_{x=1}^N GMS(r, d)_x \right)^2} \quad (2.11)$$

GMSD could improve the performance of SSIM, meanwhile, maintain a low computational complexity.

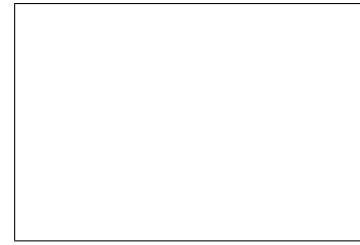
Considering image entropy [76, 77], different pooling strategies [103], using image singular values [57, 58], analysis in wavelet domain [9], and taking color into account [104], have been other innovations for FR IQA of singly distorted images.



(a) original image



(b) gradient magnitude of the original image



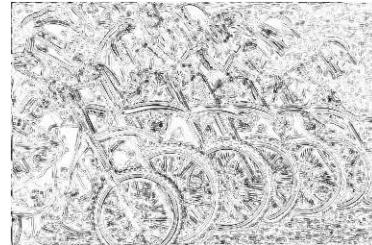
(c) gradient magnitude similarity of the original image



(d) blurred image



(e) gradient magnitude of the blurred image



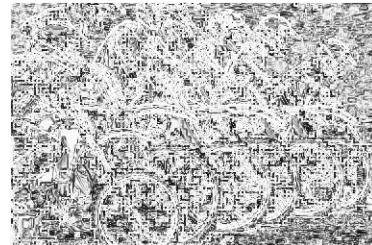
(f) gradient magnitude similarity of the blurred image



(g) image with Jpeg artifact



(h) gradient magnitude of the Jpeg-compressed image



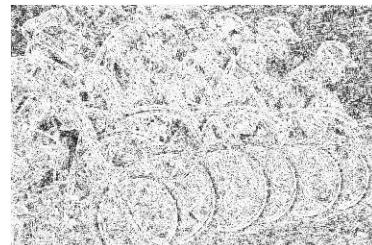
(i) gradient magnitude similarity of the Jpeg image



(j) noisy image



(k) gradient magnitude of the noisy image



(l) gradient magnitude similarity of the noisy image

Figure 2.2: Different versions of an image with different distortions along with the gradient magnitude map and the map of gradient magnitude similarity with the original image.

2.2 No-Reference IQA Methods

The common approach of FR metrics is to define the quality score using mathematical relations [47]. In the NR scenario, the distorted image is usually described with a feature vector and a model maps the vector to a quality score. This model can be a mathematical relation or be obtained using machine learning. The challenge for many NR IQA algorithms is to devise features that are expressive and fast to compute. Another strategy for feature extraction, is the use of machine learning techniques, such as *convolutional neural networks-CNNs*. With CNNs, the feature vector can be achieved automatically. In what follows, the framework of *feature-based* methods is presented, then the methods that are based on automatic feature extraction are reviewed.

2.2.1 Feature-Based NR IQA Methods

A common scheme for NR assessment, is extracting quality-relevant features. Then, with the use of machine learning methods, such as *support vector regression-SVR*, the features are mapped to quality scores. The required training data for machine learning can be obtained from the subjective scores in the image quality databases. The performance of these methods is dependent on the ability of the features to express the quality aspects of the image. The frameworks for training and testing of these methods is illustrated in Fig. 2.3 and Fig. 2.4.

If a dataset contains N pictures and their corresponding human scores, it can be considered as a set of ordered pairs like (I_i, MOS_i) , where I_i is the i^{th} image and MOS_i is the subjective score of I_i for $i \in \{1, 2, \dots, N\}$. These ordered pairs are the ‘samples’. If a feature vector is extracted from I_i it will be named: \vec{f}_i . If L samples of the dataset are selected for train, and, M samples for test, a model can be trained on L samples (Fig. 2.3) and tested on the other M samples (Fig. 2.4). It must be noted that $L + M = N$. The more the scores of the model are correlated with MOS s, the better is the performance of the method. The learning model in this framework can be a SVR [85] or a neural network. Although this model is effective in method’s performance, but the main contribution to accuracy and speed, is made by the feature extraction mechanism.

Natural scene statistics is a common criterion for extracting features. Consider the two pictures in Fig. 2.5 of ‘bikes’ and ‘monarch’ butterfly. If these images are transformed to MSCN² domain, their frequency histogram will obey the normal distribution (Fig. 2.5c)

²Calculation of the *Mean-Subtracted & Contrast-Normalized* image will be explained in the following paragraphs.

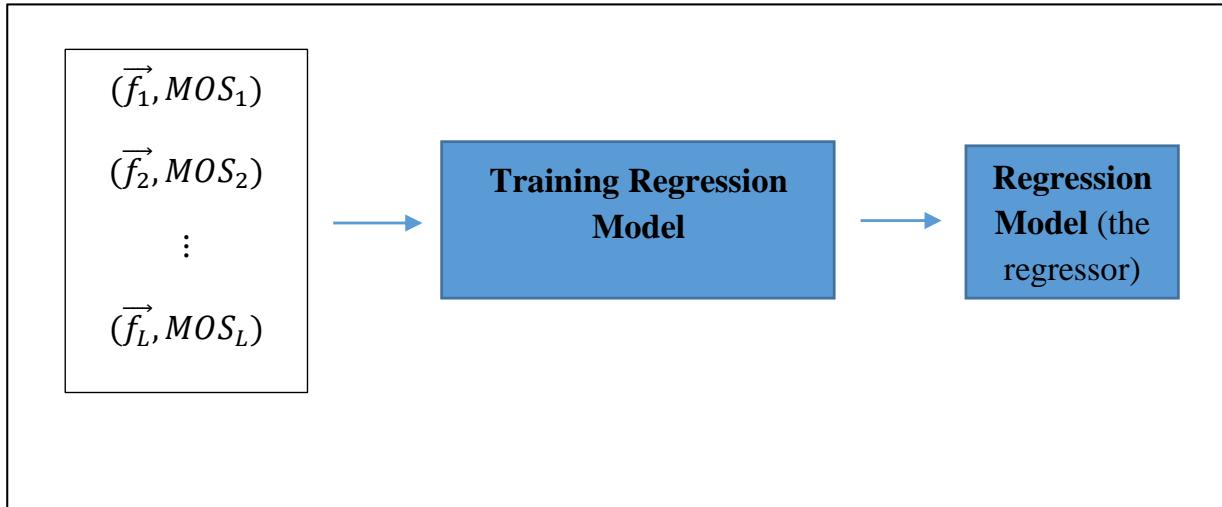


Figure 2.3: Training feature-based methods

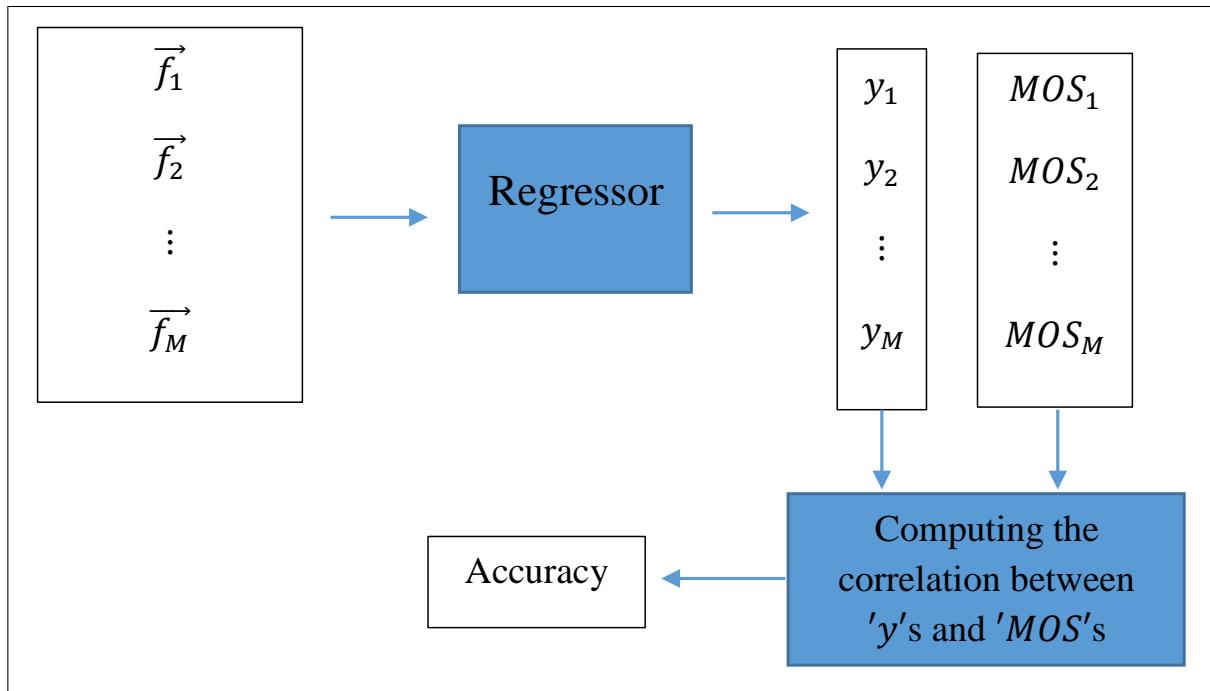


Figure 2.4: Testing feature-based methods (The ‘Regressor’ is the output of the train process.)

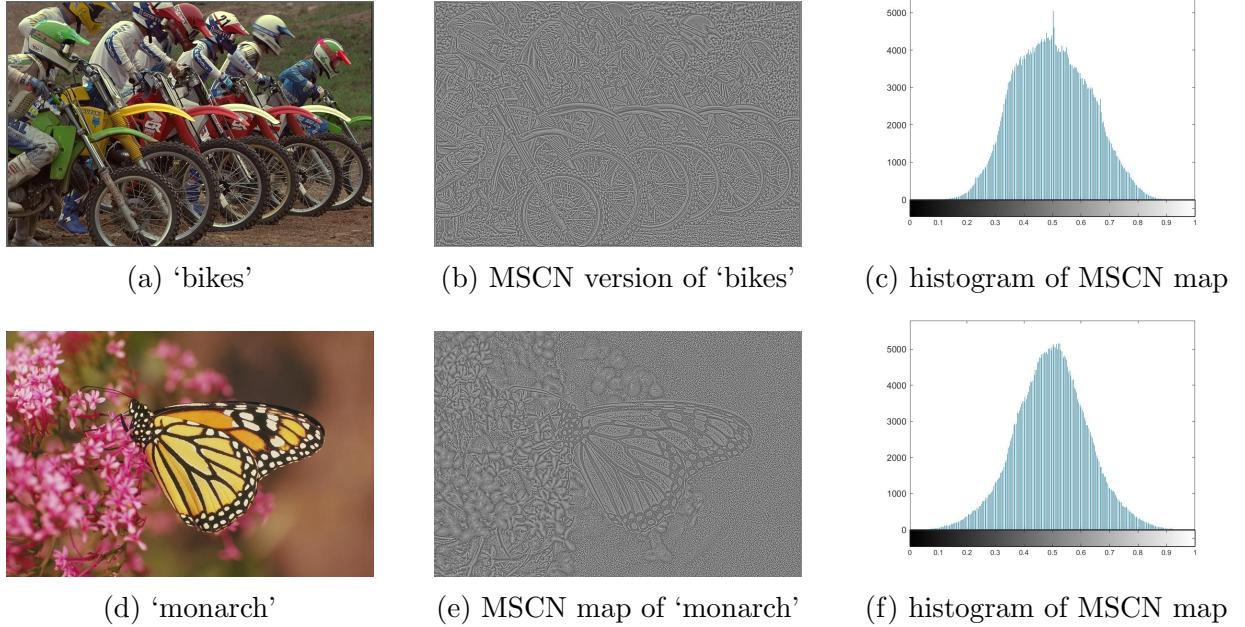


Figure 2.5: Two images from LIVE dataset and their MSCN map and the histogram of values in the MSCN domain

and Fig. 2.5f). Ruderman [72] demonstrated that this distribution will be normal for any natural image, regardless of its content. If an image is distorted, it deviates from the natural state and it is expected that its distribution in the MSCN domain deviates from the normal distribution. In Fig 2.6 the histogram of MSCN maps are shown for the distorted versions of ‘bikes’ and ‘monarch’. It is seen that despite the different contents of the two images, the distributions change with respect to type, and intensity of the distortions. For example, the Jpeg histogram has the same shape for both scenes (Fig. 2.6j and Fig. 2.6k). Therefore, the MSCN distribution of a picture can be used to detect distortion. The MSCN value of the pixel at (x, y) of the grayscale image I is represented with $\hat{I}(x, y)$ and obtained from the following relation [62]:

$$\hat{I}(x, y) = \frac{I(x, y) - \mu}{\sigma} \quad (2.12)$$

Where μ and σ are the mean and variance in a 3×3 neighborhood of $I(x, y)$.

The distribution of MSCN values can also be fitted with a generalized Gaussian distribution (GGD) [74]. GGD has parameters, such as mean (m), variance (σ), and shape (α).

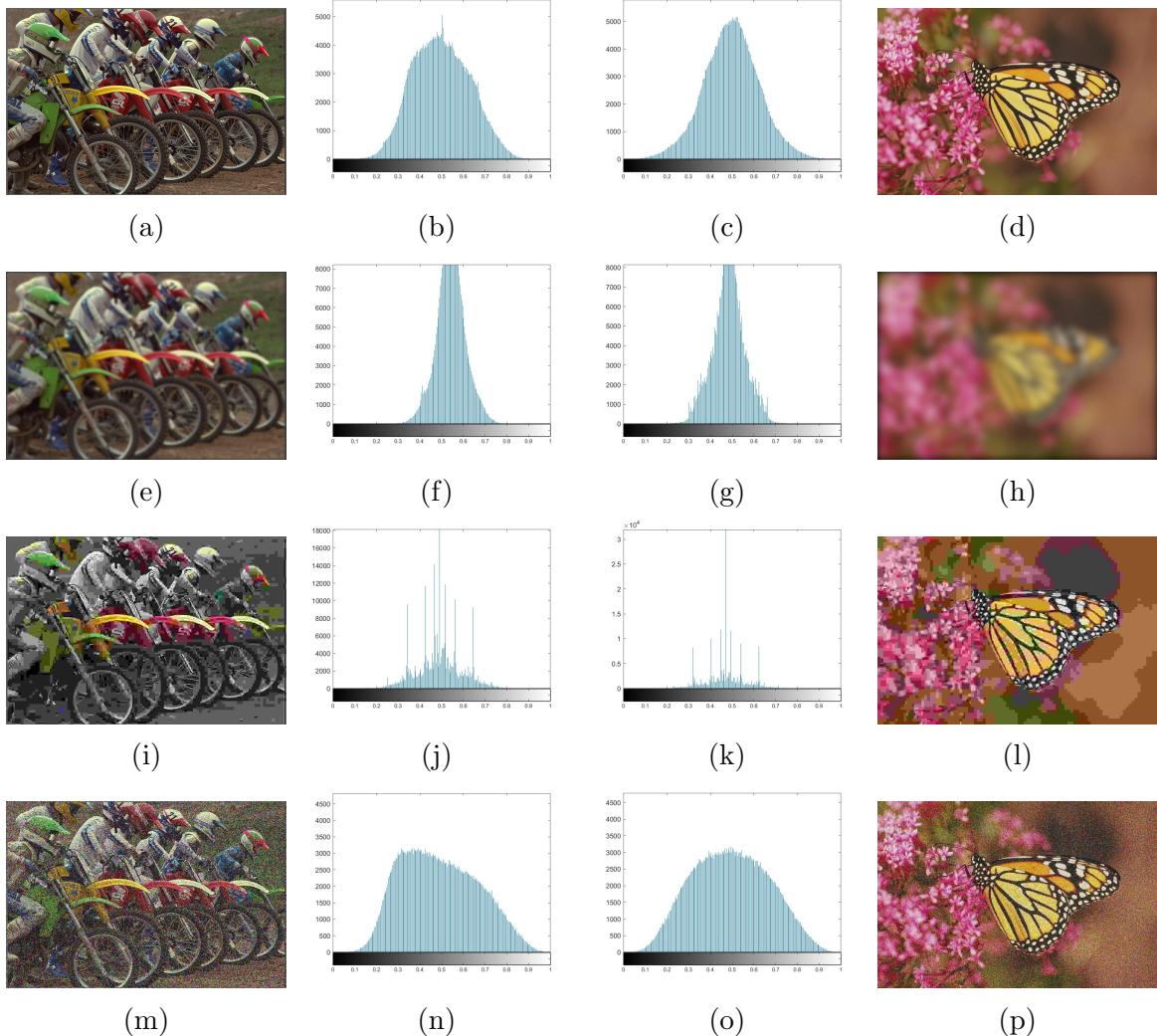


Figure 2.6: MSCN distributions for distorted versions of two images

The estimated parameters for an image, can be considered as the elements of a feature vector that describes the quality of the image.

