Full-Reference and Non-Reference Image Quality Assessment Based on Optimization Technique

（最適化技術に基づくフルリファレンスおよびノンリファレンス画像品質評価）

Yadanar Khaing

A Dissertation Submitted to
the Graduate School of Science and Engineering
in Partial Fulfillment of the Requirements for the Degree of
DOCTOR OF ENGINEERING
in
Mathematics, Electronics and Informatics

Supervisor: Professor Tetsuya Shimamura

Saitama University, Japan

September 2019

Contents

List of Figures

List of Tables

List of Symbols

$FR-IQA$	full-reference image quality assessment
$BIQA$	blind image quality assessment
IQA	image quality assessment
MOS	Mean Opinion Score
PSO	Particle Swarm Optimization
NSS	Natural Scene Statistics
CNN	convolutional neural network
IR	Image restoration
$SSIM$	Structural Similarity
HVS	Human Visual System
$MS-SSIM$	Multi-scale Structural Similarity
MAD	Most Apparent Distortion
$FSIM$	Feature Based Structural Similarity
PC	phase congruency
GM	gradient magnitude
$NR-IQA$	No-reference Image Quality Assessmen
$BLIINDS$	Blind image integrity notator using DCT statistics
DCT	discrete cosine transform
$BRISQUE$	Blind/Referenceless image spatial quality evaluator algorithm
$PSNR$	peak signal-to-noise ratio
NQM	noise quality measure
UQI	universal image quality index
$IW-SSIM$	information content weighted $SSIM$
VIF	visual information fidelity
IFC	information fidelity criterion
$RFSIM$	Riesz-transform based feature similarity
$SR-SIM$	spectral residual based similarity
VSI	visual saliency induced index
GSM	gradient similarity
$DIIVINE$	Distortion Identification based Image Verity and Integration Evaluation
SVR	support vector regression
$CORNIA$	Codebook Representation for No-reference Image quality Assessment

SOM	semantic obviousness metric
$GSSIM$	Gradient based structural similarity
$VSNR$	Visual signal to noise ratio
VGS	Visual gradient similarity
$LOGPSNR$	log $PSNR$
$PSNR-HVS$	$PSNR$ based on HVS
$DOGSSIM$	Difference of Gaussian $SSIM$
$ESIM$	Edge Similarity
$SURF$	Speed Up Robust Features
n	spatial index
k	the DCT coefficient frequency index
$D[n,k]$	DCT coefficient of a patch
θ	orientation
$\tilde{D}_\theta[n,k]$	Normalized DCT patch
$R_\theta[n]$	the Renyi entropy for the particular image patch
$E[R_\theta]$	the average per orientation for all patches
$var(E[R_\theta])$	the variance across all four orientations
$max(E[R_\theta]))$	the maximum entropy across the four orientations
$BIRD$	Biggest Index Ranking Difference
GA	Genetic Algorithm
μ_{ref}	mean intensity of reference image
μ_{dst}	mean intensity of distorted image
σ_{ref}	standard deviation of reference image
σ_{dst}	standard deviation of distorted image
$l(I_{ref},I_{dst})$	luminance comparison function of reference and distorted images
$c(I_{ref},I_{dst})$	contrast comparison function of reference and distorted images
$s(I_{ref},I_{dst})$	structure comparison function of reference and distorted images
$S(I_{ref},I_{dst})$	overall similarity measure of reference and distorted images
$\sigma_{ref,dst}$	the correlation coefficient between reference and distorted images
α	positive constant
β	positive constant
γ	positive constant
M_s	multi-scale
$c_i(I_{ref},I_{dst})$	constrast comparison function at the $i-th$ scale
$s_i(I_{ref},I_{dst})$	structure comparison function at the $i-th$ scale

$l_{M_s}(I_{ref}, I_{dst})$	luminance comparison function at the $M_s - th$ scale
α_i	positive constant
α_{M_s}	positive constant
β_i	positive constant
γ_i	positive constant
Ł	luminance image
CSF	contrast sensitivity function
rms	root mean square
Z	tatal number of blocks
$C_{ref}(z)$	the rms contrast of reference image for block z
$C_{dst}(z)$	the rms contrast of distorted image for block z
$\mu_{ref}(z)$	the mean block of the z in the reference image
$\tilde{\sigma}_{ref}(z)$	the minimum of the standard deviation of the four sub-blocks in z
$\sigma_{er}(z)$	standard deviation of block z
$\varrho(z)$	local distortion visibility map
δ	threshold value
d_{detect}	the perceived distortion due to visual detection
$\omega(z)$	the local MSE of block z
M_p	the set of pixels in block z
$G_{s,o}(r, \phi)$	2-D log-Gabor filter bank with frequency responses
s	spatial scale
o	spatial orientation
r	normalized radial frequency
ϕ	normallized orientation
r_s	normalized center frequency
$\eta(z)$	local statistical difference map
d_{appear}	perceived distortion
PC_{ref}	phase congruency of reference image
PC_{dst}	phase congruency of distorted image
$S_{PC}(x)$	similarity measure between PC_{ref} and PC_{dst}
$S_G(x)$	similarity measure between G_{ref} and G_{dst}
$S_L(x)$	the final similarity measure between I_{ref} and I_{dst}

Acknowledgements

First and foremost, I would like to express my deep gratitude to my supervisor Prof. Tetsuya Shimamura for his continuous support, patience, kindness, motivation, enthusiasm, and understanding throughout my Ph.D. study. His guidance helped me in all the time of research and writing of this thesis. Moreover, I am very grateful to him for pushing me to do better, to improve myself, and test my limits not only as a student but also as a whole person. Furthermore, I would like to extend my wholehearted thanks to him for giving me an exciting opportunity as his teaching assistant in Saitama University and being able me to balance study and fun during this three years. I could not have imagined having a better advisor and mentor like him.

My sincere thanks also go to Dr. Yosuke Sugiura for sharing his expertise so willingly, and for being so dedicated to his role as my secondary mentor.

I am also grateful to my dissertation committees, Prof. Takashi Komuro, Prof. Yoshinori Kobayashi, Prof. Jun Ohkubo, for their kind support, valuable feedback, and profound ideas to this research work.

Similarly, my intense gratitude goes to JASSO and JGC-S scholarship foundations for their financial support in my journey towards this Ph.D. degree.

I would like to extend thanks to the Ministry of Education, Myanmar that allowed me to study abroad.

My special thanks also go to all lab members for stimulating discussions, advices, encouragement, kindness and all priceless memory we have had throughout this three years.

Finally, my heartfelt thanks go to my parents and siblings for their unconditional love, boundless patience, amazing warmth and endless support.

Abstract

In this study, we proposed some objective image quality estimation methods for full-reference image quality assessment($FR{-}IQA$) and no-reference or blind image quality assessment ($BIQA$). First, we proposed a new combination technique for full-reference image quality assessment ($FR-IQA$) by utilizing three better-recognized image quality assessment (IQA) methods. For selecting the IQA methods, we firstly pick up the most appropriate IQA index for image quality databases and then add other two indices which have the most dissimilar features with the first index. Indeed, the combination of multiple IQA measures naturally emerges because of the shortcomings of single IQA indices for different types of distortions.

Over the last several decades, IQA has been a topic of intense research in image processing field and it is the process of determining the level of accuracy of digital images. Nowadays, huge amount of images are daily produced for several purposes, for example, forecasting weather, finding diseases and monitoring criminals. For these reasons, it is very importance to keep the quality of such images at an acceptable visual level at the end-users after the production and transmission. Furthermore, accurate measurement of the image quality is an important step in many image-based applications. To achieve this goal, effective IQA algorithms are necessary and have recently become a very hot research topic.