In fact, this is what the method BRISQUE [62] does. Beside MSCN coefficients, NSS of other domains have been explored. BIQI [63] uses NSS in wavelet domain and BLIINDS-II [73] computes the features based on the entropy and average of coefficients of discrete cosine transform (DCT). Also, other distributions have been used for describing the histograms such as Gaussian scale mixtures (GSM) [30].

2.2.2 Automatic Feature Extraction

Convolutional neural networks can also extract feature vectors [4]. The common architecture of these networks pools multiple 2D-feature maps after many convolutional layers to a 1D-vector. Then, fully connected layers of neurons classify or regress the vector. Methods like CNN [39], MEON [52], DIQaM-NR [6], and Rank-IQA [50] are examples of such networks.

Using CNNs, the feature extraction process is embedded in the network and it is difficult to interpret the computed features. This is the same for dictionary-based methods [38]. In these methods, such as CORNIA [98], HOSA [93], QAF [102], SOM [105], and QAC [96], the distorted image is coded in a feature vector, based on a dictionary that contains the features of all good images. This dictionary is built by clustering the features of a set of good quality images.

Automatic features are shown to be expressive. In neural networks, these features are jointly optimized with the model that computes the scores, which increases the accuracy. However, the computational complexity of these methods is high and they demand large numbers of training samples. As mentioned earlier, these features are not interpretable and their relation with quality aspects of the image cannot be analyzed.

2.3 FR methods for Multiple Distortions

The existence of multiple distortions in an image, makes quality prediction difficult [8]. Even if we can measure the severity of each distortion correctly, their *joint effect* on the perceived quality will be difficult to predict, since their combination is not linear and depends on the type of distortions [25]. The accurate measurement of distortions' intensity in the presence of other distortions, is also controversial [48] and necessitates to consider

the *mutual effect* of the distortions on each other. As an example, assume adding a fixed amount of noise to two images; a sharp image and its blurry version. Normally the noise will be more sensible in the blurry version. Similarly, it is expected that blurring a noisy image must alleviate the effect of noise [48]. However, subjective experiments [40] show that this is true in some cases and completely wrong for many common samples.

These observations show the difficulty of assessing multiple distortions and reveal that accurate prediction of each combination or even permutation of distortions can be quite different. Thus, optimization for a specific permutation may result in poor performance for others. Some of methods for multiple distortions, tried to estimate or model the joint and mutual effect of the distortions. In the rest of the section, different methods for IQA of multiply-distorted images are reviewed and analyzed.

2.3.1 Combination of Scores

Combining the scores of four singly-distorted metrics, was the solution for FR IQA of multiple distortions, proposed by Chetouani [14]. In this method, four scores are computed for the distorted image with methods VIF [77], IFC [76], WASH [71], and SSIM [91]. They are then fed to an SVR as a 1×4 feature vector to be mapped to the quality score.

The four single distortion methods that are employed here, consider different criteria. SSIM measures structural similarity, VIF and IFC take images' entropy into account and WASH measures blurriness. With this combination, it has been tried to capture the effect of multiple distortions.

A similar approach was taken by Okarma in [67] with the difference that a nonlinear combination of the scores was used to pool the scores of the single metrics to the final score. The coefficients and powers of this nonlinear combination were optimized using the subjective scores of multi-distortion IQA datasets.

Due to the difficulty of modeling the joint effect, Chetouani used SVR in order to let machine learning to estimate the correct combination based on the individual metrics. However, when using a trainable model, its ability to generalize to samples outside the training dataset must be tested as well. Meanwhile, FR methods are normally based on mathematical relation and compute the quality directly.

Another issue is that the methods in [14] and [67] need to execute multiple independent metrics which is a huge computational load. This makes them very slow in comparison to the conventional FR metrics.

2.3.2 Similarity of QWT Subbands

Li et al. [43] used quaternion wavelet transform (QWT) [64] for extracting image structural information. At each scale, an image in QWT domain is decomposed into four subbands, LL , LH , HL , and HH . Each of these subbands has a magnitude, M , and three phases, θ , ϕ , and ψ (Fig. 2.7). According to Fig. 2.7, subbands of QWT have fine characteristics for detecting distortions. The magnitude of LL , is an approximation of the entire image and roughly shift invariant. Phases of LL , demonstrate the horizontal, vertical and diagonal structures. The magnitudes of LH , HL , and HH , extract the edges in different directions. In their method, called QWT-IQA, Li et al. compare the magnitude and phases of subbands of the reference and distorted image. Then they compute the quality scores of each subband, which are named: Q_{LL} , Q_{LH} , Q_{HL} , and Q_{HH} .

The magnitude and phases are denoted as in Fig. 2.7 and the corresponding terms in reference or distorted image, are distinguished with R and D subscripts. As an example, the computation of Q_{LL} will be explained. To compare LL_{M_R} and LL_{M_D} , their point-wise similarity will be calculated:

$$LL_M S(x, y) = sim(LL_{M_D}(x, y), LL_{M_R}(x, y), c) \quad (2.13)$$

$LL_M S$ is the similarity map of LL_{M_D} and LL_{M_R} . Similarly, $LL_\theta S$, $LL_\phi S$, and $LL_\psi S$ can be calculated. To collapse the information in $LL_M S$, $LL_\theta S$, $LL_\phi S$, and $LL_\psi S$ into one map, their point-wise multiplication is calculated:

$$LLS(x, y) = LL_M S^{10}(x, y) \times LL_\theta S(x, y) \times LL_\phi S(x, y) \times LL_\psi S(x, y) \quad (2.14)$$

LLS is the similarity map for the LL subband of the reference and the distorted image. The power of 10 for $LL_M S$ and 1 for $LL_\theta S$, $LL_\phi S$, and $LL_\psi S$, is obtained experimentally. For pooling LLS to one score, a weighted average of its values is computed. The weight of each pixel, $W(x, y)$, is the maximum of LL_{M_R} and LL_{M_D} :

$$W(x, y) = max\{LL_{M_R}(x, y), LL_{M_D}(x, y)\} \quad (2.15)$$

This weighting strategy is an effort to simulate an HVS property that relates to higher vision accuracy in brighter scenes [24]. Pixels with larger values, correspond to brighter spots and contribute more in determining the quality score. The score of LL is called Q_{LL} and obtained from relation 2.16. The scores for other subbands can be computed similarly.

$$Q_{LL} = \frac{\sum_x \sum_y LLS(x, y) \times W(x, y)}{\sum_x \sum_y W(x, y)} \quad (2.16)$$

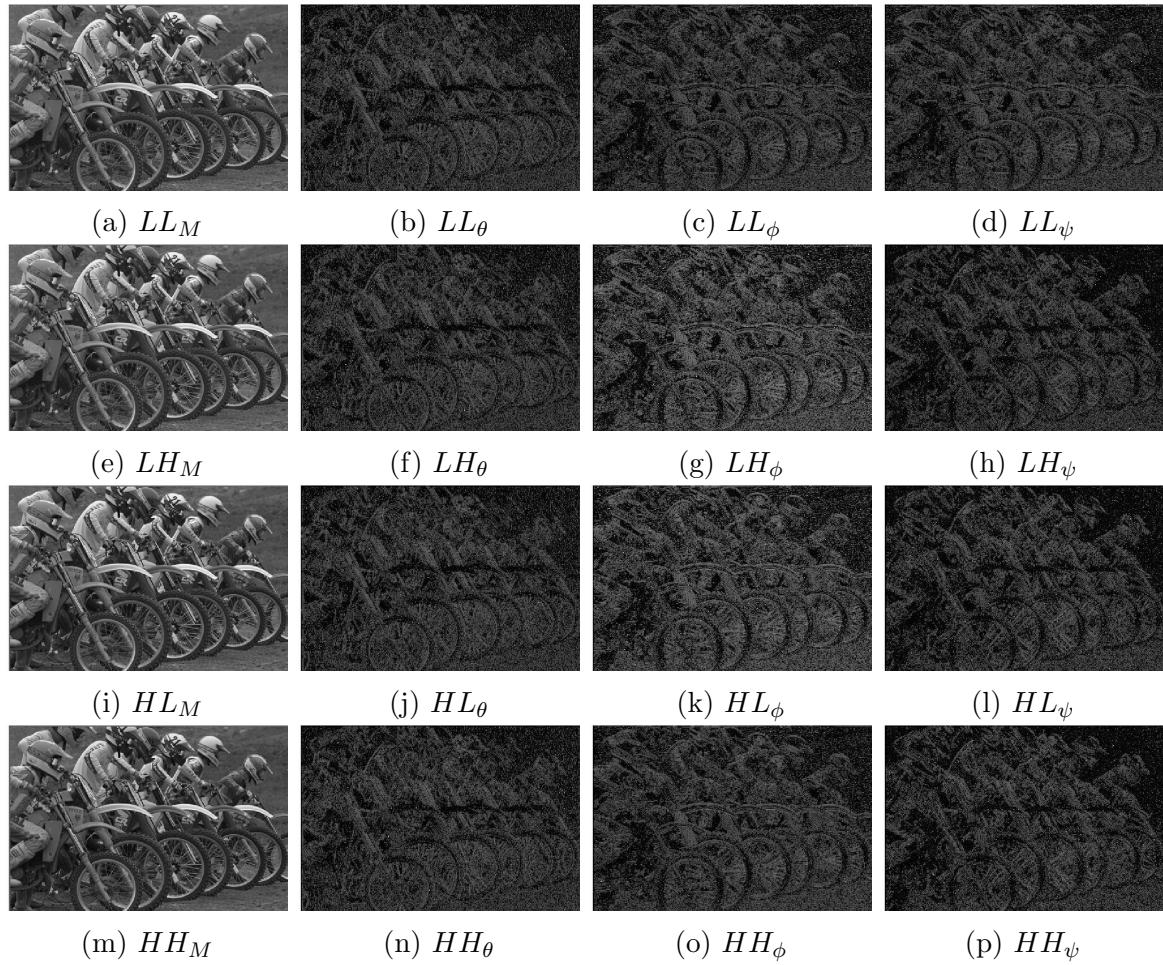


Figure 2.7: The QWT subbands of ‘bikes’ in one scale

According to 2.17, the final score is the linear combination of subbands' score. The values of coefficients a , b , c , and d were optimized on MLIVE [37] dataset.

$$Q = a \times Q_{LL} + b \times Q_{LH} + c \times Q_{HL} + d \times Q_{HH} \quad (2.17)$$

QWT has been used for single-distortion assessments [11, 83, 80]. This transform is used in QWT-IQA for multiple distortions, however its time complexity is not suitable for real-time applications.

2.3.3 Measuring Mutual Information

IFC [76] is a FR method for single distortions which is defined based on the mutual information between the reference and the distorted image. IFC assumes that the distorted image is always a noisy and blurry version of the reference. The severity of noise and blur is estimated via NSS. This way, IFC measures the amount of information that can be extracted from the reference, but not from the distorted image. VIF [77] is the improved version of this algorithm.

In MDIQA [109], IFC is modified for multiple distortions. IFC extracts NSS from the reference and the distorted in wavelet domain. Due to its model of distortion, IFC considers the mutual effect. In MDIQA, the wavlet subbands are filtered with contrast sensitivity function [65], prior to NSS extraction.

The implementations that authors provided for IFC are complex and MDIQA inherits the same complexity, yet with a minor increase in accuracy for multiple distortions.

2.4 RR Methods for Multiple Distortions

The methods presented so far, assumed that the reference image is one of the inputs to the algorithms. In the RR case, instead of the entire image, a feature vector of the image is available along with the distorted signal. The same features will be extracted from the distorted version and will be compared with the features of the reference. Similar to the FR scenario, the RR methods have been proposed specifically for multiply-distorted images and they will be reviewed in this section.

2.4.1 Combining Scores

Similar to [14], a method was proposed in [13] to combine the features of distortion-specific measures for constructing the feature vector of the reference and the test images. NR methods for measuring the severity of blur, ringing effect, blocking effect, and noise extract features from the reference image and build the reference feature vector. The same algorithms are applied to the distorted image and build the distorted feature vector. The concatenation of these two vectors is fed to an SVR for quality prediction.

The SVR will estimate the mutual and joint effect of the measured distortions, however, like the FR case, the execution of multiple methods is computationally expensive.

2.4.2 Using the Internal Generative Mechanism

Mahmoudpour and Schelkens used the internal generative model [19] for assessing multiple distortions [54]. In this model, the image is decomposed to two subbands; the *predicted* and the *residual*. These two subbands are invariant to specific artifacts. The predicted image is variant to structural degradations, such as blur, but invariant to noise. On the other hand, the residual image is variant to noise, but is not sensitive to blur.

The predicted image is transformed to Shearlet domain to extract the edges of the image. The residual subband is described with Renyi entropy [21]. The final feature vector is obtained by subtracting the distorted vector from the reference vector and mapped to the score using SVR. This algorithm is improved in RRSEA [55].

2.5 NR methods for Multiple Distortions

There has been more studies for the NR assessment of multiply-distorted images comparing to the FR and RR problems. In this section the methods are introduced with a review on their pros and cons.

2.5.1 Five-Steps and Six-Steps Methods

The first method proposed for NR assessment of multiply-distorted images was based on independent measurement of each distortion and the combining the measures into a single score for the entire image [28]. In the method called “FISBLIM”, Gu et al. mentioned

that when HVS is exposed to a multiply-distorted scene, it primarily notes the noise of the image [24]. There are mechanisms in the HVS that make the noise insensible [16] and then eye notices other distortions.

In the first step, they estimate the noise severity with SINE [113] and denoise the image based on the estimated intensity. Then they measure the blur and blocking effect of the noiseless image. The linear combination of these assessments gives the final score. In this framework, the severity of noise, blur and blocking effect can be measured with any distortion-specific method. Free energy principle [19] was incorporated in SISBLIM [29] to model human perception of distortion joint effect.

Although FISBLIM and SISBLIM are NR methods, they are not based on machine learning and compute the scores directly and could outperform conventional methods in assessing multiple distortions. These methods assume that the distortions of an image are a subset of noise, blur and Jpeg blocking artifact. This presumption limits their applications to images outside this set.

2.5.2 Combining Quality Parameters

It was shown in [44] that relevant perceptual features can describe the quality of images instead of measuring the distortions. Mean pixel value, image contrast, mean value of gradient magnitude, and a measure of texture complexity were the elements of the feature vector. Feature extraction is done in the spatial domain and the phase congruency of Fourier domain. In LQAF-RF [53] a random forest replaced SVR for mapping the features to quality in order to estimate the joint effect of the distortions. In [45] authors analysed the image in multiple scales which improved the accuracy, but caused a slower execution.

2.5.3 Dictionary-Based Approach

BRISQUE [62], BIQI [63], and SRNSS [32] are NR methods for singly-distorted images, that each of them extracts 36, 18, and 24 features out of an image, respectively. If we apply all of them to an image, we will have 87 features. Authors in BoWSF [51] measured the correlation of each of these 87 features with the severity of specific distortions in presence of other distortions. More correlated features were assumed to be more robust to the mutual effect of the distortions. So they ended up with a subset of features, called the ‘selected features’.

If we consider a set of N good-quality images, we can clip P patches from each of them and obtain $N \times P$ good-quality patches (Fig. 2.8). By extracting the selected features

from good-quality patches, we have $N \times P$ vectors. K -means clustering of these vectors will result in K vectors that are the center of each cluster. These K center-vectors will serve as the dictionary that describes a good quality image. For a test image, P patches are extracted and described by the selected features in BoWSF. The Euclidean distance of these vectors with the K centers, composes the feature vector that describes the entire image. Lasso regression [81] is then used to map the vector to score.

It was shown in [42] that different distortions appear differently in smooth or detailed image regions. For example, noise is less perceivable in crowded image regions. Clipping patches from various areas of an image, gives BoWSF the chance to consider the effect of distortion on different regions of image.

2.5.4 Learning to Measure Distortion Combinations

Zhang and Chandler [107] artificially distorted the images of Berkeley dataset [59] to create a set of multiply distorted images. Each image in their synthesized dataset falls into one of these seven categories:

- Only blurred
- Only noisy
- Only compressed with Jpeg algorithm
- Blurred and noised
- Blurred and compressed
- Compressed and noisy
- Compressed, noisy, and blurred

The images of each category vary in severity of the distortions. Like BoWSF [51], they measured the correlation of different features with different distortion levels. They experimented the features of GWH-GLBP [46], C-DIIVINE [108], and DESIQUE [106], then chose the most correlated ones as the ‘selected features’.

A classifier was trained to categorize an image into one of the seven mentioned classes based on the selected features. For each class, three SVRs were trained to measure the severity of blur, noise, and compression loss.

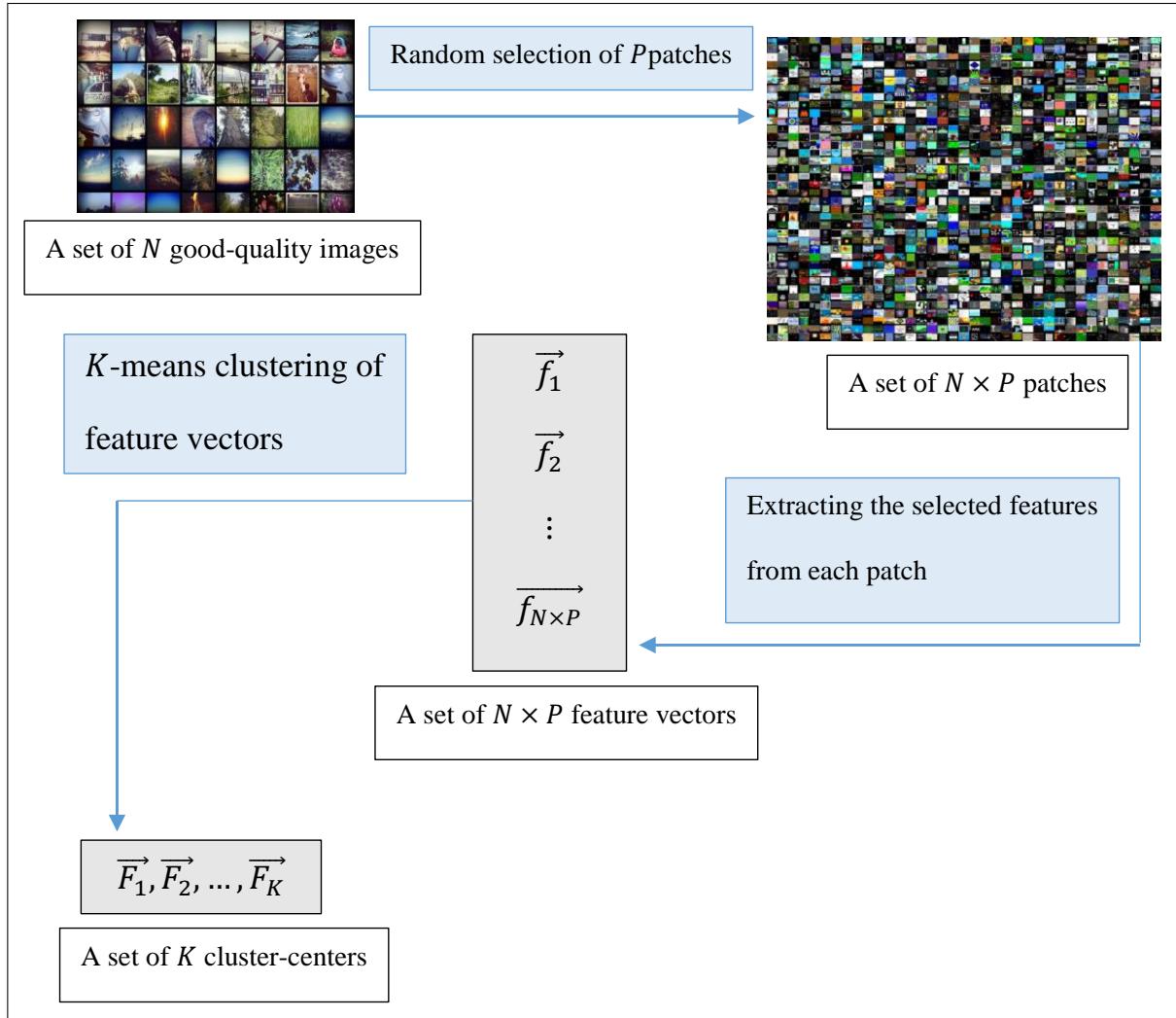


Figure 2.8: The procedure of building a dictionary

Fig. 2.9 illustrates the framework of this method, named ‘MUSIQUE’. The classifier uses the selected features to output seven probabilities, $P(i)$ s, where $i \in \{1, 2, \dots, 7\}$ and $\sum_{i=1}^7 P(i) = 1$. $P(i)$ is the probability that the image belongs to i^{th} distortion category. The selected features are also fed to seven groups of SVRs. In the group i , three SVRs are trained: one for estimating B_i (level of blur), one for N_i (level of noise), and one for J_i (level of Jpeg compression). That is for each \mathcal{D}_i , where $\mathcal{D} \in \{B, N, J\}$ and $i \in \{1, 2, \dots, 7\}$ a SVR is trained using the labeled distorted images. The final distortion parameters, B , N , and J , are obtained by the weighted average of \mathcal{D}_i s and $P(i)$ s:

$$N = \sum_{i=1}^7 P(i) \times N_i \quad (2.18)$$

$$B = \sum_{i=1}^7 P(i) \times B_i \quad (2.19)$$

$$J = \sum_{i=1}^7 P(i) \times J_i \quad (2.20)$$

The same pooling strategy in [7] maps the three parameters to a quality score.

Although these models were trained merely on the synthesized database, they can be generalized to existing multiple-distortion datasets with acceptable accuracy. However, similar to FISBLIM, MUSIQUE is designed for particular distortions and is also a complicated and complex method. If we want to train MUSIQUE for four different distortions, the number of classes will increase from 7 to 15.

2.5.5 Convolutional Neural Networks

Kang et al. [39] proposed a deep architecture for IQA. The test image is transformed to MSCN domain and then divided to 32×32 patches. Each patch is mapped to a score by the network and the total score is the average of the patches. The images and scores in the subjective datasets are used for training the network. The architecture is shown in Fig. 2.10.

The result of convolving a 32×32 image with a 7×7 filter with $stride = 1$, is a 26×26 feature map. By filtering the input patch with 50 filters, we will have 50 feature maps of dimension 26×26 . The maximum of each map is stored in a 1×50 vector (max-pooling). The minimum of the maps is stored in another 1×50 vector (min-pooling). All

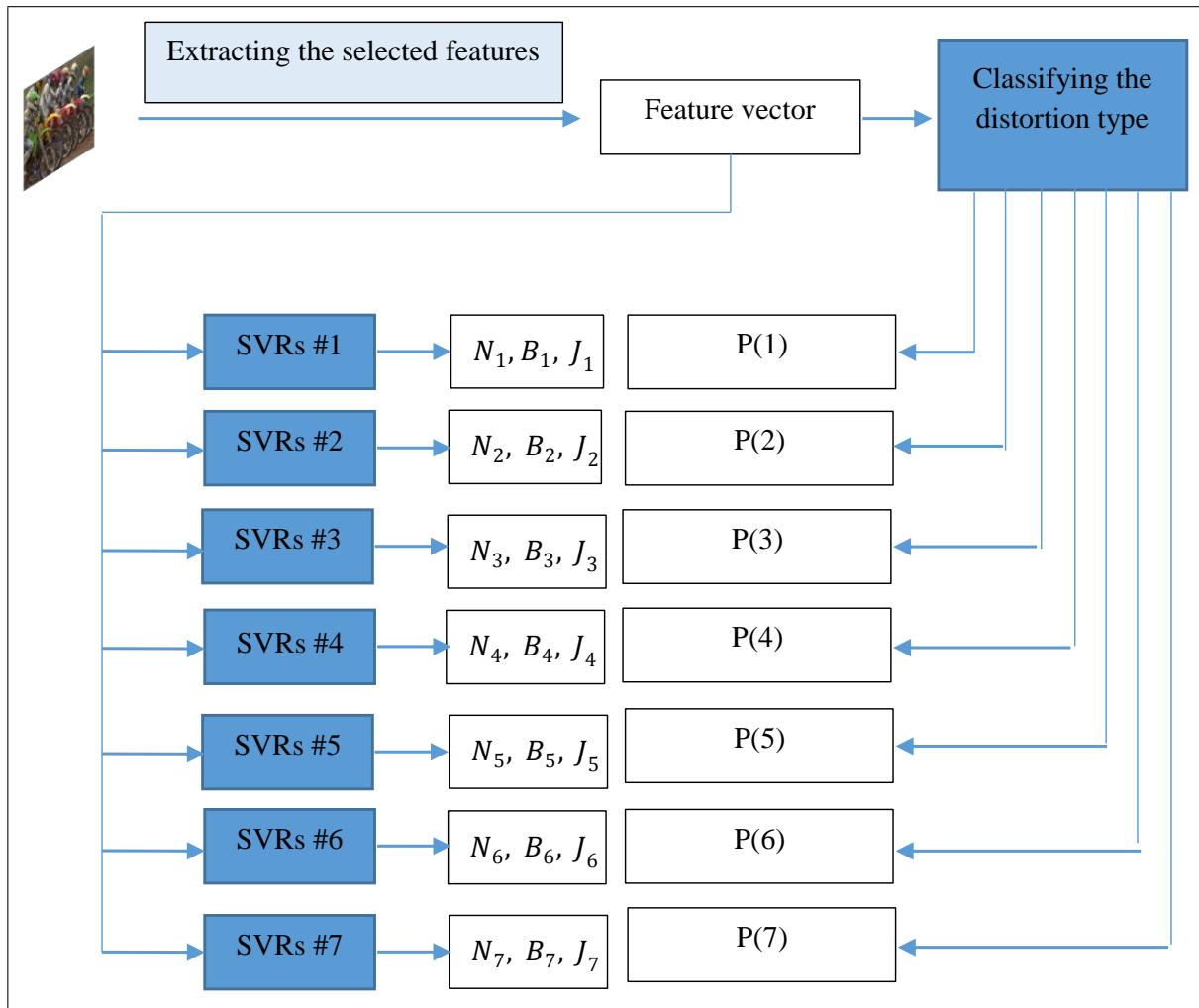


Figure 2.9: Learning the distortion parameters

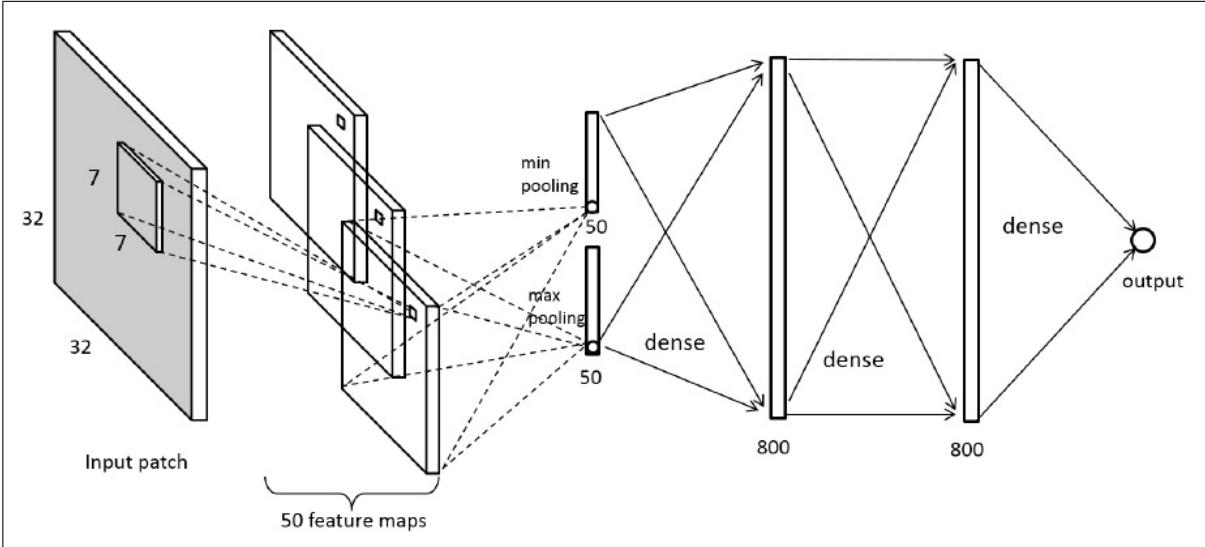


Figure 2.10: The architecture of Kang’s network

the elements of these two vectors are connected to the 800 neurons of the next layer. The output of these neurons is also connected to all of the 800 neurons in the next layer. The linear combination of these 800 outputs, is the final quality score.

It can be inferred, that the two pooled vectors are the extracted features for the patch and the last fully-connected layers do the regression. Fu et al. [20] increased the number of filters from 50 to 400. They also used average-pooling, instead of max and min-pooling to collapse the feature maps. The resulted architecture was trained and tested on a subject-rated multiple distortion dataset. The modified network was more accurate for multiply-distorted images.

EONSS [87] is another CNN, proposed for multiple distortions. Different from previous networks, EONSS is not trained with subject-rated images. Deep neural networks need numerous training samples, hence the authors, trained EONSS on artificially distorted images that their scores were obtained from full-reference methods. This way they were not limited in the number of training samples, since both images and labels were synthesized. However, they tested the method on subjective datasets. Despite of being ‘opinion-free’), the network could outperform other methods in terms of accuracy, but not speed. The architecture of EONSS is plotted in Fig. 2.11.