Basically, there are two types of image quality assessment called subjective image quality assessment and objective image quality assessment. Subjective Image Quality Assessment is the most reliable way to evaluate the visual quality of digital images perceived by the human observers. In practice, however, subjective image quality assessment is time consuming and very expensive. Thus, it is often used to construct image quality datasets and provide the ground-truth mean opinion scores (MOS) for evaluating objective quality measures. Objective image quality assessment is automatically estimating the quality of images by algorithms instead of humans using MOS provided by human observers and it is more handy than the subjective IQA . To carry out this requirement, many researchers proposed various not only single but also combination IQA methods in recent times. However, all existing single and combination IQA methods still have some shortcomings to be able to get the highest performance for full-reference IQA . Therefore, we consider a simple and robust combination method that are suitable for all image databases. In our combination, we firstly pick up the most correlated IQA method for all types

of distortions by applying the algorithm that is used to select the most appropriate method for combination. After choosing the first combined IQA method, we choose the one which has the biggest index ranking difference with the first one as the second combined IQA method, since it has the most different characteristics comparing to the first chosen combined IQA method. Following the same way, we decide the third one. After selecting the most appropriate methods, we combine the three methods by employing the weighting factors, exponentaited coefficients and constant values. Then, we optimize these parameter values by using the Particle Swarm Optimization (PSO). Experimental results verified that the proposed method gives the best performance for various databases and outperforms other state-of-the-art not only traditional single methods but also previous combination methods.

On the other hand, it is very difficult to get the information of reference images for image quality estimation in reality. This gives a motivation to consider blind image quality assessment ($BIQA$) methods which are able to measure the quality of distorted images without referencing the original images. Therefore, many researchers develop numerous $BIQA$ approaches using Natural Scene Statistics (NSS) based features. In most NSS based $BIQA$ methods, features are extracted by the wavelet transform and they are usually very slow due to the use of computationally expensive image transformations. Thus, more recent techniques promote extracting features from the spatial domain, which leads to a significant reduction in computation time. However, all existing $BIQA$ methods have still restrictions to get the highest performance. To overcome the restrictions, we consider to construct a very simple and robust end-to-end learning mechanism using convolutional neural network (CNN). One of CNN s advantages is that it can take raw images as input and incorporate feature learning into the training process. Thus, in our work, we take distorted images labelled with Mean Opinion Score (MOS) as inputs and output the related score for each image. Experimental results demonstrated that our proposed method outperforms other state-of-the-art ones.

Chapter 1

Introduction

1.1 Background

Digital Image processing involves changing the nature of an image in order to either improve its pictorial information for human interpretation, or render it more suitable for autonomous machine perception [1]. However, the recorded image is often a degraded version of the original scene because of imperfections due to the different processes: acquisition, encoding, decoding, transmission, conversion, quantification and digital storage, etc. As a consequence different variety of degradations and distortions exist that need to be taken into account, covering for instance, essentially blurring effects (e.g. camera motion, out of focus) and noise (e.g. scatter effects, sensor noise), which generally results in loss of visual quality. Image restoration (IR) is one of the most important research fields in image processing that is a part of inverse problems, for which the main purpose is to enhance the quality of an observed image by mitigating the previously mentioned defects and imperfections [2].

Various IR approaches exist in the literature and, for each particular and specific application, a guideline should be provided for the choice of the suitable IR technique, following different criteria: type of degradation, robustness against additive noise, degree of a prior knowledge required, computational complexity and visual quality performance. On the other hand, image quality assessment (IQA) plays an important role in many applications in image processing, such as image acquisition, transmission, enhancement and restoration.

Image quality assessment can be defined as to assess or to measure the quality of an image in accordance or in reference to the original image. For example, in

image compression, if the captured image contains distortions then it would not match with the original image that is stored in the dataset. So, finding the quality of the image in those areas is very necessary. Usually subjective rating methods are used for calculating the quality of the image. In this subjective rating, Observers rated the image quality. The images are given to the experts. Based on the time requirements available, they give scores to the image. Subjective results can provide accurate results, but it is time consuming and also a costly process. This is the reason for the development of objective image quality assessment (IQA) algorithms that will predict the quality of the image automatically. Figure 1.1 is the block diagram of the objective IQA measures.

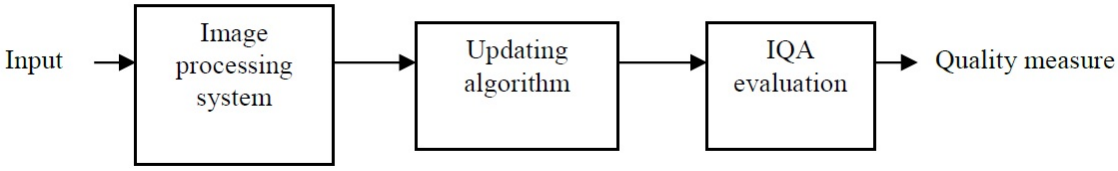


Figure 1.1: The block diagram for objective IQA

1.1.1 Full-reference Image Quality Assessment ($FR-IQA$)

In $FR-IQA$ metrics, the perfect quality reference image is fully available for quality evaluation process. The following are widely cited methods in the literature, and have been reported to have good performance by researchers.

Structural Similarity ($SSIM$)

A sophisticated tool for image quality evaluation is SSIM [6] that measures the similarity between two images and considered to be correlated with the quality perception of the human visual system (HVS). $SSIM$ principle is based on the modeling any image distortion as a combination of luminance distortion, contrast distortion, and loss of correlation. The detailed explanation of $SSIM$ is described in Chapter 2.

Multi-scale Structural Similarity ($MS-SSIM$)

The multi-scale structural similarity measure [41] is a more flexible method than the other single scale methods. The existing $SSIM$ algorithm is a single scale approach. By using this multi-scale method, image details with different resolutions can be incorporated. Low-pass filtering and down sampling are the two main operations used in this method. The original and the distorted images are iteratively low-pass filtered and then down sampling will be done on that by a factor of 2. For this multi-scale operation, the original image is taken as scale 1. The highest scale is scale M, so a total of M-1 iterations are taken place. Similar to the $SSIM$ method, three comparisons have been done here i.e., contrast comparison, luminance comparison and the structure comparison. Only luminance comparison is performed on scale M. Other two are performed on the intermediate scales. The final quality measure is the combination of these three comparisons. This is a convenient image quality metric than the other single scale approaches.

Most Apparent Distortion (MAD)

Most apparent distortion [20] is mainly based on the property of HVS when judging the quality of the image. It consists of two strategies for quantifying the information contained in images. Detection based strategy and appearance based strategy. In detection based strategy, there are two steps that have to be done. First, find the locations in which distortions are visible. This can be done by converting the original and the distorted images into luminance images and then converting it into perceived luminance. Applying a contrast sensitivity function to this will give the visibility map. In the second step, combine this visibility map with the local errors. This detection based strategy mainly used for high quality images. For low quality images detection based strategy can be used. A log-Gabor decomposition method can be adopted for decomposing the original image into different subbands. Comparing the computed local subband statistics would give the result of appearance based strategy. The final prediction can get by combining the two strategies.

Feature Based Structural Similarity ($FSIM$)

The $FSIM$ which is feature based similarity index [12] is mainly based on the low level features of an image. Most of the methods are based on structural similarity. Two features are used here for the calculation of similarity index, phase congruency

(PC) and gradient magnitude (GM) where PC is the primary feature and GM is the secondary feature. PC is a dimensionless measure. For the computation of $FSIM$, PC and GM have to be extracted from the image. The PC extraction can be done by using a log Gabor filter. The gradient operators can be Sobel, Prewitt and Schar operator. The input images can be gray scale or color images. If it is a color image, PC and GM can be extracted from the luminance. One of the main advantage of $FSIM$ is that it can support color images. Figure 1.2 is the block diagram of the $FSIM$ algorithm.

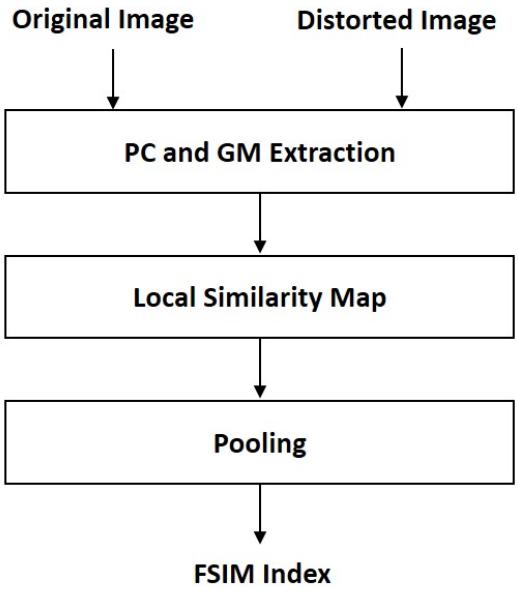


Figure 1.2: The block diagram of the FSIM algorithm

1.1.2 No-reference Image Quality Assessment ($NR-IQA$)

In many real-world applications, such as image communication systems, the reference image is not available and the quality evaluation is solely based on the test image. $NR-IQA$ is a more difficult task in comparison to $FR-IQA$ method. However, human beings usually can efficiently assess the quality of a test image without using any reference image. This is probably due to the fact that our brain holds a lot of information about how an image should or should not look like in real world [67]. Some $NR-IQA$ methods can be found in [68–73].