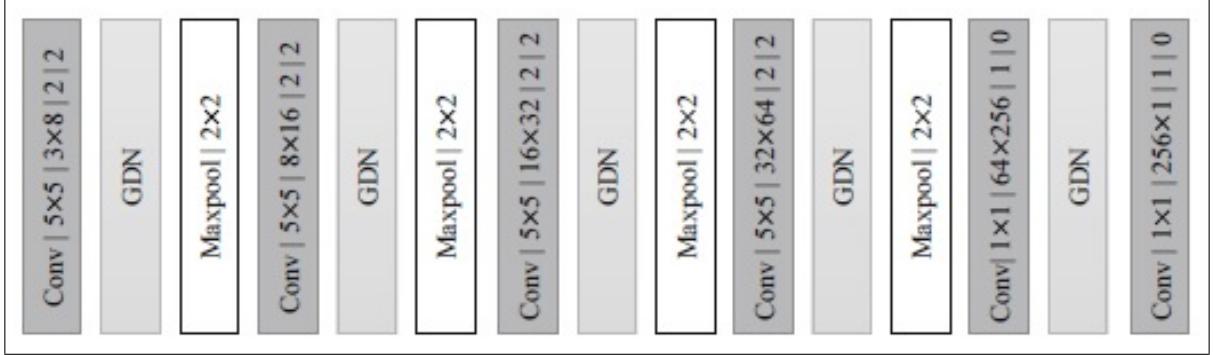


Figure 2.11: The architecture of EONSS

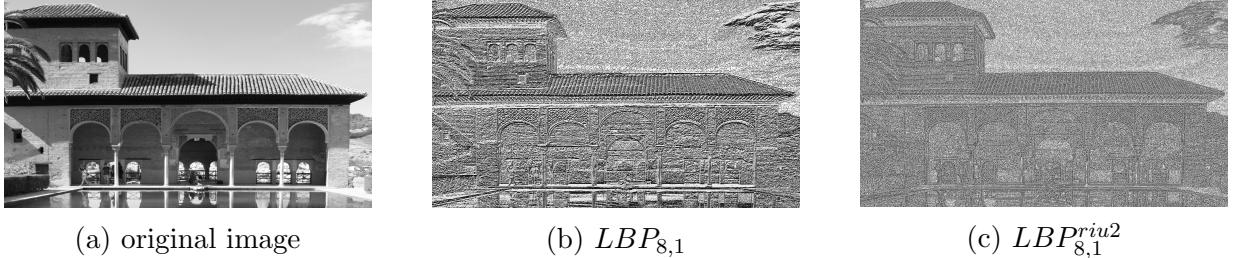


Figure 2.12: An image from MDID13 dataset along with its LBP map variants

2.5.6 Using Local Binary Patterns

The method GWH-GLBP [46] employed local binary patters (LBPs) to describe multiple distortions. LBP is an expressive descriptor in texture recognition [66]. Consider 8 adjacent pixels of a center pixel. The value of the center pixel is either less or greater than the value of each neighbor. If a bit is assigned to this situation, an 8-bit binary number describes the value of the center pixel relative to its neighbors. The decimal equivalent of this binary number, will represent the center pixel in the LBP map. An image and its LBP map are demonstrated in Fig. 2.12a-2.12b.

In the above explanations, 8 neighbors were considered that located at the radius 1 of the center pixel. If P is the number of neighbors and R is the radius, the LBP of pixel $I(x, y)$, can be formulated as:

$$LBP_{P,R}(x, y) = \sum_{p=0}^{P-1} f \left(I \left(x + R \cos \left(\frac{2p\pi}{P} \right), y + R \sin \left(\frac{2p\pi}{P} \right) \right), I(x, y) \right) \cdot 2^p \quad (2.21)$$

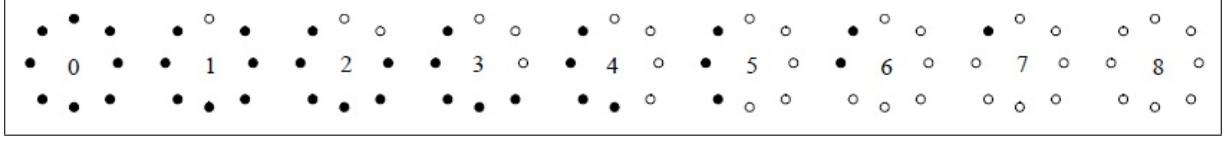


Figure 2.13: The nine possible binary patterns with 2 or less transitions

Where f outputs either 1 or 0 based on this definition:

$$f(\alpha, \beta) = \begin{cases} 0 & \alpha - \beta < 0 \\ 1 & \alpha - \beta \geq 0 \end{cases} \quad (2.22)$$

It is apparent that $LBP_{8,1}(x, y)$ can posses one of the $2^8 = 256$ possible values, but before converting it to a decimal number, we can count the number of transitions from 0 to 1 or from 1 to 0. If we categorize the binary patterns with respect to the number of their transitions, the class number of each pattern will be invariant to rotation. All binary patterns in Fig. 2.13 (figure form [66]) have either two or zero transitions. It is shown in [66] that these nine patterns have the best performance for describing images, which are invariant to rotation and luminance changes. The invariant LBP is called $LBP_{P,R}^{riu2}$ and formulated as:

$$LBP_{P,R}^{riu2}(x, y) = \begin{cases} \sum_{p=0}^{P-1} f(I(x + R\cos(\frac{2p\pi}{P}), y + R\sin(\frac{2p\pi}{P})), I(x, y)) & \mathcal{U}(LBP_{P,R}(x, y)) \leq 2 \\ P + 1 & o.w. \end{cases} \quad (2.23)$$

Where $I(x, y)$ and $f(., .)$ are the same as 2.21 and 2.22; and $\mathcal{U}(LBP_{P,R}(x, y))$ counts the transitions of the binary pattern of $LBP_{P,R}(x, y)$. The $LBP_{P,R}^{riu2}$, for $P = 8$, can have 10 distinct values. Fig. 2.12c shows an example.

It was shown in GHW-GLBP, that the LBP in gradient domain is effective in describing distortions. In this method, the gradient magnitude of the image is first computed and then described with an $LBP_{8,1}^{riu2}$ operator. Fig. 2.14 shows distorted versions of a scene along with their gradient magnitude and $LBP_{8,1}^{riu2}$ maps.

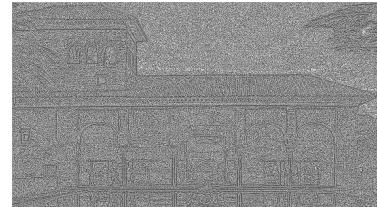
The histogram of $LBP_{8,1}^{riu2}$ will have 10 bins, where the height of each bin is equal to the frequency of the corresponding value (the middle column of Fig. 2.15). In GWH-GLBP, the frequency of the values is weighted by the value of gradient magnitude. The gradient magnitude of all (x, y) pixels with $LBP_{8,1}^{riu2} = k$, are summed up, and this summation determines the height of the histogram for bin k . Formally, if the $LBP_{8,1}^{riu2}$ which



(a)



(b)



(c)



(d)



(e)



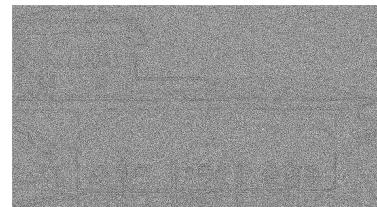
(f)



(g)



(h)



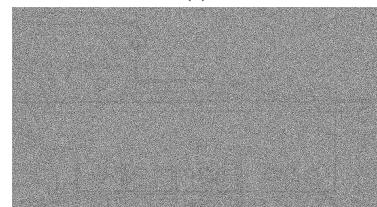
(i)



(j)



(k)



(l)

Figure 2.14: Images of different quality (left column) with their gradient magnitude map (middle column) and the LBP computed from gradient magnitude (right column)

is computed from the gradient magnitude map is addressed as $GMLBP_{8,1}^{riu2}$ and \vec{h}_{gw} is the weighted histogram of $GMLBP_{8,1}^{riu2}$, then \vec{h}_{gw} will have 10 bins. The height of bin k , where $k \in \{0, 1, \dots, 9\}$, is obtained from 2.24.

$$\vec{h}_{gw}(k) = \sum_{x=1}^N GM(x) \times f(GMLBP_{8,1}^{riu2}(x), k) \quad (2.24)$$

Where N is the number of pixels, x is their index, and $f(.,.)$ is defined in 2.25. The weighted histograms are demonstrated in the rightmost column of Fig. 2.15. Weighting the histograms is also applied in HOG [18] which, in a sense, encodes the dependency between $LBP_{8,1}^{riu2}$ values and gradient magnitude in the feature vector.

$$f(\alpha, \beta) = \begin{cases} 0 & \alpha = \beta \\ 1 & \alpha \neq \beta \end{cases} \quad (2.25)$$

By considering the global precedence principle [36], histograms are computed in various scales (5 scales), in order to mimic the primitive vision. These histograms are concatenated to compose the feature vector. The mapping is done with a SVR.

With the use of LBP, GWH-GLBP can also take the texture into account. It is shown in [33, 95, 27] that the relation between image structures of different granularities is meaningful for IQA. This relation can be expressed in terms of statistical dependency of variables that describe the structures. This happens in the GWH-GLBP, because LBP is more sensitive to micro-structures and gradient magnitude is more sensitive to macro-structures.

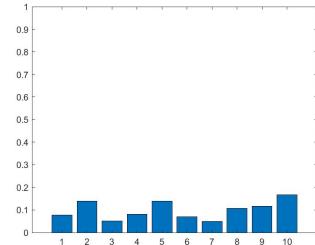
These factors caused the feature set of GWH-GLBP to be acceptably expressive without being trained for a particular combination of distortions. Instead of gradient magnitude map, [61] computed the LBP map on the phase congruency image. Which is another way to consider different structures. These strategies are employed in other metrics for IQA of multiple distortions.

2.5.7 LBP for the MSCN Image

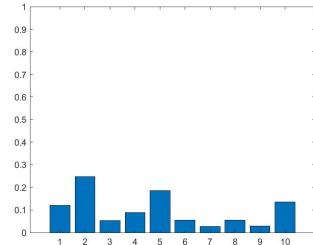
In addition to the features used in GWH-GLBP, Dai et al. [17] also described the MSCN map with a LBP operator. It is shown in [41] that the structural information of an image can be divided to the first-order and second-order information. The authors of [17] consider the gradient magnitude as the first-order descriptor and MSCN values as the extractor of the more detailed structures of the second order.



(a)



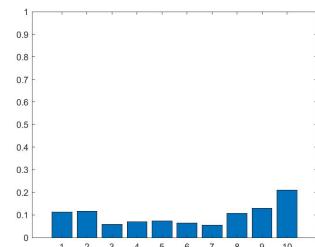
(b)



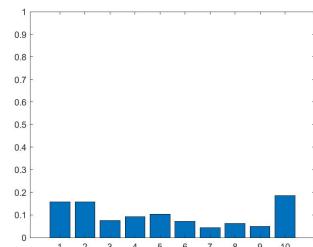
(c)



(d)



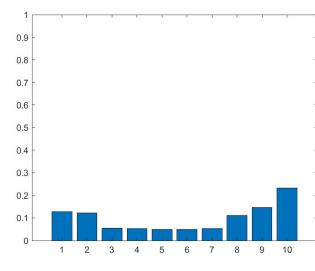
(e)



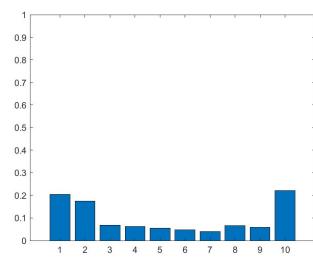
(f)



(g)



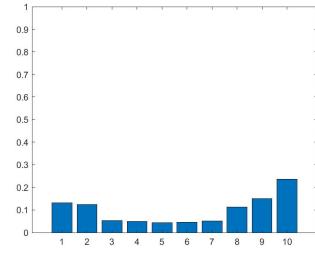
(h)



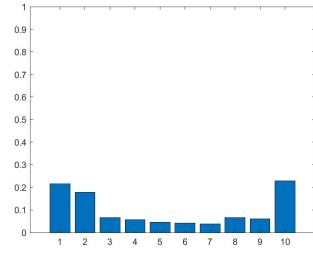
(i)



(j)



(k)



(l)

Figure 2.15: Images of different quality (left column) with their simple histogram (middle column) and their weighted histogram (right column)

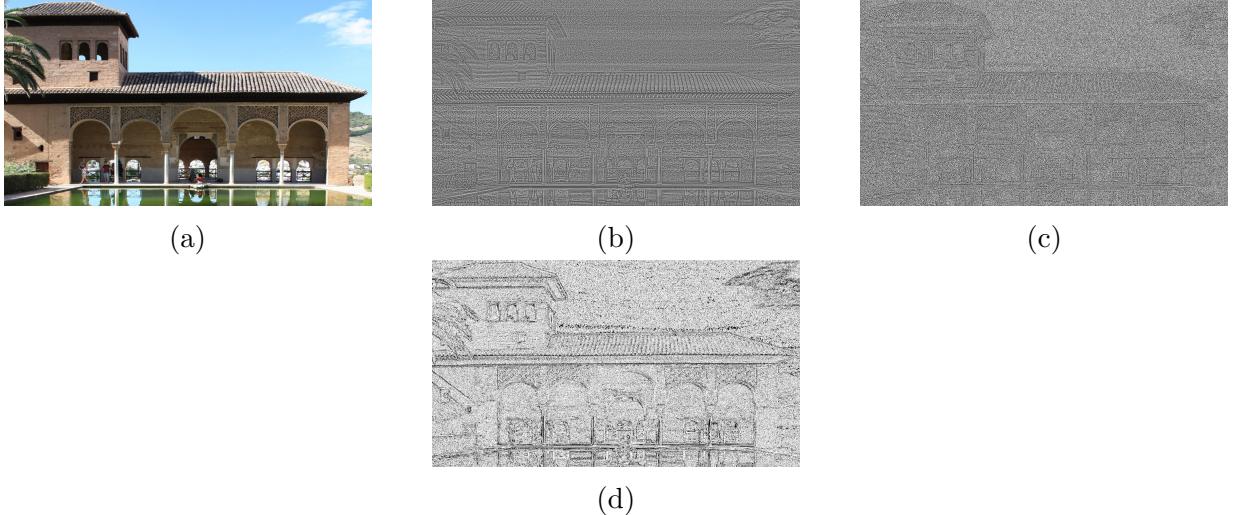


Figure 2.16: An image with its $LBP_{8,1}^{riu2}$ (c) and $GCSLBP_{8,1}$ (d) maps; extracted from the MSCN values (b)

Hence, two sets of features are extracted from the image, corresponding to the first and second order image structures. The first-order features are the weighted histogram of $LBP_{8,1}^{riu2}$, which is computed in the gradient domain. It is similar to [46], only different in the filter used for computing the gradients. The second-order features are the simple histogram of $GCSLBP_{8,1}$, that are computed in the MSCN domain. The difference of $GCSLBP_{P,R}$ with $LBP_{P,R}$ is that, instead of subtracting neighbor pixels from the center pixel, center-symmetric pairs of pixels are compared with each other.

In a neighborhood with P members, $\frac{P}{2}$ pixels are in opposite locations, hence $GCSLBP_{P,R}$ can have $2^{\frac{P}{2}}$ distinct values. In Fig. 2.16d the $GCSLBP_{8,1}$ is demonstrated for a sample. It is seen that structures in $GCSLBP_{8,1}$ are more meaningful in comparison to $LBP_{8,1}^{riu2}$ (Fig. 2.16c). Similar to GWH-GLBP, the histograms are extracted from multiple scales (3 scales). Considering the second order features could increase the accuracy while preserving an acceptable time complexity.

Instead of simply mapping the feature vectors with SVR, the authors applied a random subspace method [34] to the feature space which improved the accuracy. As seen in [53], different mapping strategies can be effective for the same feature sets. This is also seen in [111], where a deep neural network handles the regression task for multiply-distorted images.

2.5.8 LBP for Color Information

Hadizadeh and Bajic [31] employed Gaussian jets [26] for extracting image structures of various orders. Like the two previous methods, feature extraction is done in different scales (5 scales). In their method, Color-jet, they also took into account the color information of an image. This is done by using the chromatic feature maps of [84], which are invariant to luminance changes. The final description of each map is achieved by $LBP_{8,1}^{riu^2}$. The norm of Gaussian jets is used to weigh the histogram of $LBP_{8,1}^{riu^2}$ maps.

2.5.9 A New Local Binary Pattern

The approach of describing the micro and macro structures and analysing the image in multiple scales has also been taken in the work of Yue et al. [99]. Instead of scaling the image, the authors varied the radius of the LBP.

It is shown in [68] that while gazing at a point the information of the surrounding region is obtained at a lower rate as we go away from the center. An LBP was proposed in IMLBP for modeling this phenomenon. This improved LBP is then calculated for various radii. The simple histogram of the resultant maps comprises the feature vector of the image. Similar to [17], random subspace sampling is used to improve the accuracy of the method.

2.5.10 Describing Structural Information with Various Features

BOSS, the method proposed in [112], uses 2D entropy [110] to measure the amount of structural information and uses singular values to describe the structures of the image. For analysing the structures at two levels, LBP and image gradients are considered for the detailed and coarse structures, respectively.

Apart from the sophisticated structural analysis, natural scene statistics, which are extracted from a transform domain, contribute to the features that represent the image. The variance and shape parameter of a generalized Gaussian distribution represent the statistics. Similar to previous methods, BOSS extracts the features in 4 scales and employs a random forest for regression. Several techniques that are incorporated in BOSS increase its accuracy, but also its complexity.

2.6 Analysis and Conclusion

In this chapter, the general approach for visual quality assessment was introduced. The common criteria, are the structural degradation and natural scene statistics.

Considering the texture, is common in assessing multiple distortions, which is done by using LBP or its variants. Another trend, is to jointly consider the fine and coarse details of the image. Simulating the multi-scale characteristics of the image is also effective for multiple distortions. Although methods such as FISBLIM, SISBLIM, and MUSIQUE are effective for the considered distortions, they are only applicable for a limited set of distortion combinations.

Suitable run-time has been given up to accurate prediction in many methods. Also, features that can describe a variety of distortion combinations are desirable. The accuracy of many methods is largely dependent on their training data, which demands method with better generalization. By analyzing the image structures at two levels, we try to address these issues with an algorithm in the next chapter.

Chapter 3

The Proposed Method

The study of the previous methods revealed that fine details and textures convey valuable information for assessing multiple distortions. Therefore, we consider such features and propose a FR and a NR method for assessing the quality of multiply-distorted images.

For preserving the speed of execution, it is desired that expressing the texture does not add a significant computational load to the algorithm. On the one hand, gradient directions are implicitly computed for calculating the gradient magnitude and on the other hand, gradient directions are more sensitive to fine details than gradient magnitude. So, we describe the macro-structures with gradient magnitude and the micro-structures with gradient direction.

By combining the similarity of gradient magnitude with the similarity of gradient direction in a FR method, we could increase the accuracy for multiply distorted images in an acceptable time.

To employ gradient direction in NR manner, the histogram of LBP values were used to construct the features vector. The relation between micro and macro structures is an effective feature of the current methods. In the proposed method, this relation is considered by weighting the histograms with gradient magnitude. Experiments show that the resultant generality of the method justifies the added features. The proposed FR and NR algorithms are presented in subsequent sections.

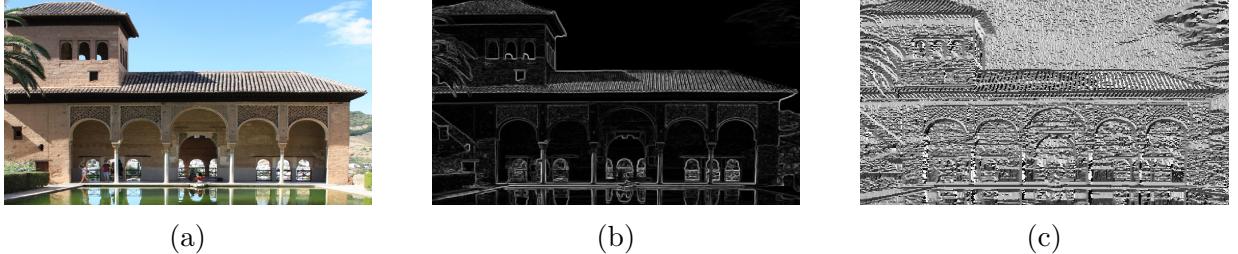


Figure 3.1: Magnitude and direction of gradient vectors for an image in (b) and (c), respectively.

3.1 Gradient Direction Similarity

Gradient directions are a good representative for human detection [18] and texture [92]. Fig. 3.1c shows the gradient directions for a sample image. It is seen that direction is more sensitive to fine details. In [97] gradient magnitude similarity is used for assessing single distortions. This similarity is defined in 2.10. Gradient magnitude is attributed to extract the coarse structures of an image [95] and we incorporate gradient directions for studying the effect micro-structures on the perceived quality of multiply-distorted images. If the direction of a gradient vector is defined as 2.9, the gradient direction similarity map of the images, r and d , is called GDS and obtained as:

$$GDS(r, d)_{(x,y)} = \frac{2 \times GD_r(x, y) \times GD_d(x, y) + c}{GD_r(x, y)^2 + GD_d(x, y)^2 + c} \quad (3.1)$$

Where, GD_r and GD_d are the gradient direction maps of r and d , and c is a constant for avoiding the instability of the fraction. It is seen in Fig. 3.2 that gradient direction is sensitive to image artifacts. With the point-wise multiplication of GDS and GMS , the gradient similarity map is obtained, which we address it as GS :

$$GS(r, d)_{(x,y)} = GMS(r, d)_{(x,y)} \times GDS(r, d)_{(x,y)} \quad (3.2)$$

For collapsing the values of GS to a single score, a weighted average of its values is computed. The weight of each pixel is equal to maximum of gradient magnitude from either the reference or the distorted image:

$$W(x, y) = \max \{GM_r(x, y), GM_d(x, y)\} \quad (3.3)$$

So the quality score is defined as below:

$$Q = \frac{\sum_x \sum_y W(x, y) \times GS(x, y)}{\sum_x \sum_y W(x, y)} \in [0, 1] \quad (3.4)$$

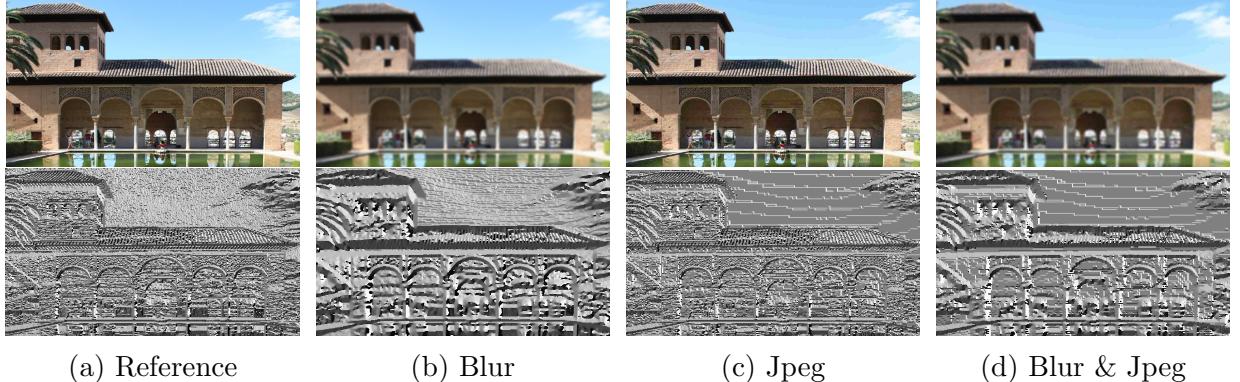


Figure 3.2: Sensitivity of gradient direction to various artifacts

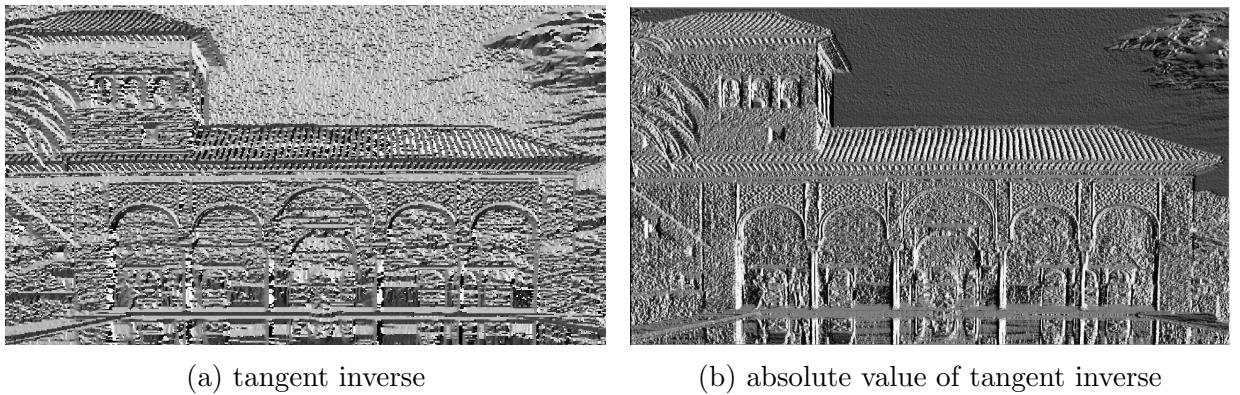


Figure 3.3: Two alternatives for representing gradient directions

A more accurate prediction is achieved with computing the direction similarity and weighting with gradient magnitude. Instead of considering θ and $-\theta$ as two different directions, we can use the absolute value of 2.9 as the ‘gradient direction’:

$$GD_I(x, y) = \left| \tan^{-1} \left(\frac{g_v}{g_h} \right) \right| \quad (3.5)$$

Fig. 3.3 shows the difference. It is seen that Fig. 3.3b has a higher contrast. [18] also uses the same representation and this increases the accuracy up to one percent per dataset.

So the magnitude of gradient can be accompanied by gradient direction in a full-reference manner that we called it gradient similarity [12]. In the next section we use gradient direction in a NR method.

3.2 The NR Method

The proposed NR method is a feature-based method in which, a vector is extracted from the image and a model maps it to a score. The algorithm for extracting the features and the model are explained in the following sub-sections.

3.2.1 Extracting Direction-Based Features

Considering its effect in increasing the accuracy of the FR method, gradient direction sounds to be a good descriptor for quality. Therefore, we were motivated to incorporate it in NR assessment. The purpose here is to code the quality-relevant aspects of an image in a feature vector. Simple histogram of pixel values is not suitable, due to redundancy and a high-dimension.

To shrink the size of feature vector, the values can be quantized. If v is the value of a pixel in the GD , then $v \in [0, 180]$. If we want to quantize v to 11 values, the truncated v is defined as:

$$v_q = \lfloor \frac{v}{18} \rfloor \quad (3.6)$$

and $v_q \in \{0, 1, 2, \dots, 10\}$. Then the histogram of GD will have 11 bins. Fig. 3.4 shows such histogram for various qualities of two different scenes. The quality decreases from top to bottom.

The features that describe quality, must be independent from image content and correlated with distortions. We see in Fig. 3.4 that quantized directions are independent from the scene, but do not change with the perceived quality. The bins are quite similar for the second and third row and do not distinguish their quality.

by weighting the bins with gradient magnitude, we can encode their relation in the histogram. In Fig. 3.5 we see that this is effective for sensitivity to quality. Such weighting considers micro and macro structures which have their own sensing mechanisms in HVS [26].

In a local region, the direction of gradient vectors are highly coherent and they are usually computed within 8×8 windows [18]. So in order to extract the local structural deviation, we compute the gradient magnitude of GD with 3.7 and name it GM_{GD} . GM_{GD} increases with this deviation (Fig 3.6). Also, this derivation whitens the information in GD .

$$GM_{GD}(x, y) = \sqrt{\left(\frac{\partial}{\partial x} GD(x, y) \right)^2 + \left(\frac{\partial}{\partial y} GD(x, y) \right)^2} \quad (3.7)$$