Blind image integrity notator using DCT statistics algorithm ($BLIINDS$)

An image quality estimation algorithm based on a single-stage DCT framework, known as blind image integrity notator using DCT statistics ($BLIINDS$), is introduced in [74]. In the algorithm, DCT is first employed to 17x17 image patches centered at each pixel in the given image. Four features are then extracted at two scales representing the information of image contrast and image structure. The first feature, i.e., the mean value of the local DCT contrast, is computed by averaging the local DCT contrast values from all image patches. The local DCT contrast is computed as the average of the ratio of the non-DC DCT coefficient magnitudes in the local patch normalized by the DC coefficient of that patch [74]. The second feature is determined by first calculating the kurtosis of each DCT image patch and then averaging the lowest tenth percentile of the computed value. The remaining two features, the DCT coefficient entropy variance across four orientations and the maximum DCT coefficient entropy across four orientations, are defined based on anisotropy measurement on each of the DCT patches. Specifically, DCT image patches are initially generated at four different orientations: 0°, 45°, 90° and 135°. Each DCT patch is then normalized as follows:

$$\tilde{D}_\theta[n,k] = \frac{D_\theta[n,k]^2}{\sum_k D_\theta[n,k]^2} \tag{1.1}$$

where θ is one of the four orientations, $D[n,k]$ denotes DCT coefficients of a certain patch; k indicates the DCT coefficient frequency index ($1 < k \leq 17$), and n represents the spatial index where the patch is computed. Next, the Renyi entropy for the particular image patch is computed as:

$$R_\theta[n] = -\frac{1}{2} log\left(\sum_k \tilde{D}_\theta[n,k]^3\right) \tag{1.2}$$

By defining $E[R_\theta]$ as the average per orientation for all patches of orientation θ, the variance across all four orientations and the maximum entropy across the four orientations are then determined as $var(E[R_\theta])$ and $max(E[R_\theta]))$ respectively [74].

Improved blind image integrity notator using DCT statistics algorithm (BLIINDS-II)

The proposed algorithm, known as BLIINDS-II, provides improvement to the $BLIINDS$ algorithm in the use of a general statistical model and in the performance. The framework of BLIINDS-II algorithm is illustrated in Figure 1.3.

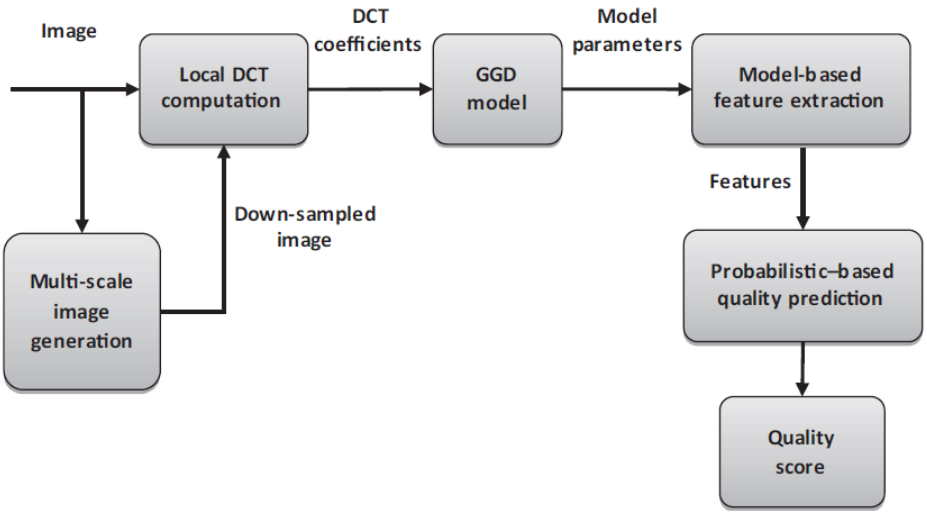


Figure 1.3: BLIINDS-II framework [29]

An input image is first divided into 5x5 blocks with a 2-pixel overlap between the blocks. Then, the blocks are subjected to local DCT computation where the coefficient extraction is carried out locally in the spatial domain. The DCT block is then partitioned into two configurations: first, the block is partitioned into three oriented sub-regions to indicate directional information and second, the block is partitioned in such a way to reflect the low-frequency, mid-frequency and high-frequency sub-bands respectively. While sample statistics such as kurtosis and entropy are used as features in the previous $BLIINDS$ algorithm, the BLIINDS-II algorithm refines the technique by modelling image features using a generalized Gaussian distribution.

Blind/Referenceless image spatial quality evaluator algorithm ($BRISQUE$)

An NSS -based $BIQA$ algorithm that works in the spatial domain is presented in [30]. In contrast to previous $NR-IQA$ algorithms, no transformation (DCT , DWT, etc.) is required for the algorithm referred as blind/reference less image

spatial quality evaluator ($BRISQUE$). Using a spatial NSS framework, the empirical distribution of locally normalized luminance coefficients and pairwise products of these coefficients is utilized to design 18 features for image quality prediction. These features are selected based on observations that the coefficients possess statistical properties which are varied with the existence of distortion and the signs of adjacent coefficients also show a regular structure which can change in the presence of distortion. By quantifying these changes, it is assumed that the type of distortion affecting the image and its perceptual quality can be predicted.

1.2 Problem Statements and Objectives

Today, IQA research has emerged as an active subdiscipline of image processing, and numerous algorithms for IQA have been researched and developed over the last several decades. Furthermore, many of the resulting techniques and algorithms have begun to benefit a wide variety of applications. Variations of IQA algorithms have proved useful for applications such as image and video coding, digital watermarking, unequal error protection, image denoising, image synthesis, and various other areas (e.g., for predicting intelligibility in sign language video). Though the remarkable progress has been made, we still have yet to achieve full evaluations of quality up to now.

According to the availability of the original reference image, the objective methods are classified into: (i) full-reference where the quality is estimated using the original image and its degraded version, (ii) reduced-reference where only some informations of the original image are used and (iii) no-reference (or blind), where only the degraded image is exploited [3].

All three types of IQA algorithms can perform quite well at predicting quality. Some of today's best-performing full-reference algorithms have been shown to generate estimates of quality that correlate highly with human ratings of quality, typically yielding Spearmans and Pearsons correlation coefficients in excess of 0.9. Research in no-reference and reduced-reference IQA is much less mature; however, recent methods have been shown to yield quality estimates which also correlate highly with human ratings of quality, sometimes yielding correlation coefficients which rival the most competitive full-reference methods. However, a perfect method that performs well with all types of distortions for any database is still an open issue. Thus, considering on the availability of reference information of input images, in this study we intends at:

- Developing a combination method of dissimilar feature-scores for full-reference image quality assessment ($FR-IQA$) using particle swarm optimization (PSO) algorithm.
- Developing an end-to-end learning mechanism using convolutional neural network (CNN) for blind image quality assessment ($BIQA$).

1.3 Overviews

Until recently, many $FR-IQA$ and $NR-IQA$ ($BIQA$) methods have been proposed.