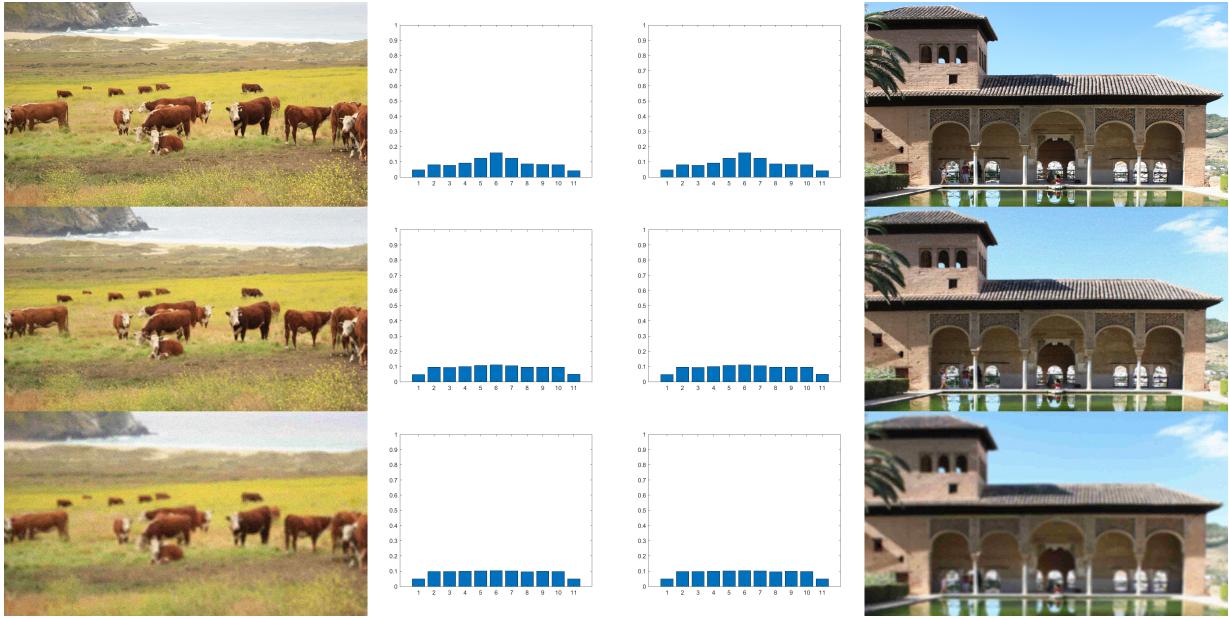


Figure 3.4: Histogram of direction values after quantization

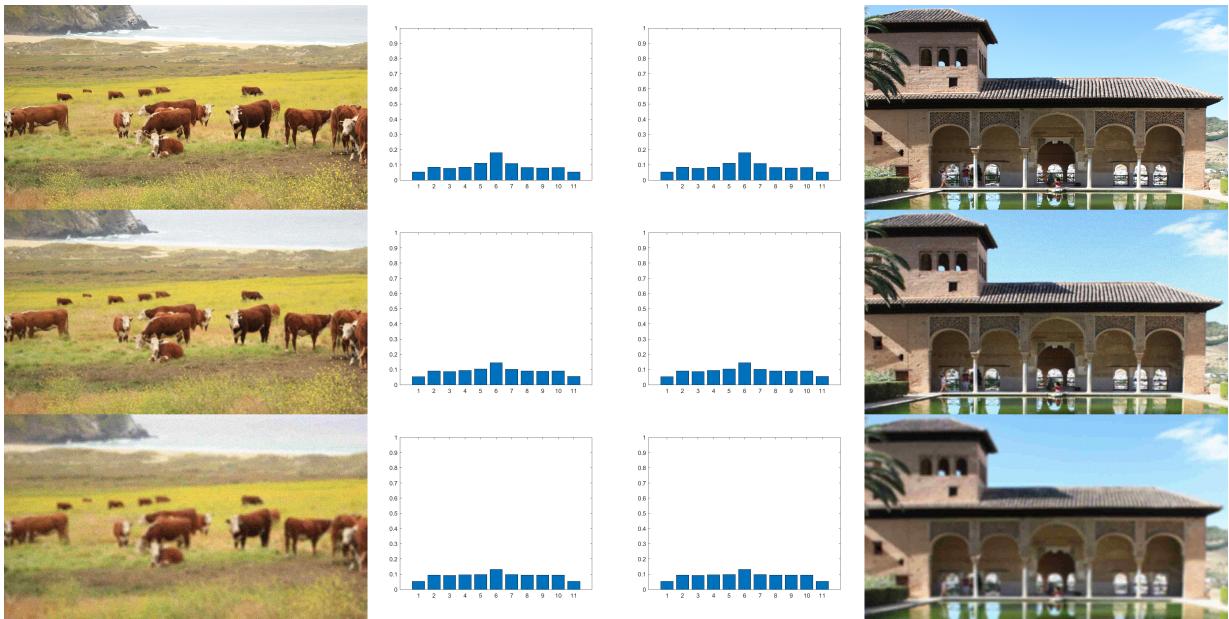


Figure 3.5: Weighted histogram of direction values

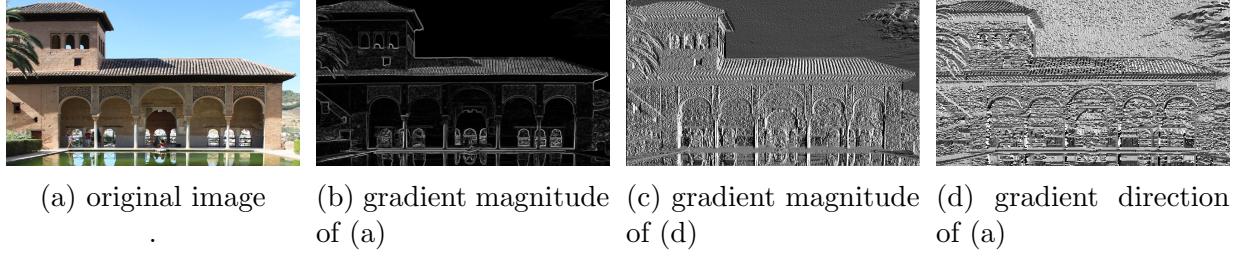


Figure 3.6: Gradient magnitude and direction for (a) in (b) and (d), respectively. Derivative of gradient direction in (c)



Figure 3.7: Difference of $LBP_{P,R}^{riu2}$ and $MLBP_{P,R}^{riu2}$. (a) in $LBP_{P,R}^{riu2}$ the difference of n_c and P n_i s is computed, where n_i s are located in a neighborhood of radius R and arcs with angles $\theta = \frac{360}{P}$ (b) a neighborhood of size 3×3 is shown around each n_i . The ‘median’ of each neighborhood is considered for subtracting from n_c

To reduce the dimensions and extract NSS, we employed LBPs instead of quantizing the maps. LBP is expressive and describes textures. It was pointed out in IMLBP [99] that the sampling rate of HVS drops when getting away from the gazing point [68]. Hence, they changed the operation of LBP to model this phenomenon.

In $LBP_{P,R}^{riu2}$, the difference of the center pixel, n_c and the neighbors, n_i is computed, where $i = 1, 2, \dots, P$ (Fig. 3.7a). To simulate the HVS mechanism, a neighborhood of n_i is considered with size $(R - 1) \times (R - 1)$ (Fig. 3.7b [99]). The median value of this neighborhood is addressed as M_{n_i} . Then we subtract n_c from M_{n_i} s and name the new operator ‘ $MLBP_{P,R}^{riu2}$ ’.

$MLBP_{8,1}^{riu2}$ can obtain one of 10 possible values in $\{0, 1, 2, \dots, 9\}$. The maps of gradient

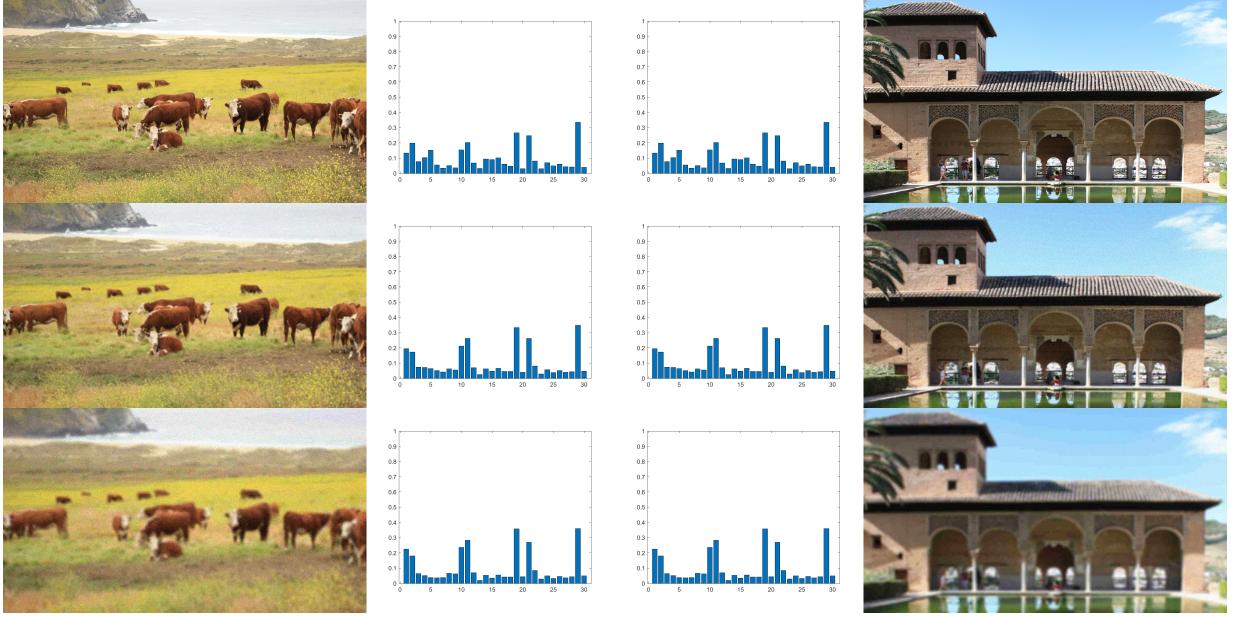


Figure 3.8: Proposed features for different qualities of two scenes

direction will be described by this operator.

To extract the feature vector of a test image, it is firstly decomposed to GM , GD , and GM_{GD} maps. The computation of pixels of these maps is formulated in 2.8, 2.9, and 3.7. Then, $LBP_{8,1}^{riu2}$ is computed for GM , and $MLBP_{8,1}^{riu2}$ is computed for GD and GM_{GD} . The histograms of the three maps are weighted by GM . The final feature vector is the concatenation of these histograms. The same scenes in Fig. 3.4 are described with these features in Fig. 3.8, where we see the bars fluctuate according to the variation of quality.

For modeling the global precedence principle [36], feature extraction takes place in multiple scales. Each histogram has 10 bins, so with computing 3 histograms, we end up with a 30-element vector. We compute the vector for 3 scales, so the final feature vector will be of dimension 1×90 .

3.2.2 Computing the Quality Score

The workflow of feature-based methods is described in Fig. 2.3 and 2.4. If $\vec{f}_{1 \times 90}$ is the extracted feature vector, a model regresses the final score using the features. This model can be a simple arithmetic relation, a neural network, or a SVR.

Although, the method's performance is mainly determined by the expressiveness of the features, for multiple distortions, the model turns out to be very effective [17, 99, 111]. Here we use a SVR [78].

SVR model is a function, F that combines the elements of \vec{f} to the quality score as in 3.8, where $\vec{\omega}$ and \vec{b} are the weights. If (\vec{f}_i, y_i) is a sample from a dataset and y_i is the human label, for $i \in \{1, \dots, N\}$, then the SVR computes the optimum values for $\vec{\omega}$ and \vec{b} [10].

$$F(\vec{f}) = \vec{\omega} \cdot \vec{f}^T + \vec{b} \quad (3.8)$$

3.3 Summary

In this chapter we used the direction of gradient vectors to describe the quality of images in a FR and a NR scenario. In the next chapter we evaluate the speed and accuracy of the proposed methods.

Chapter 4

Experimental Results

As discussed in Chapter 1, the ultimate goal of objective IQA algorithms, is to replace human assessments, so the accuracy and speed of these algorithms must be evaluated. Evaluations are done using the subject-rated images in IQA datasets and according to a few testing conventions. In this chapter we introduce the datasets and evaluation criteria, then we will report and compare the performance of the proposed methods.

4.1 IQA Datasets

Three datasets are common for multiply distorted images, MLIVE [37], MDID2013 [29], and MDID2016 [79]. Each image in MLIVE falls into one of the following six categories:

- It is not distorted (it is a reference image).
- It is blurred (singly-distorted).
- It has blocking effect because of JPEG compression (singly-distorted).
- It is noisy (singly-distorted).
- It is blurry with JPEG artifact (multiply-distorted).
- It is blurry and noisy (multiply-distorted).

So only two multiple-distortion are present in MLIVE, [blur+JPEG] and [blur+noise]. The severity of these distortions will vary for different distorted versions. There are two folders in MLIVE, “blurjpeg” and “blurnoise”. In each folder, there are 15 reference images and 225 distorted images, it will be 15 distorted versions for each reference. The type and severity of the artifacts in each contaminated version is summarized in table 4.1. Table 4.2 gives the same information for folder “blurnoise”. The reference images are identical in the two folders. The order of applying the artifacts in multiply-distorted images was in a way to consider the mutual effect, that is, the images are first blurred and then the second artifact is applied.

MDID2013 has 12 reference images that are contaminated with 3 artifacts. Each distorted image in MDID2013 is simultaneously blurred, noisy, and compressed with lossy JPEG. There are 3 various levels for each noise, so there will be $3 \times 3 = 27$ distorted versions for each reference.

The number of distorted images increases to 1600 in MDID2016. They are the result of applying 5 types of distortions to 20 references. There are 80 distorted versions for each reference. Type and severity of distortions varies in these 80 versions. ‘blur’, ‘noise’, ‘JPEG’, ‘JPEG2000’, and ‘contrast change’ are the set of all distortions. For each of the 80 versions, a random selection of them is applied with a random severity.

4.2 Evaluation Criteria

The performance of an algorithm is evaluated from the two aspects of speed and accuracy. For measuring the accuracy, statistical correlation indexes, such as SROCC, PLCC, and KROCC, are used. These indexes are calculated for the scores of the algorithm and the subjective scores. SROCC and KROCC measure the scores monotonicity, PLCC measures the accuracy, and RMSE can measure the consistency of predictions.

For evaluating the methods that use trainable models, a 80-20 division of the dataset is usually considered, that is, they train the model on 80% of the dataset and leave the other 20% for test. In IQA datasets, the distorted versions of a reference image comprise a ‘scene’. Considering MDID2013 as an example, it has 12 scenes that each of them has 27 members (there are 12 references that each of them has 27 distorted versions). When training a model on MDID2013, 10 ($= \lceil 0.8 \times 12 \rceil$) scenes will be randomly selected for training and the other 2 scenes for testing. For cross-validation, the 80-20 division will be performed 1000 times. The median of the 1000 test results will be reported as the final performance.

To further assess an algorithm, the train portion in the 80-20 division can be reduced. It means that we can test the algorithm for 70-30, 60-40, 50-50, or 40-60 divisions. If the proposed features are effective, the model's performance will be less affected by reducing the training data.

The generalization ability is also vital. If the features are effective, the over-fit problem will be alleviated and the model can be generalize to the samples outside the dataset. It is observed that the performance of the models drop significantly when they are trained on one dataset and tested on another one.

The run-time of the algorithms is usually reported as the average time for images of a dataset with a certain spatial resolution. For FR methods, the time required to compute the score will be measured and for NR algorithm, the time for extracting the feature vector will be reported as the computational complexity.

4.3 Evaluating the FR Method

The accuracy of the proposed FR method for MLIVE, MDID2013, and MDID2016 datasets is reported in table 4.3. The algorithm is tested for the two modes of 180° and 360° (180° designates that the absolute value of tangent inverse is considered). The accuracy of single-distortion methods is also included for comparison. The run-time of algorithms is shown in table 4.4.

For a fair comparison of computational complexity, the author-provided codes are required which may not be available. For such cases, we reported a relative run-time. As in [87], if a method has reported its run-time along with a third method, and we have the code of the third method, we can compare the speed of our algorithms in relevance to the common code. In the second row of the table 4.4, we have provided the relative run-time of the algorithms to that of the PSNR. For Chetouani's [13] method, we considered the time needed for extracting the feature vector. The experiments were carried out on a Windows laptop with a 2.0GHz CPU and 12GB of RAM in MATLAB. MLIVE images were used for measuring the average speed.

Chetouani2015 [13] extracts a feature vector out of two images and uses a trainable model, hence, it is capable of estimating the joint effect and achieving a high accuracy. The other methods compute the score using mathematical relations and directly model the joint and mutual effects. Therefore, its expected that they be inferior to the accuracy of Chetouani2015 [13]. However, the method in [13] needs to execute VIF, IFC, WASH, and SSIM for each image. Any time [13] runs, our algorithm can assess 150 images. The same

holds for MDIQA and VIF. Despite a higher accuracy, the proposed method is 74 times and 170 times faster than MDIQA and VIF, respectively.

Incorporating gradient directions caused the improvement of prediction monotonicity for all datasets in comparison to GMSD. QWT-IQA is more accurate on MLIVE, but it is considerably more complex. However, its accuracy is inferior to the proposed method for MDID2013 and MDID2016. Clearly the proposed FR method has been successful in achieving an acceptable trade-off between complexity and speed.

4.4 Evaluating the NR Method

Similar to the FR method, the proposed NR algorithm will be assessed regarding its speed and accuracy. Since, this algorithm uses a trainable model, we evaluate and compare its generality and robustness in separate sections.

4.4.1 Performance on Individual Datasets

In this experiment, 80% of the scenes in a dataset are used for training and 20% for testing. For cross validation, the mentioned procedure is performed 1000 times. The scenes for train and test are randomly selected each time. The median of these 1000 tests will be reported. In table 4.5, we see the SROCC, PLCC, RMSE, and KROCC indexes for the proposed method. The results for other method is quoted from authors' reports. The weighted average of correlation indexes is provided in the last three rows, where the weight of each dataset equals its number of images.

BOSS [112] has the superior performance among the compared methods and its model accurately predicts the quality of MDID2013. According to the results, MDID2016 seems challenging because non of the previous methods has reached 90% precision. The proposed method has the best performance for MDID2016, which outnumbers the other datasets in terms of distortions and scenes. Although it is inferior to BOSS, the proposed method surpasses other algorithms in the overall evaluation.

4.4.2 Impact of Training Data

To inspect the impact of reducing the training data, we performed the previous experiments with the following portions of train-test data: 70-30, 60-40, 50-50, and 40-60. As it is seen

in table 4.6, the proposed features are much more robust to reducing the training data in comparison to [46]. Although the SVD¹-based method in [112] outperforms the proposed method, our features are still comparable to the rich feature set of BOSS [112].

4.4.3 Generalization Ability

The real-world images may differ from those of the datasets, so it is very important that a trained model be applicable to samples outside the training set. To test this, considering the three multiply-distorted datasets, the following experiment was conducted. We trained the model on a whole dataset, a test it using the other two. Unlike the previous tests, the train-test procedure is done only once, not 1000 times. Table 4.7 shows the SROCC indexes. It is demonstrated that the proposed algorithm has the superior generalization ability. Although BOSS [112] employs a variety of complex features, the gradient direction-based features are less dependent on the training data.

4.4.4 Computational Complexity

The average time for extracting a feature vector for MLIVE images of size 1280×720 is reported in table 4.8. The major opponent of our method in terms of accuracy is BOSS [112]. It is better in learning individual datasets and inferior to our method regarding generality. It is seen from table 4.8 that the proposed method is simultaneously accurate and fast. No time complexity is reported for BOSS[112] and no implementatin is provided by the authors. However, according to its use of SVD and DCT statistics, it is computationally more intense than a spatial-domain method.

4.5 Future Works and Conclusion

Modeling human perception of image quality is a fundamental problem in artificial intelligence (AI) and computer vision, which is applicable in many image processing tasks. With the ease of internet access and acquisition or sharing digitized images, much information are delivered via pictures. This demands the monitoring of visual quality of digital images which entails their objective assessment.

¹Singular Value Decomposition

The complexity of human visual sensation and perception, limits the direct mathematical modeling of quality which in turn, makes machine learning techniques a suitable choice. Similar to other AI problems, human knowledge in a specific domain, can make learning easier for the machine. Therefore, image quality researches acknowledge efficient and expressive features, regardless of the advances in automatic methods.

In some applications, images usually contain multiple distortions and their assessment is different from the case of singly-distorted images and new methods have been proposed for handling multiple distortions. By studying these methods, it was revealed that some trends were common in successful algorithms. Considering image texture, representing the relation between bi-order structures, analyzing the image in multiple scales, and mapping the features with sophisticated methods seemed to be effective in assessing multiple distortions. In this research, we proposed a NR and a FR method that could successfully fill the gaps in the performance of previous algorithms.

It is apparent that promoting the models that map the features to the final score can be very effective for multiple distortions. This may stem in modeling the joint effect of simultaneous artifacts. Therefore, inspecting different regression techniques, such as multi-layer perceptrons or multiple SVR models can be considered for future studies.

Table 4.1: Severity of Blur and JPEG for 15 Distorted Versions of a Reference Image in ‘blurjpeg’ Folder. Six of these 15 versions are singly-distorted (the severity of one distortion is zero) and 9 are multiply-distorted

	Version #1	Version #2	Version #3	Version #4	Version #5	Version #6	Version #7	Version #8	Version #9	Version #10	Version #11	Version #12	Version #13	Version #14	Version #15
Blur Level	1	2	3	1	2	3	1	2	3	1	2	3	0	0	0
JPEG Level	0	0	0	1	1	1	2	2	2	3	3	3	1	2	3

Table 4.2: Severity of Blur and Noise for 15 Distorted Versions of a Reference Image in ‘blurnoise’ Folder

	Version #1	Version #2	Version #3	Version #4	Version #5	Version #6	Version #7	Version #8	Version #9	Version #10	Version #11	Version #12	Version #13	Version #14	Version #15
Blur Level	1	2	3	1	2	3	1	2	3	1	2	3	0	0	0
Noise Level	0	0	0	1	1	1	2	2	2	3	3	3	1	2	3

Table 4.3: The Accuracy of the Proposed Method and Other FR Algorithms

Dataset	Index	PSNR	SSIM [91]	GMSD [97]	VIF [77]	Chetouani2015 [13]	MDIQA [109]	QWT-IQA [43]	Proposed Method (360)		Proposed Method (180)	
									SROCC	PLCC	RMSE	KROCC
MLIVE (blur+JPEGG)	SROCC	0.66621	0.84488	0.8470	0.8788	0.9059	0.9054	0.9003	0.8609	0.8697		
	PLCC	0.7246	0.7971	0.8947	0.9052	0.9371	0.9273	0.9262	0.9097	0.9173		
	RMSE	13.2046	19.6387	8.5598	8.1427	6.7182	7.1264	7.2260	7.9582	7.6291		
	KROCC	0.4775	0.6520	0.6559	0.6922	0.7394	0.7309	0.7225	0.6730	0.6836		
MLIVE (blur+noise)	SROCC	0.7088	0.8760	0.8366	0.8807	0.9203	0.9149	0.9058	0.8709	0.8692		
	PLCC	0.7752	0.8333	0.8674	0.8492	0.9377	0.9298	0.9174	0.8572	0.8538		
	RMSE	11.7851	10.3120	9.2817	9.8499	6.3548	7.0071	7.4238	9.6079	9.7125		
	KROCC	0.5290	0.6867	0.6405	0.6930	0.7610	0.7473	0.7291	0.6815	0.6774		
MLIVE (ALL)	SROCC	0.6771	0.8604	0.8448	0.8823	0.9071	-	0.9043	0.8684	0.8726		
	PLCC	0.7398	0.8125	0.8808	0.8723	0.9305	-	0.9203	0.8584	0.8616		
	RMSE	12.7238	11.0238	8.9533	9.2487	6.8669	-	7.4036	9.7019	9.5986		
	KROCC	0.5003	0.6695	0.6548	0.6970	0.7433	-	0.7294	0.6828	0.6857		
MDID2013	SROCC	0.5604	0.4873	0.8283	0.8444	-	-	-	0.7794	0.8477	0.8599	
	PLCC	0.5647	0.5425	0.8307	0.8218	-	-	-	0.7896	0.8397	0.8512	
	RMSE	0.0419	0.0427	0.0283	0.0290	-	-	-	0.0312	0.0276	0.0267	
	KROCC	0.3935	0.3360	0.6240	0.6438	-	-	-	0.5617	0.6448	0.6632	
MDID2016	SROCC	0.5784	0.8328	0.8613	0.9306	0.9500	0.9031	0.8299	0.8871	0.8987		
	PLCC	0.6164	0.8457	0.8543	0.9367	0.9547	0.8934	0.8379	0.8975	0.9090		
	RMSE	1.7350	1.1757	1.1452	0.7717	0.6581	0.9901	1.2026	0.9715	0.9186		
	KROCC	0.4119	0.6446	0.6790	0.7714	0.8046	0.7209	0.6355	0.7068	0.7222		

Table 4.4: The Algorithms' Run-Time

	PSNR	SSIM [91]	GMSD [97]	VIF [77]	Chetouani2015 [13]	MDIQQA [109]	QWT-IQA [43]	Proposed Method (360)	Proposed Method (180)
Absolute Run-Time (s)	0.0041	0.0271	0.0233	2.4458	5.3271	-	-	0.0321	0.0338
Relative Run-Time (to PSNR)	1	7	6	445	969	1035	11	8	8
Accuracy	0.5784	0.8328	0.8613	0.9306	0.9500	0.9031	0.8299	0.8871	0.8987

Table 4.5: The Performance of NR Methods on Individual Datasets (MLIVE, MDID2013, and MDID2016).

Dataset	Index	FISBLIM [28]	SISBLIM [29]	Monogenic [111]	Jet-LBP [31]	BOSS [112]	GWH-GLBP [46]	Proposed Method
MLIVE	SROCC	0.8574	0.8572	0.943	0.953	0.9529	0.9437	0.9432
	PLCC	0.8808	0.8652	0.951	0.956	0.9549	0.9494	0.9496
	RMSE	8.9540	9.4836	5.747	5.510	5.6436	5.8660	5.8726
	KROCC	0.6700	0.6606	0.747	-	0.8184	0.7978	0.7968
MDID2013	SROCC	0.776	0.6934	0.907	0.921	0.9446	0.8967	0.9141
	PLCC	0.7802	0.7099	0.919	0.920	0.9502	0.9121	0.9273
	RMSE	0.0318	0.0358	0.019	0.016	0.0158	0.0197	0.0177
	KROCC	0.5737	0.4947	0.698	-	0.8010	0.7219	0.7393
MDID2016	SROCC	0.588	0.6545	-	0.8536	0.8969	0.8901	0.9008
	PLCC	0.4969	0.6313	-	0.8591	0.8997	0.8903	0.9050
	RMSE	1.9122	1.7089	-	1.1220	0.9824	0.9902	0.9354
	KROCC	0.4306	0.4717	-	-	0.7101	0.7016	0.7196
Weighted Average	SROCC	0.6647	0.6982	-	0.8816	0.9140	0.9012	0.9107
	PLCC	0.6083	0.6864	-	0.8858	0.9171	0.9045	0.9165
	KROCC	0.4955	0.5106	-	-	0.7430	0.7226	0.7369

Table 4.6: The Results for Changing the Ratio of Training Data.

Dataset	Ratio	Proposed Method				GWH-GLBP [46]				BOSS [112]			
		SROCC	KROCC	PLCC	RMSE	SROCC	KROCC	PLCC	RMSE	SROCC	KROCC	PLCC	RMSE
MDID2013	80-20	0.9141	0.7393	0.9273	0.0177	0.8967	0.7219	0.9121	0.0197	0.9446	0.8010	0.9501	0.0153
	70-30	0.8639	0.6708	0.8769	0.0241	0.8317	0.6449	0.8585	0.0242	0.9337	0.7773	0.9371	0.0177
	60-40	0.8524	0.6528	0.8614	0.0257	0.6642	0.4668	0.6623	0.0377	0.9298	0.7678	0.9320	0.0182
	50-50	0.8411	0.6408	0.8457	0.0270	0.6319	0.4690	0.6160	0.0379	0.9236	0.7576	0.9266	0.0192
	40-60	0.8296	0.6262	0.8322	0.0279	0.4260	0.2926	0.4666	0.0446	0.9088	0.7331	0.9104	0.0209
MLIVE	80-20	0.9432	0.7968	0.9496	5.8726	0.9437	0.7978	0.9494	5.8660	0.9529	0.8184	0.9549	5.6436
	70-30	0.9370	0.7795	0.9403	6.4530	0.8073	0.6397	0.8149	11.1371	0.9403	0.7901	0.9457	6.1659
	60-40	0.9191	0.7527	0.9239	7.1566	0.7924	0.5974	0.7842	12.2011	0.9413	0.7880	0.9443	6.2453
	50-50	0.9135	0.7432	0.9183	7.4310	0.6573	0.4697	0.6477	14.4161	0.9367	0.7795	0.9371	6.5934
	40-60	0.8990	0.7202	0.9053	8.0136	0.3272	0.2283	0.4131	17.4939	0.9267	0.7615	0.9286	6.9858
MDID2016	80-20	0.9008	0.7196	0.9050	0.9354	0.8901	0.7016	0.8903	0.9902	0.8969	0.7101	0.8997	0.9824
	70-30	0.8932	0.7069	0.8961	0.9746	0.8268	0.6231	0.8304	1.2316	0.8932	0.7098	0.8951	0.9877
	60-40	0.8853	0.6951	0.8888	1.0099	0.8178	0.6175	0.8185	1.2597	0.8836	0.7005	0.8895	0.9908
	50-50	0.8758	0.6816	0.8791	1.0502	0.8135	0.6092	0.8134	1.2839	0.8801	0.6945	0.8804	1.0626
	40-60	0.8639	0.6658	0.8671	1.0985	0.7058	0.5360	0.7164	1.5307	0.8771	0.6895	0.8753	1.0816

Table 4.7: SROCC Indexes for Cross-Dataset Evaluation.

Dataset for Train Dataset for Test		Proposed Method	GWHI-GLBP [46]	Jet-LBP [31]	BOSS [112]
MLIVE	MDID2013	0.8211	0.6587	0.7466	0.7939
	MDID2016	0.7588	0.5524	0.7307	0.5986
MDID2013	MLIVE	0.7199	0.204	0.4062	0.7756
	MDID2016	0.7138	0.3099	0.1583	0.4207
MDID2016	MLIVE	0.8851	0.7505	0.8866	0.7680
	MDID2013	0.8220	0.7403	0.8087	0.7122

Table 4.8: Methods' Time Complexity

	Proposed Method	GWH-GLBP [46]	Jet-LBP [31]
Absolute Run-Time (s)	0.9218	0.2769	2.2537
Relative Run-Time (to GM-LoG [95])	5	1.5	12
Accuracy	0.9107	0.9012	0.8816

References

- [1] The Zettabyte Era: Trends and Analysis - Cisco, 2018.
- [2] Juha Alakarhu. Image sensors and image quality in mobile phones. In *Proc. 2007 International Image Sensor Workshop*, 2007.
- [3] Louis Anegekuh, Lingfen Sun, Emmanuel Jammeh, Is Haka Mkwawa, and Emmanuel Ifeachor. Content-Based Video Quality Prediction for HEVC Encoded Videos Streamed Over Packet Networks. *IEEE Transactions on Multimedia*, 2015.
- [4] Vijay Badrinarayanan, Alex Kendall, and Roberto Cipolla. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12):2481–2495, 2017.
- [5] Rushikesh Borse and Prerana Markad. Competitive analysis of existing image quality assessment methods. In *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on)*, pages 1440–1444. IEEE, 2014.
- [6] Sebastian Bosse, Dominique Maniry, Klaus-Robert Muller, Thomas Wiegand, and Wojciech Samek. Deep Neural Networks for No-Reference and Full-Reference Image Quality Assessment. *IEEE Transactions on Image Processing*, 27(1):206–219, jan 2018.
- [7] Damon M. Chandler. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 2010.
- [8] Damon M. Chandler. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research. *ISRN Signal Processing*, 2013.
- [9] Damon M. Chandler and Sheila S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 2007.

- [10] Chih-Chung Chang and Chih-Jen Lin. Libsvm: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, 2(3):27, 2011.
- [11] Qiwei Chen, Yi Xu, Chuan Li, Ning Liu, and Xiaokang Yang. An image quality assessment metric based on quaternion wavelet transform. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2013.
- [12] Pooryaa Cheraaqaee, Azadeh Mansouri, and Ahmad Mahmoudi-Aznaveh. Incorporating gradient direction for assessing multiple distortions. In *2019 4th International Conference on Pattern Recognition and Image Analysis (IPRIA)*, pages 109–113. IEEE, 2019.
- [13] Aladine Chetouani. A Reduced Reference Image Quality assessment for Multiply Distorted Images. In *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pages 1–4. IEEE, nov 2015.
- [14] Aladine Chetouani. An Image Quality Metric with Reference for Multiply Distorted Image. pages 477–485. Springer, Cham, oct 2016.
- [15] Shyamprasad Chikkerur, Vijay Sundaram, Martin Reisslein, and Lina J. Karam. Objective video quality assessment methods: A classification, review, and performance comparison. *IEEE Transactions on Broadcasting*, 57(2 PART 1):165–182, 2011.
- [16] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 2007.
- [17] Tao Dai, Ke Gu, Li Niu, Yong bing Zhang, Weizhi Lu, and Shu Tao Xia. Referenceless quality metric of multiply-distorted images based on structural degradation. *Neurocomputing*, 290:185–195, 2018.
- [18] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005.
- [19] Karl Friston. The free-energy principle: A unified brain theory?, 2010.
- [20] J Fu, H Wang, and L Zuo. Blind image quality assessment for multiply distorted images via convolutional neural networks. *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1075–1079, 2016.