In $FR-IQA$, peak signal-to-noise ratio ($PSNR$) is the simplest approach for IQA but it is not well correlated with human evaluation. This technique, therefore, often serves as a bottom model for comparison. In [4], Damera-Venkata et al. introduced noise quality measure (NQM) in which a distorted image is modelled using a linear frequency distortion and an additive noise injection. In [5], Wang et al. proposed universal image quality index (UQI) and it evaluates quality of an image using loss of correlation, luminance distortion, and contrast distortion. Further extension of UQI , structural similarity ($SSIM$), was proposed by Wang et al. [6]. A multi-scale $SSIM$, $MS-SSIM$, was presented in [41]. Wang and Li in [8] proposed information content weighted $SSIM$ ($IW-SSIM$) approach as an extension of $MS-SSIM$. In their work, local information was measured using statistical models of natural scenes. Statistical properties of natural environment are also utilized in visual information fidelity (VIF) [9] measure and information fidelity criterion (IFC) [10]. In [11], Riesz-transform based feature similarity ($RFSIM$) was proposed and it is computed by comparing Riesz-transform features at key locations between the distorted image and its reference image. Authors of feature similarity index ($FSIM$) [12] developed an approach which uses phase congruency and image gradient magnitude as low-level local features. In [13], spectral residual based similarity ($SR-SIM$) using visual saliency map was proposed. A visual saliency to calculate a local quality map of the distorted image is used in visual saliency-induced index (VSI) [14]. The gradient similarity (GSM) measure [15] estimates image quality taking into consideration structure and contrast changes, as well as luminance distortions. In [16], image structural degradation was considered and determined using local binary patterns. In SURF-SIM [17], multi-scale differences between features detected and described

by Speed Up Robust Features ($SURF$) approach are combined with a pooling strategy. An IQA measure that evaluates images taking into account inter-patch and intra-patch similarities was described in [18]. In their work, authors used modified normalized correlation coefficient and image curvature.

Development of $FR-IQA$ measures can also involve different combination strategies. For example, Liu and Yang [19] combined SNR, $SSIM$, VIF , and $VSNR$ using canonical correlation analysis. A most apparent distortion algorithm (MAD) [20] adopts two strategies for IQA in which a local luminance and a contrast masking evaluate high-quality images. Changes in the local statistics of spatial-frequency components are used for images with a low quality. Three IQA metrics, $MS-SSIM$, VIF and R-SVD, were non-linearly combined by Okarma in [21,22]. A non-linear fusion of IQA measures was also investigated in [23]. In [24], up to seven IQA models were combined using a regularized regression. Peng and Li in [25] presented an approach based on conditional Bayesian mixture of experts model in which a support vector machines classifier was used for prediction of the type of distortion, and then $SSIM$, $VSNR$, and VIF with k-nearest-neighbour regression were fused. In [26], the authors presented adaptive combination of IQA measures with an edge-quality based on preservation of edge direction. In [60], a combination of local and global distortion measures was considered using saliency maps, gradient and contrast information. Recently, many complex combination approaches have been introduced, and therefore, in this study, we proposed a nonlinear combination method employing Particle Swarm Optimization (PSO) [52] algorithm to develop a well-performing method which combined dissimilar feature-scores for $FR-IQA$.

On the other hand, we also considered to measure the perceptual quality of an image without accessing to the reference image because we may not get the original image in practice. A typical approach to $NR-IQA$ is to model statistics of natural images and regress parametric deviations from this model to perceived image degradations. As these parameters and its deviations may depend on the distortion type, the $DIIVINE$ framework [28] identifies the distortion type affecting an image in a first step and uses a distortion-specific regression scheme to estimate the perceived quality in a second step. The statistical features are calculated based on an oriented subband decomposition. BLIINDS-II [29] uses a generalized Gaussian density function to model block DCT coefficients of images. $BRISQUE$ [30] proposes a $NR-IQA$ approach that utilizes an asymmetric generalized Gaussian distribution to model images in the spatial domain. The modeled image features

here are differences of spatially neighbored, mean subtracted and contrast normalized image samples. NIQE [31] extracts features based on a multivariate Gaussian model and relates them to perceived quality in an unsupervised manner. In order to cope with more complex and authentic distortion types FRIQUEE [32], [33] employs a deep belief network of 4 layers trained to classify bins of 10 different distortion ranges. Input to the network is a set of handcrafted feature maps and the feature representation on the last hidden layer is extracted to be input to support vector regression (SVR) for quality prediction. $CORNIA$ [34] is one of the first purely data-driven $NR-IQA$ methods combining feature and regression training. Here, a codebook is constructed by k-means clustering of luminance and contrast normalized image patches. Soft-encoded distances between visual codewords and patches extracted from distorted images are used as features that are pooled and regressed using SVR for estimating image quality. This approach is refined to the semantic obviousness metric (SOM) [35], where object-like regions are detected and the patches extracted from these regions are input to $CORNIA$. Similarly to $CORNIA$, QAF [36] constructs a codebook using sparse filter learning based on image log-Gabor responses. As log-Gabor responses are often considered a low level model of the HVS , conceptually, QAF also belongs to the bottom-up domain. Motivated by the recent success of CNNs for classification and detection tasks and the notion that the connectivity patterns in these networks resemble those of the primate visual cortex, [37] proposes a shallow CNN consisting of 1 convolutional layer, 1 pooling layer and 2 fully-connected layers, that combines feature extraction and regression. Quality is estimated on contrast normalized image patches and patchwise quality is pooled to imagewise quality by averaging. BIECON [38] proposes an interesting approach for data augmentation and tackles CNN -based $NR-IQA$ in 2 steps: First, a local quality is estimated based on normalized image patches employing a CNN of 2 convolutional, 2 pooling and 5 fully-connected layers. This network is trained to replicate a conventional $FR-IQA$ such as $SSIM$ or GMSD within a no-reference framework. Second, mean values and the standard deviations of the extracted patchwise features are regressed to an imagewise quality estimate employing a perceptron with one hidden layer. Preliminary results on the application of deeper neural networks, trained end-to-end, for IQA have been presented in [39], [40]. However, all existing $BIQA$ methods have still shortcomings to get the highest performance. Thus, in this study, we also proposed a very simple and robust end-to-end learning mechanism using CNN for $BIQA$.

1.4 Organization of the thesis

The background, problem statements and objectives, and also overviews of this study have been mentioned in this chapter. The rest of the thesis is organized as follows. In Chapter 2, the combination method of dissimilar feature-scores for $FR-IQA$ is proposed. In Chapter 3, an end-to-end learning mechanism using CNN for $BIQA$ is described. Finally, the concluding remarks and remaining issues are drawn in Chapter 4.

Chapter 2

Combination of Dissimilar Feature-Scores for Image Quality Assessment Using Particle Swarm Optimization Algorithm

In this chapter,a new combination technique for full-reference image quality assessment (IQA) is proposed by utilizing three better-recognized IQA methods. For selecting the IQA methods, we firstly pick up MAD as the most appropriate IQA index for image quality databases and then add other two indices $MS-SSIM$ and $FSIM$ which have the most dissimilar features with the first index MAD . The parameter values employed in the new IQA score are optimized using the particle swarm optimization algorithm. By experiments, it is validated that the proposed method gives the best performance for various databases and outperforms the other state-of-the-art methods.

2.1 Related Work

Nowadays, we are sharing photos via social media, sending and receiving photo messages and transmitting live videos everyday for various reasons. These media facilities are feasible through digital cameras and photo editing systems. Digital images are, however, degraded by various types of distortions during processing. Thus, we need to measure the quality of the images by image quality assessment (IQA). To fulfil this requirement, numerous methods for IQA have been searched

and proposed over the last two decades.

The objective IQA is more handy than the subjective IQA . We can easily adjust the parameters of the image processing system by utilizing the objective IQA value in the system. Traditional IQA measures; mean squared error (MSE) and peak signal-to-noise ratio ($PSNR$) are the simplest and widely used. However, these are not so matched to human eyes. In 2004, structural similarity (SSIM) [6] was designed to improve the traditional metrics such as $PSNR$ and MSE. SSIM fails to give a satisfactory correlation with human visual system (HVS) in the case of blurred images. The improved version of SSIM, multi-scale SSIM ($MS-SSIM$) [41], has better quality prediction accuracy compared to the original SSIM. Gradient based structural similarity ($GSSIM$) [42] is also an improved version of SSIM where the contrast and structure components of SSIM are replaced by the gradient based contrast and structure components, respectively. $GSSIM$ provides better performance than SSIM especially for blurred images. Visual information fidelity (VIF) [9] is based on the amount of information shared by the reference and distorted images. VIF outperforms many of the existing full reference IQA algorithms. However, the main drawback of VIF is its computational complexity. Visual signal to noise ratio ($VSNR$) [43] is a wavelet based approach. It is good for full reference images but its computation is quite complex.

Visual gradient similarity (VGS) [44] assesses the image quality by using magnitudes and directions of gradient vectors and evaluates contrast changes by using the intensity of gradient vectors. VGS is effective for image denoising, contrast change and JPEG compression images but weak for local block-wise distorted images. In another metric $LOGPSNR$ [45], images are filtered by the Laplacian of Gaussian (LOG) filter and then the image quality is measured by $PSNR$. $LOGPSNR$ is effective for Gaussian noise, high frequency noise and Gaussian blur images but weak for intensity shift images.

In recent times, furthermore, several new image quality measures have been proposed as better alternative indices. Examples of twelve best-recognized IQA measures up to now are $SSIM$ [6], $MS-SSIM$ [41], VIF [9], $VSNR$ [43], VGS [44], $LOGPSNR$ [45], MAD [20], $FSIM$ [12], NQM [4], IFC [9], $PSNR$, $PSNR-HVS$ [47]. Each matrix provides an improvement relative to the traditional metrics, but the degree of improvement is limited by their insufficient consideration of HVS properties. This results in the fact that each single index has some shortcomings for certain distortion types. Thus, development of the fusion of multiple IQA techniques emerges by nature. The two scores combination

as shown in [45] is effective for some distortion types of TID2008 database but it cannot assess Meanshift images accurately. Even if one more score is added to [45] as in [48], the performance of the three scores combination is a slight improvement relative to the two scores combination. Furthermore, these combination methods cannot work well on other different databases. Thus, a combination method that performs well with all types of distortions for any database is still an open issue.

Among twelve best-recognized IQA indices described in above, MAD is the most appropriate index for all six publicly available image quality databases [49]. In MAD , local luminance and contrast masking are used to estimate detection-based perceived distortion in high-quality images, whereas changes in the local statistics of spatial-frequency components are used to estimate appearance-based perceived distortion in low-quality images. Feature similarity ($FSIM$) [12] is used to understand low-level features which are minor details of the images like lines or dots. Low-level features convey important visual information and are crucial to image understanding. Intuitively, in the combination of multiple methods as a new score, combining of methods which have similar features cannot produce significant profit comparing to the original single index. For instance, the two scores combination of $PSNR$ and $PSNR - HVS$ cannot give us the better performance comparing to each single index because both indices can predict well on only additive noise distortion type images.

Recent development of full-reference IQA measures involves different combination strategies. M. Liu and X. Yang [19] combined SSIM, SNR, $VSNR$ and VIF using canonical correlation analysis. K. Okarma [21, 22] non-linearly combined three IQA metrics, $MS - SSIM$, VIF and R-SVD [50], as a full-refeference image quality metric. In [25], P. Peng and Z. N. Li proposed an approach based on conditional Bayesian mixture of experts model which utilized a support vector machines classifier to predict the distortion type and combined SSIM, $VSNR$ and VIF with k-nearest-neighbour regression. In [24], up to seven IQA models were combined using a regularized regression. In 2013, multi-method fusion technique was introduced in [49] and M. Oszust developed linearly combined similarity measures in [51]. Although these combinations achieved good evaluation results, the combinations of different numbers of different IQA measures for each database are not straightforward in practice. As far as we know, therefore, all existing combination methods still have some shortcomings to be able to get the highest performance for $FR - IQA$.

To be able to overcome this major challenges of current research trend, in

this chapter, we introduce a novel idea of combination of dissimilar feature-scores applying Particle Swarm Optimization (PSO) [52] algorithm. We firstly select MAD index as the most correlated method for all types of distortions applying the Biggest Index Ranking Difference ($BIRD$) algorithm [49] that is used to select the most appropriate method for combination. After choosing the first combined IQA method, we choose the one which has the biggest index ranking difference with the first one as the second combined IQA method, since it has the most different characteristics comparing to the first chosen combined IQA method. Following the same way, we decide the third, the fourth, and the fifth combined IQA methods and so on. Although we experimented the performance comparison for combination of up to twelve scores, the number of combined IQA scores which is the best for all databases is three to balance the performance and complexity. Hence, in our combination work, we combine MAD index with other two indices, $MS-SSIM$ and $FSIM$, which have the most dissimilar features with MAD by employing the exponentiated coefficients and weighted constant values. Furthermore, we optimize these values by using the PSO algorithm which is based on the swarm intelligence. Amazingly, the calculation of PSO is very simple and the speed of searching is very fast. Compared with the other developing calculations, as an example, Genetic Algorithm (GA), it occupies the bigger optimization ability and it can be completed easily. Thus, the main contribution of the proposed method is that it is very simple but the performance is incredible compared to other state-of-the-art ones.

The organization of this chapter is as follows. Section 2.2 to 2.7 explains about the $SSIM$, $MS-SSIM$, MAD , $FSIM$, $FSIM_C$ and PSO algorithm, respectively. Section 2.8 presents the proposed method in detail. In Section 2.9, we evaluate the performance of the proposed method by experiments. Finally, Section 2.10 concludes the chapter.

2.2 Structural Similarity (SSIM) Measure

The $SSIM$ algorithm [6] assumes that HVS is highly adapted for extracting structural information from a scene. Thus, this algorithm attempts to model the structural information of an image. The $SSIM$ algorithm is based on the fact that pixels of a natural image demonstrate strong dependencies and these dependencies carry useful information about the structure of a scene. Therefore, a method that is capable of measuring structural information change can provide a good approx-

imation of perceived image distortion. The $SSIM$ algorithm defines image degradation as perceived change in structural information. In [6], it is stated that the structure of the objects in a scene is independent of local luminance and contrast. Thus, to extract the structural information, the effect of illumination should be separated. In this algorithm, structural information in an image is defined as those traits that represent the structure of objects in that image, independent of the local luminance and contrast. The $SSIM$ algorithm performs similarity measurement in three steps: luminance comparison, contrast comparison, and structure comparison. Firstly, the luminance of each image signal is compared. The estimated mean intensity is computed as follows:

$$\mu_{ref} = \frac{1}{MN} \sum_{j=1}^{N} \sum_{i=1}^{M} I_{ref}(i,j) \tag{2.1}$$

The luminance comparison function, $l(I_{ref}, I_{dst})$, is a function of μ_{ref} and μ_{dst} .

Secondly, the contrast of each image signal is compared. For estimating the contrast, standard deviation is being used. An unbiased estimate of standard deviation in discrete form is as follows:

$$\sigma_{ref} = \sqrt{\frac{1}{MN-1} \sum_{j=1}^{N} \sum_{i=1}^{M} (I_{ref}(i,j) - \mu_{ref})^2} \tag{2.2}$$

The contrast comparison function, $c(I_{ref}, I_{dst})$, is a function of σ_{ref} and σ_{dst} . Thirdly, the structure of each image signal is compared. Structure comparison function, $s(I_{ref}, I_{dst})$, is a function of $[I_{ref} - \mu_{ref}]/\sigma_{ref}$ and $[I_{dst} - \mu_{dst}]/\sigma_{dst}$. Finally, three comparison functions are combined and an overall similarity measure is produced. $S(I_{ref}, I_{dst})$, is the overall similarity measure of $l(I_{ref}, I_{dst})$, $c(I_{ref}, I_{dst})$ and $s(I_{ref}, I_{dst})$. Definitions of $l(I_{ref}, I_{dst})$, $c(I_{ref}, I_{dst})$ and $s(I_{ref}, I_{dst})$ are as follows:

$$l(I_{ref}, I_{dst}) = \frac{2\mu_{ref}\mu_{dst} + T_1}{\mu_{ref}^2 + \mu_{dst}^2 + T_1} \tag{2.3}$$

$$c(I_{ref}, I_{dst}) = \frac{2\sigma_{ref}\sigma_{dst} + T_2}{\sigma_{ref}^2 + \sigma_{dst}^2 + T_2} \tag{2.4}$$

$$s(I_{ref}, I_{dst}) = \frac{\sigma_{ref,dst} + T_3}{\sigma_{ref}\sigma_{dst} + T_3} \tag{2.5}$$

where T_1, T_2 and T_3 are positive stabilizing constants to prevent the denominator

from becoming too small. They are given by:

$$T_1 = (t_1 D)^2, T_2 = (t_2 D)^2, T_3 = T_2/2 \tag{2.6}$$

where t_1 and t_2 are small scalar constants. In (2.5) $\sigma_{ref,dst}$ is the correlation coefficient between the reference and distorted images. In the discrete form, $\sigma_{ref,dst}$ can be estimated by:

$$\sigma_{ref,dst} = \frac{1}{MN-1} \sum_{j=1}^{N} \sum_{i=1}^{M} (I_{ref}(i,j) - \mu_{ref})(I_{dst}(i,j) - \mu_{dst}) \tag{2.7}$$