- [21] Salvador Gabarda and Gabriel Cristóbal. Blind image quality assessment through anisotropy. *Journal of the Optical Society of America A*, 2007.
- [22] Aarushi Gaur and Krystian Mikolajczyk. Ranking images based on aesthetic qualities. In *Proceedings - International Conference on Pattern Recognition*, 2014.
- [23] Deepti Ghadiyaram and Alan C. Bovik. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, 25(1):372–387, 2016.
- [24] Rafael C.. Gonzalez and Richard Eugene Woods. *Digital image processing*. 2008.
- [25] J S Goodman and D E Pearson. Multidimensional scaling of multiply-impaired television pictures. *IEEE Trans Syst Man Cybern*, 9(6):353–356, 1979.
- [26] Lewis D. Griffin. The second order local-image-structure solid. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007.
- [27] Ke Gu, Leida Li, Hong Lu, Xiongkuo Min, and Weisi Lin. A fast reliable image quality predictor by fusing micro- and macro-structures. *IEEE Transactions on Industrial Electronics*, 2017.
- [28] Ke Gu, Guangtao Zhai, Min Liu, Xiaokang Yang, Wenjun Zhang, Xianghui Sun, Wanrong Chen, and Ying Zuo. FISBLIM: A FIve-Step BLInd Metric for quality assessment of multiply distorted images. In *SiPS 2013 Proceedings*, pages 241–246. IEEE, oct 2013.
- [29] Ke Gu, Guangtao Zhai, Xiaokang Yang, and Wenjun Zhang. Hybrid no-reference quality metric for singly and multiply distorted images. *IEEE Transactions on Broadcasting*, 60(3):555–567, 2014.
- [30] Praful Gupta, Anush Krishna Moorthy, Rajiv Soundararajan, and Alan Conrad Bovik. Generalized Gaussian scale mixtures: A model for wavelet coefficients of natural images. *Signal Processing: Image Communication*, 66:87–94, aug 2018.
- [31] Hadi Hadizadeh and Ivan V. Bajic. Color Gaussian Jet Features For No-Reference Quality Assessment of Multiply-Distorted Images. *IEEE Signal Processing Letters*, 23(12):1717–1721, dec 2016.
- [32] Lihuo He, Dacheng Tao, Xuelong Li, and Xinbo Gao. Sparse representation for blind image quality assessment. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1146–1153, 2012.

- [33] M. Heydari, P. Cheraaqee, A. Mansouri, and A. Mahmoudi-Aznaveh. A low complexity wavelet-based blind image quality evaluator. *Signal Processing: Image Communication*, 74, 2019.
- [34] Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1998.
- [35] Y Horita, K Shibata, Y Kawayoke, and Z Sazzad. MICT image quality assessment database.
- [36] Howard C. Hughes, George Nozawa, and Frederick Kitterle. Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, 1996.
- [37] Dinesh Jayaraman, Anish Mittal, Anush K. Moorthy, and Alan C. Bovik. Objective quality assessment of multiply distorted images. In *Conference Record - Asilomar Conference on Signals, Systems and Computers*, pages 1693–1697, 2012.
- [38] Frederic Jurie and Bill Triggs. Creating efficient codebooks for visual recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, 2005.
- [39] Le Kang, Peng Ye, Yi Li, and D Doermann. Convolutional Neural Networks for No-Reference Image Quality Assessment. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1733–1740, 2014.
- [40] Vishwakumara Kayargadde and Jean-Bernard Martens. Perceptual characterization of images degraded by blur and noise: model. *Journal of the Optical Society of America A*, 13(6):1178, 1996.
- [41] Jonas Larsson, Michael S. Landy, and David J. Heeger. Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *Journal of Neurophysiology*, 2006.
- [42] G. E. Legge and J. M. Foley. Contrast masking in human vision. *Journal of the Optical Society of America*, 1980.
- [43] Chaofeng Li, Yifan Li, Yunhao Yuan, Xiaojun Wu, and Qingbing Sang. Quaternion wavelet transform based full reference image quality assessment for multiply distorted images. *PLOS ONE*, 13(6):e0199430, jun 2018.

- [44] Chaofeng Li, Yu Zhang, Xiaojun Wu, Wei Fang, and Li Mao. Blind multiply distorted image quality assessment using relevant perceptual features. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 4883–4886. IEEE, sep 2015.
- [45] Chaofeng Li, Yu Zhang, Xiaojun Wu, and Yuhui Zheng. A multi-scale learning local phase and amplitude blind image quality assessment for multiply distorted images. *IEEE Access*, 2018.
- [46] Qiaohong Li, Weisi Lin, and Yuming Fang. No-reference quality assessment for multiply-distorted images in gradient domain. *IEEE Signal Processing Letters*, 23(4):541–545, 2016.
- [47] Weisi Lin and Manish Narwaria. Perceptual image quality assessment: recent progress and trends. In *Visual Communications and Image Processing 2010*, volume 7744, page 774403. International Society for Optics and Photonics, 2010.
- [48] L Linde. Similarity of distorted pictures: on the interaction between edge blur and random noise. *FOA Rep. C*, (53004-H9), 1981.
- [49] Lixiong Liu, Hongping Dong, Hua Huang, and Alan C. Bovik. No-reference image quality assessment in curvelet domain. *Signal Processing: Image Communication*, 29(4):494–505, 2014.
- [50] Xialei Liu, Joost van de Weijer, and Andrew D Bagdanov. RankIQA: Learning from Rankings for No-reference Image Quality Assessment. *arXiv preprint arXiv:1707.08347*, 2017.
- [51] Yanan Lu, Fengying Xie, Tongliang Liu, Zhiguo Jiang, and Dacheng Tao. No reference quality assessment for multiply-distorted images based on an improved bag-of-words model. *IEEE Signal Processing Letters*, 22(10):1811–1815, 2015.
- [52] Kede Ma, Wentao Liu, Kai Zhang, Zhengfang Duanmu, Zhou Wang, and Wangmeng Zuo. End-to-End Blind Image Quality Assessment Using Deep Neural Networks, 2017.
- [53] Mengzhu Ma and Chaofeng Li. Blind multiply distorted image quality assessment using an ensemble random forest. In *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pages 1–5. IEEE, oct 2017.

- [54] Saeed Mahmoudpour and Peter Schelkens. Reduced-reference image quality assessment based on internal generative mechanism utilizing shearlets and Rényi entropy analysis. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–5. IEEE, may 2017.
- [55] Saeed Mahmoudpour and Peter Schelkens. Reduced-reference quality assessment of multiply-distorted images based on structural and uncertainty information degradation. *Journal of Visual Communication and Image Representation*, 2018.
- [56] Redzuan Abdul Manap and Ling Shao. Non-distortion-specific no-reference image quality assessment: A survey. *Information Sciences*, 2015.
- [57] Azadeh Mansouri, Ahmad Mahmoudi Aznaveh, Farah Torkamani-Azar, and J. Afshar Jahanshahi. Image quality assessment using the singular value decomposition theorem. *Optical Review*, 2009.
- [58] Azadeh Mansouri and Ahmad Mahmoudi-Aznaveh. SSVD: Structural SVD-based image quality assessment. *Signal Processing: Image Communication*, 2019.
- [59] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *Proceedings of the IEEE International Conference on Computer Vision*, 2:416–423, 2001.
- [60] Xikui Miao, Hairong Chu, Hui Liu, Yao Yang, and Xiaolong Li. Quality assessment of images with multiple distortions based on phase congruency and gradient magnitude. *Signal Processing: Image Communication*, 2019.
- [61] Xikui Miao, Hairong Chu, Hui Liu, Yao Yang, and Xiaolong Li. Quality assessment of images with multiple distortions based on phase congruency and gradient magnitude. *Signal Processing: Image Communication*, 2019.
- [62] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 21(12):4695–708, 2012.
- [63] Anush Krishna Moorthy and Alan Conrad Bovik. A two-step framework for constructing blind image quality indices. *IEEE Signal Processing Letters*, 17(5):513–516, 2010.

- [64] B. Muraleetharan and K. Thirulogasanthar. Coherent state quantization of quaternions. *Journal of Mathematical Physics*, 2015.
- [65] King N Ngan, Kin S Leong, and Harcharan Singh. Cosine transform coding incorporating human visual system model. In *Visual Communications and Image Processing*, volume 707, pages 165–171. International Society for Optics and Photonics, 1986.
- [66] Timo Ojala, Matti Pietikäinen, and Topi Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.
- [67] K. Okarma and K. Quality Assessment of Images with Multiple Distortions using Combined Metrics. *Elektronika ir Elektrotechnika*, 20(6):128–131, jun 2014.
- [68] Denis G. Pelli and Katharine A. Tillman. The uncrowded window of object recognition, 2008.
- [69] Nikolay Ponomarenko, Lina Jin, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, and C. C. Jay Kuo. Image database TID2013: Peculiarities, results and perspectives. *Signal Processing: Image Communication*, 30:57–77, 2015.
- [70] Nikolay Ponomarenko, Vladimir Lukin, Alexander Zelensky, Karen Egiazarian, Jaakko Astola, Marco Carli, and Federica Battisti. TID2008-A database for evaluation of full-reference visual quality assessment metrics. *Advances of Modern ...*, 10(January 2016):30–45, 2009.
- [71] M. Reenu, Dayana David, S. S. Aneesh Raj, and Madhu S. Nair. Wavelet based sharp features (WASH): An image quality assessment metric based on HVS. In *Proceedings - 2nd International Conference on Advanced Computing, Networking and Security, ADCONS 2013*, 2013.
- [72] Daniel L. Ruderman. The statistics of natural images. *Network: Computation in Neural Systems*, 5(4):517–548, 1994.
- [73] Michele a Saad, Alan C Bovik, and Christophe Charrier. Blind image quality assessment: a natural scene statistics approach in the DCT domain. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 21(8):3339–52, 2012.

- [74] Karnran Sharifi and Alberto Leon-Garcia. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, 5(1):52–56, 1995.
- [75] H R Sheikh, Z. Wang, L. Cormack, and A.C. Bovik. LIVE Image Quality Assessment Database Release 2.
- [76] Hamid Rahim Sheikh, Alan Conrad Bovik, and Gustavo de Veciana. An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 14(12):2117–2128, 2005.
- [77] H.R. Hamid Rahim Sheikh and Alan C. A.C. Bovik. Image information and visual quality. *IEEE Transactions on Image Processing*, 15(2):430–444, 2006.
- [78] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
- [79] Wen Sun, Fei Zhou, and Qingmin Liao. MDID: A multiply distorted image database for image quality assessment. *Pattern Recognition*, 61:153–168, 2017.
- [80] Lijuan Tang, Leida Li, Kezheng Sun, Zhifang Xia, Ke Gu, and Jiansheng Qian. An efficient and effective blind camera image quality metric via modeling quaternion wavelet coefficients. *Journal of Visual Communication and Image Representation*, 49:204–212, nov 2017.
- [81] Robert Tibshirani. Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1996.
- [82] Farah Torkamani-Azar and Jussi Parkkinen. Image quality assessment using block-based weighted SVD. *Signal, Image and Video Processing*, 12(7):1337–1344, 2018.
- [83] Albekaye Traoré, Philippe Carré, and Christian Olivier. Reduced-reference metric based on the quaternionic wavelet coefficients modeling by information criteria. In *2014 IEEE International Conference on Image Processing (ICIP)*, pages 526–530. IEEE, 2014.
- [84] Joost Van De Weijer, Theo Gevers, and Andrew D. Bagdanov. Boosting color saliency in image feature detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006.

- [85] V N Vapnik. *The Nature of Statistical Learning Theory*, 1995.
- [86] VQEG. Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment, 2000.
- [87] Zhongling Wang, Shahrukh Athar, and Zhou Wang. Blind Quality Assessment of Multiply Distorted Images Using Deep Neural Networks. In *International Conference on Image Analysis and Recognition*, pages 89–101. Springer, 2019.
- [88] Zhou Wang and Alan C. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 2002.
- [89] Zhou Wang and Alan C Bovik. Mean Squared Error : Love It or Leave It ? *IEEE Signal Processing Magazine*, 2009.
- [90] Zhou Wang, Alan C. Bovik, and Ligang Lu. Why is image quality assessment so difficult? *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 2002.
- [91] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [92] Jinjian Wu, Weisi Lin, Guangming Shi, Yazhong Zhang, Weisheng Dong, and Zhibo Chen. Visual Orientation Selectivity Based Structure Description. *IEEE Transactions on Image Processing*, 2015.
- [93] Jingtao Xu, Peng Ye, Qiaohong Li, Haiqing Du, Yong Liu, and David Doermann. Blind image quality assessment based on high order statistics aggregation. *IEEE Transactions on Image Processing*, 25(9):4444–4457, 2016.
- [94] Shaoping Xu, Shunliang Jiang, and Weidong Min. No-reference/blind image quality assessment: a survey. *IETE Technical Review*, 34(3):223–245, 2017.
- [95] Wufeng Xue, Xuanqin Mou, Lei Zhang, Alan C Bovik, and Xiangchu Feng. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features. *IEEE Transactions on Image Processing*, 23(11):4850–4862, 2014.
- [96] Wufeng Xue, Lei Zhang, and Xuanqin Mou. Learning without human scores for blind image quality assessment. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 995–1002, 2013.

- [97] Wufeng Xue, Lei Zhang, Xuanqin Mou, and Alan C. Bovik. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, 23(2):668–695, 2014.
- [98] Peng Ye, J Kumar, Le Kang, and D Doermann. Unsupervised feature learning framework for no-reference image quality assessment. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 1098–1105, 2012.
- [99] Guanghui Yue, Chunping Hou, Ke Gu, Nam Ling, and Beichen Li. Analysis of Structural Characteristics for Quality Assessment of Multiply Distorted Images. *IEEE Transactions on Multimedia*, 2018.
- [100] Guangtao Zhai, Jianfei Cai, Weisi Lin, Xiaokang Yang, Wenjun Zhang, and Minora Etoh. Cross-dimensional perceptual quality assessment for low bit-rate videos. *IEEE Transactions on Multimedia*, 2008.
- [101] Fan Zhang, Lin Ma, Songnan Li, and King Ngi Ngan. Practical image quality metric applied to image coding. *IEEE Transactions on Multimedia*, 2011.
- [102] Lin Zhang, Zhongyi Gu, Xiaoxu Liu, Hongyu Li, and Jianwei Lu. Training quality-aware filters for no-reference image quality assessment. *IEEE Multimedia*, 21(4):67–75, 2014.
- [103] Lin Zhang, Ying Shen, and Hongyu Li. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, 2014.
- [104] Lin Zhang, Lei Zhang, Xuanqin Mou, and David Zhang. FSIM: A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20(8):2378–2386, 2011.
- [105] Peng Zhang, Wengang Zhou, Lei Wu, and Houqiang Li. SOM: Semantic obviousness metric for image quality assessment. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 07-12-June, pages 2394–2402, 2015.
- [106] Yi Zhang and Damon M. Chandler. An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes. *Image Quality and System Performance*, 8653:86530J, 2013.

- [107] Yi Zhang, Damon M. Chandler, and Senior Member. Opinion-Unaware Blind Quality Assessment of Multiply and Singly Distorted Images via Distortion Parameter Estimation. *IEEE Transactions on Image Processing*, PP(c):1, nov 2018.
- [108] Yi Zhang, Anush K. Moorthy, Damon M. Chandler, and Alan C. Bovik. C-DIIVINE: No-reference image quality assessment based on local magnitude and phase statistics of natural scenes. *Signal Processing: Image Communication*, 2014.
- [109] Yin Zhang, Xuehan Bai, Junhua Yan, Yongqi Xiao, Wanyi Zhang, C R Chatwin, and R C D Young. A full-reference image quality assessment for multiply distorted image based on visual mutual information. *Journal of Imaging Science and Technology*, 2019.
- [110] Liping Zheng, Hua Jiang, Quanke Pan, and Guangyao Li. Medical image segmentation based on an improved 2D entropy. In *ICCIT 2009 - 4th International Conference on Computer Sciences and Convergence Information Technology*, 2009.
- [111] Wujie Zhou, Lu Yu, Yaguan Qian, Weiwei Qiu, Yang Zhou, and Ting Luo. Deep blind quality evaluator for multiply distorted images based on monogenic binary coding. *Journal of Visual Communication and Image Representation*, 2019.
- [112] Yu Zhou, Leida Li, Jinjian Wu, Ke Gu, Weisheng Dong, and Guangming Shi. Blind quality index for multiply distorted images using biorder structure degradation and nonlocal statistics. *IEEE Transactions on Multimedia*, 2018.
- [113] Daniel Zoran and Yair Weiss. Scale invariance and noise in natural images. In *Proceedings of the IEEE International Conference on Computer Vision*, 2009.