Finally, structural similarity index is defined as:

$$SSIM(I_{ref}, I_{dst}) = [l(I_{ref}, I_{dst})]^{\alpha} [c(I_{ref}, I_{dst})]^{\beta} [s(I_{ref}, I_{dst})]^{\gamma} \tag{2.8}$$

where α , β and γ are positive constants chosen to indicate the relative importance of each component. The universal quality index (UQI) [52,53] is a special case of the $SSIM$ index when: $T_1 = T_2 = T_3 = 0$ and $\alpha = \beta = \gamma = 1$. The block diagram of the $SSIM$ algorithm is presented in Figure 2.1.

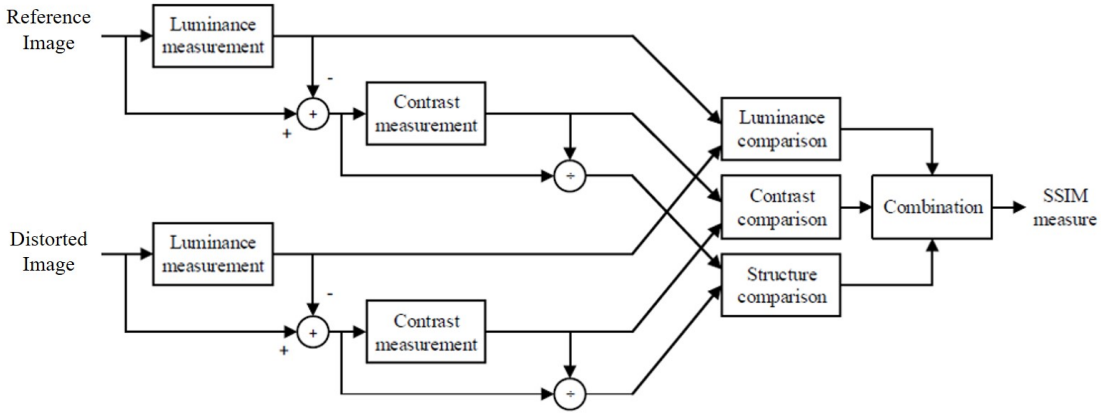


Figure 2.1: The block diagram of the SSIM algorithm

2.3 Multi-scale structural similarity (MS-SSIM) Measure

The $SSIM$ algorithm described earlier is considered a single-scale approach that achieves its best performance when applied at an appropriate scale. Furthermore,

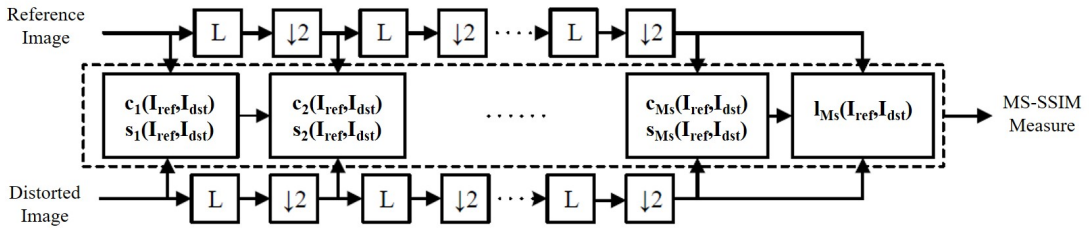


Figure 2.2: The block diagram of the MS-SSIM algorithm. L: low-pass filter; ↓2: downsampling by factor of 2.

choosing the right scale depends on the viewing conditions, e.g., viewing distance and the resolution of the display. Therefore, this algorithm lacks the ability to adapt to these conditions. This drawback of the $SSIM$ algorithm motivated researchers to design a multi-scale structural similarity index ($MS-SSIM$) [41]. The advantage of the multi-scale methods, like $MS-SSIM$, over single-scale methods, like $SSIM$, is that in multi-scale methods image details at different resolutions and viewing conditions are incorporated into the quality assessment algorithm. The block diagram of the $MS-SSIM$ algorithm is presented in Figure 2.2. After taking the reference and distorted images as input, this algorithm performs low-pass filtering and downsampling (by factor of 2) in an iterative manner. At each scale, (2.4) and (2.5) are calculated. However, (2.3) is computed only at M_s $-th$ scale. The final $MS-SSIM$ index is calculated using the following equation:

$$MS-SSIM(I_{ref},I_{dst}) = [l_{M_s}(I_{ref},I_{dst})]^{\alpha_{M_s}} \prod_{i=1}^{M_s} [c_i(I_{ref},I_{dst})]^{\beta_i} [s_i(I_{ref},I_{dst})]^{\gamma_i} \tag{2.9}$$

where $c_i(I_{ref},I_{dst})$ and $s_i(I_{ref},I_{dst})$ are the contrast and the structure comparison function at the $i-th$ scale respectively, and $l_{M_s}(I_{ref},I_{dst})$ is the luminance comparison function at the M_s-th scale. Moreover, α_{M_s} , β_i and γ_i are positive constants chosen to indicate the relative importance of each component. In [41], $\alpha_i = \beta_i = \gamma_i$ for all j, and $\sum_{i=1}^{M_s} \gamma_i = 1$.

2.4 Most apparent distortion (MAD) Measure

MAD algorithm [20] assumes that HVS employs different strategies when estimating the quality of images. It is described in [20] that when HVS attempts to view images containing near-threshold distortions, it tries to move past the image,

looking for distortions. This approach is called detection-based strategy. Moreover, it is also stated in [20] that when HVS attempts to view images containing clearly visible distortions, it tries to move past the distortions, looking for image's subject matter. This approach is called appearance-based strategy. For estimating distortions in detection-based strategy, local luminance and contrast masking are used. Moreover, for estimating distortions in appearance-based strategy, variations in local statistics of spatial frequency components are being employed.

Detection-based strategy

It is argued in [20] that when HVS views high quality images, it tries to look beyond image's subject matter, looking for distortions. Detection-based strategy consists of two stages: determining the locations of visible distortions, and computing perceived distortion due to visual detection. First, the locations of visible distortions should be determined. In order to describe the non-linear relationship between pixel values and physical luminance of display device, MAD algorithm primarily transforms pixels of the reference and distorted images to luminance values using the following equation:

$$Ł = (\beta + \alpha I)^{\gamma} \tag{2.10}$$

where Ł is the luminance image, I is the reference (or distorted) image, and β, α and γ are device specific constants. Applying (2.10) to I_{ref} and I_{dst} yields $Ł_{ref}$ and $Ł_{dst}$ respectively. Since HVS has a non-linear response to luminance, it should be converted to perceived luminance via:

$$\hat{Ł} = \sqrt[3]{Ł} \tag{2.11}$$

where $\hat{Ł}$ denotes perceived luminance. Applying (2.11) to $Ł_{ref}$ and $Ł_{dst}$ results $\hat{Ł}_{ref}$ and $\hat{Ł}_{dst}$ respectively. After computing perceived luminance, an error image is computed:

$$\hat{Ł}_{er} = \hat{Ł}_{ref} - \hat{Ł}_{dst} \tag{2.12}$$

To describe variations in sensitivity due to spatial frequency, authors of [20] employ contrast sensitivity function (CSF) as introduced in [90] with adjustments as in [91]. CSF is applied to both, the reference and error images which yields I'_{ref} and I'_{dst} respectively. Since presence of an image's content can affect the detection of distortions, a spatial domain measure of contrast masking is employed. To

model this, first I'_{ref} is divided into blocks of size 16x16 with 75 percent overlap between neighboring blocks. Afterward, rms contrast (in the lightness domain) of each block is calculated. The rms contrast for block z of I'_{ref} is calculated by:

$$C_{ref}(z) = \tilde{\sigma}_{ref}(z)/\mu_{ref}(z) \tag{2.13}$$

where $\mu_{ref}(z)$ is the mean of block z in the reference image, and $\tilde{\sigma}_{ref}(z)$ is the minimum of the standard deviation of the four sub-blocks in z . The same procedure is done for I'_{er} with the exception that the rms contrast for this image is calculated using the following equation:

$$C_{er}(z) = \begin{cases} \sigma_{er}(z)/\mu_{ref}(z) & , \quad \mu_{ref} > 0.5 \\ 0 & , \quad otherwise \end{cases} \tag{2.14}$$

where $\sigma_{er}(z)$ is the standard deviation of block z in I'_{er}. In (2.14), the threshold of 0.5 denotes the fact that HVS is relatively insensitive to changes in extremely dark regions. After computing $C_{ref}(z)$ and $C_{dst}(z)$, a local distortion visibility map $\varrho(z)$, is computed as follows:

$$\varrho(z) = \begin{cases} ln(C_{er}(z)) - ln(C_{ref}(z)) & , \quad ln(C_{er}(z)) > ln(C_{ref}(z)) > \delta \\ ln(C_{er}(z)) - \delta & , \quad ln(C_{er}(z)) > \delta \geq ln(C_{ref}(z)) \\ 0 & , \quad otherwise \end{cases} \tag{2.15}$$

where δ is a threshold value (δ=-5, as in [20]). Second, the perceived distortion due to visual detection (d_{detect}) is calculated. d_{detect} is calculated via:

$$d_{detect} = \sqrt[2]{\frac{1}{Z}\sum_{z}\left[\varrho(z)\omega(z)\right]^2} \tag{2.16}$$

where Z is the total number of blocks, and $\omega(z)$ is the local MSE of block z of size 16x16, that can be calculated using the following equation:

$$\omega(z) = \frac{1}{16 \times 16}\sum_{i,j \epsilon M_p}(I'_{er}(i,j))^2 \tag{2.17}$$

where M_p is the set of pixels in block z. d_{detect} takes its values in the interval $[0, \infty)$. If d_{detect}=0, there are no visible distortions in the test (distorted) image. As the value of d_{detect} increases, perceived distortion increases and consequently, visual quality decreases. The block diagram of the detection-based strategy is presented

in Figure 2.3.

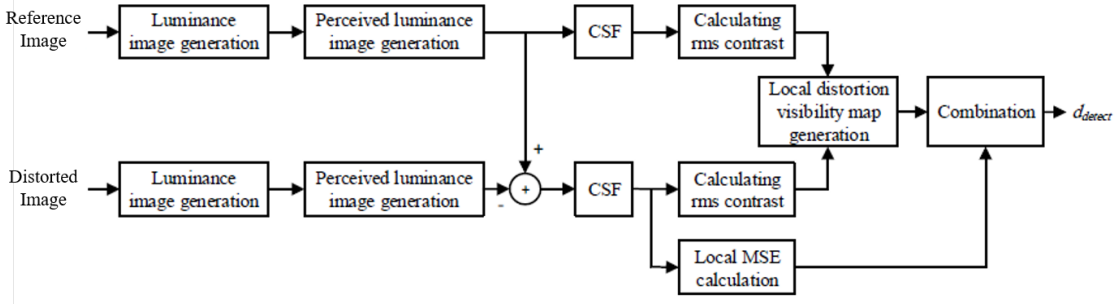


Figure 2.3: The block diagram of the detection-based strategy in the MAD algorithm. [92]

Appearance-based strategy

It is argued in [20] that when viewing low quality images, HVS tries to move past the distortions, looking for image's content. To model this strategy, MAD algorithm uses log-Gabor filter responses. Similar to detection-based strategy, this strategy is also consists of two stages; log-Gabor decomposition of the reference and distorted images, and computing the local statistical difference map. First, the reference and distorted images are decomposed into number of sub-bands via a 2-D log-Gabor filter bank with frequency responses of the form:

$$G_{s,o}(r,\phi) = exp\left[-\frac{log\left(\frac{r}{r_s}\right)}{\sqrt{2}\sigma_r}\right]^2 \times exp\left[-\frac{(\phi-\phi_0)^2}{2\sigma_0^2}\right] \tag{2.18}$$

where indices s and o correspond to spatial scale and orientation respectively, parameters r and ϕ are normalized radial frequency and orientation respectively, r_s is normalized center frequency, σ_r controls the filter's bandwidth, and ϕ_0 and σ_0 are center orientation and angular spread of the filter respectively. In [20], five scales ($s = 1, 2, ..., 5$) and four orientations ($o = 1, 2, ..., 4$) are used for log-Gabor decomposition, which result in 20 sub-bands per image. Second, a local statistical difference map, $\eta(z)$, is generated. This map is defined by comparing local sub-band statistics of the reference image with those of the distorted image. For each block of size 16x16, $\eta(z)$ is calculated by:

$$\eta(z) = \sum_{s=1}^{5}\sum_{o=1}^{4} l_s \left\{\left|\sigma_{s,o}^{ref}(z) - \sigma_{s,o}^{dst}(z)\right| + 2\left|\varsigma_{s,o}^{ref}(z) - \varsigma_{s,o}^{dst}(z)\right| + \left|\kappa_{s,o}^{ref}(z) - \kappa_{s,o}^{dst}(z)\right|\right\} \tag{2.19}$$

where $\sigma_{s,o}(z)$, $\varsigma_{s,o}(z)$, and $\kappa_{s,o}(z)$ correspond to standard deviation, skewness, and kurtosis of 16x16 sub-band coefficients associated with scale s, orientation o, and block z. In (2.19), l_s is a scale specific weight which takes into account the preference of HVS for coarser scales over fine ones. (in [20], l_s= 0.5, 0.75, 1, 5, and 6 for finest to coarsest scales, respectively). After computing $\eta(z)$, a final scalar value of perceived distortion, d_{appear} , is calculated as follows:

$$d_{appear} = \sqrt{\frac{1}{Z}\sum_{z}\eta^2(z)} \tag{2.20}$$

d_{appear} takes its values in the interval $[0, \infty)$. If $d_{appear} = 0$, there is no perceived distortion in the test image. As the value of d_{appear} increases, perceived distortion increases and consequently, visual quality decreases. The block diagram of appearance-based strategy is shown in Figure 2.4.

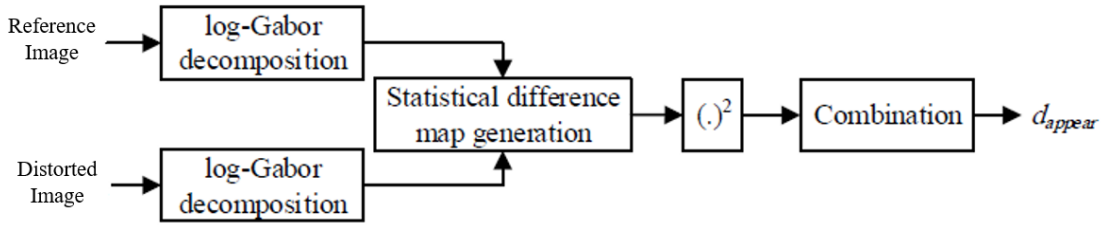


Figure 2.4: The block diagram of the appearance-based strategy in the MAD algorithm. [92]

After computing d_{detect} and d_{appear}, these two values are combined to yield an overall measure of perceived distortion. In [20], it is hypothesized that HVS uses a combination of detection based strategy and appearance based strategy for assessing the quality of images. To model the relation between these two strategies, a weighted geometric mean of d_{detect} and d_{appear} is employed that has the form:

$$MAD = (d_{detect})^{\alpha} (d_{appear})^{1-\alpha} \tag{2.21}$$

where α is a weighting constant chosen to reflect the relative importance of each term. MAD measure takes its values in the interval $[0, \infty)$. It is argued that selecting a value for α based on d_{detect} can yield good overall performance [20]. Therefore, α is calculated by:

$$\alpha = \frac{1}{1 + \beta_1 (d_{detect})^{\beta_2}} \tag{2.22}$$

where β_1 and β_2 are two constants chosen in a way that achieves the best performance of the MAD algorithm in terms of quality prediction accuracy.

2.5 Feature similarity (FSIM) Measure

The $FSIM$ algorithm [12] is based on the HVS 's perceptual characteristics because HVS understands low-level characteristics of an image, e.g., edges, dots and circles and so on [83–85]. In order to assess the quality of an image, $FSIM$ algorithm uses two kinds of features. Physiological and psychophysical experiments have demonstrated that at points with high phase congruency (PC), HVS can extract highly informative features [85–89]. Therefore, PC is used as the primary feature in the $FSIM$ algorithm. However, PC is contrast invariant and our perception of an image's quality is also affected by local contrast of that image. As a result of this dependency, the image gradient magnitude (GM) is used as the secondary feature in the $FSIM$ algorithm. Calculating $FSIM$ measure consists of two stages: computing image's PC and GM , and computing the similarity measure between the reference and test images.

PC and GM computation

The PC model states that Fourier components with maximum phase contain the points where features are perceived by HVS . This model provides a simple structure on how mammalian visual system handles detection and identification of features in an image [85–89]. First, by applying (2.18) to the reference and distorted images, a set of response vectors are created at location x, scale s , and orientation o . Second, the local amplitude of these vectors at scale s and orientation o is calculated. Moreover, the local energy at orientation o is computed. Finally, the PC value at location x is calculated using the following equation:

$$PC = \frac{\sum_o E_o(x)}{\varepsilon + \sum_s \sum_o A_{s,o}(x)} \tag{2.23}$$

where $E_o(x)$ is the local energy at orientation o, $A_{s,o}(x)$ is the local amplitude at scale s and orientation o , and ε is a positive stabilizing constant. $PC(x)$ is a real number that takes its values in the interval [0,1].

In order to compute the gradient magnitude of the reference and distorted images, three different gradient operators are employed. These operators are: Sobel operator [93], Prewitt operator [93], and Scharr operator [94].

Similarity measure computation

Consider PC_{ref} and PC_{dst} are PC maps computed for I_{ref} and I_{dst} respectively, and G_{ref} and G_{dst} are GM maps for these images. The final similarity measure between the reference and distorted images consists of two components: similarity measure between PC_{ref} and PC_{dst} or $S_{PC}(x)$, and similarity measure between G_{ref} and G_{dst} or $S_G(x)$.$S_{PC}(x)$ is calculated by the following equation:

$$S_{PC}(x) = \frac{2PC_{ref}(x)PC_{dst}(x) + T_4}{PC^2_{ref}(x) + PC^2_{dst}(x) + T_4} \tag{2.24}$$

where T_4 is a positive stabilizing constant chosen to prevent the denominator from becoming too small. $S_{PC}(x)$ takes its values in the interval $(0, 1]$.$S_G(x)$ is calculated by:

$$S_G(x) = \frac{2G_{ref}(x)G_{dst}(x) + T_5}{G^2_{ref}(x) + G^2_{dst}(x) + T_5} \tag{2.25}$$

where T_5 is a positive stabilizing constant. $S_G(x)$ takes its values in the interval (0,1].

The values of T_4 and T_5 depend on the dynamic range of PC and GM values respectively. The final similarity measure, $S_L(x)$, between I_{ref} and I_{dst} is computed as follows:

$$S_L(x) = [S_{PC}(x)]^{\alpha}[S_G(x)]^{\beta} \tag{2.26}$$

where α and β are two constants chosen to indicate the relative importance of each component (in [12], $\alpha = \beta$ =1). Our perception of an image is affected differently by different locations in an image, and a PC value at a location indicates whether that location is perceptually significant or not [89]. Therefore, if either of $PC_{ref}(x)$ and $PC_{dst}(x)$ is greater than the other, it implies that position x has a higher impact on HVS when evaluating $S_L(x)$ between I_{ref} and I_{dst}. As a result, $FSIM$ algorithm uses $PC_m(x) = max(PC_{ref}(x), PC_{dst}(x))$ as a weighting function for $S_L(x)$ in the overall similarity measure between I_{ref} and I_{dst}. Finally, the $FSIM$ index between the reference and distorted images is defined by:

$$FSIM = \frac{\sum_{x\epsilon\Omega} S_L(x)PC_m(x)}{\sum_{x\epsilon\Omega} PC_m(x)} \tag{2.27}$$

where Ω is the whole image spatial domain. The block diagram of the $FSIM$ index is presented in Figure 2.5.

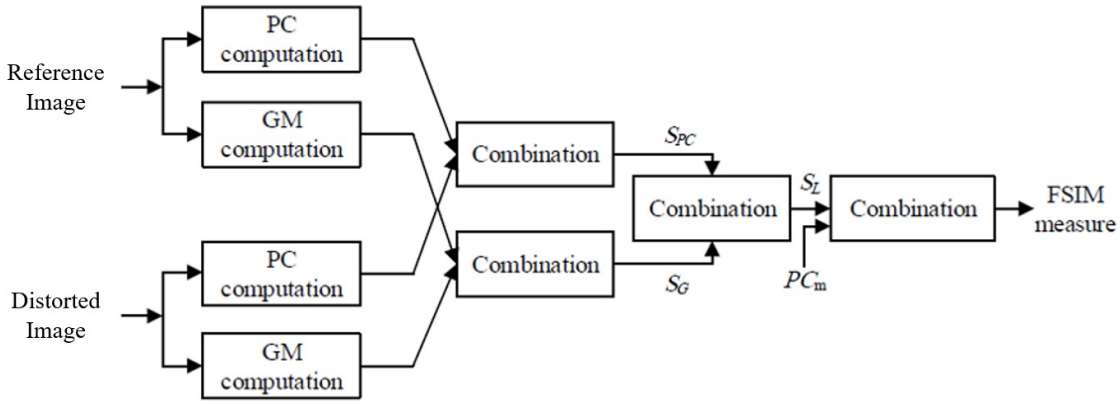


Figure 2.5: The block diagram of the FSIM algorithm. [92].

2.6 Feature similarity index for color images (FSIM$_C$)

$FSIM$ index described in Section 2.5 is designed for gray-scale images or the luminance component of color images. In order to extend $FSIM$ index to incorporate color images, first the reference RGB color image is transformed into another color space in which the luminance component can be separated from chrominance. In [12], RGB color image is transformed to YIQ color space, where Y denotes luminance component and I and Q denote chrominance components. RGB color space is transformed to YIQ color space via [95]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.274 & -0.322 \\ 0.211 & -0.523 & 0.312 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.28}$$

Suppose I_{ref} and Q_{ref} are chromatic components of the reference image, and I_{dst} and Q_{dst} are chromatic components of the distorted image. The similarity measures between chromatic components are computed as follows:

$$S_I(x) = \frac{2I_{ref}(x)I_{dst}(x) + T_6}{I^2_{ref}(x) + I^2_{dst}(x) + T_6} \tag{2.29}$$

$$S_Q(x) = \frac{2Q_{ref}(x)Q_{dst}(x) + T_7}{Q^2_{ref}(x) + Q^2_{dst}(x) + T_7} \tag{2.30}$$

where T_6 and T_7 are two positive stabilizing constant chosen to prevent the denominators from becoming too small. In [12], the values of T_6 and T_7 are set to be equal to each other. The final similarity measure between chromatic components,

$S_C(x)$, is the product of $S_I(x)$ and $S_Q(x)$:

$$S_C(x) = S_I(x)S_Q(x) \tag{2.31}$$

The $FSIM$ index for color images is calculated by:

$$FSIM_C = \frac{\sum_{x\epsilon\Omega} S_L(x)[S_C(x)]^\lambda PC_m(x)}{\sum_{x\epsilon\Omega} PC_m(x)} \tag{2.32}$$

where λ is a positive weighting constant chosen to indicate the relative importance of chromatic components. Note that for color images PC and GM are computed by their luminance component Y. Moreover, the calculation process of PC and GM for color images is the same as gray-scale images described in Section 2.5. The block diagram of the $FSIM_C$ algorithm is presented in Figure 2.6.

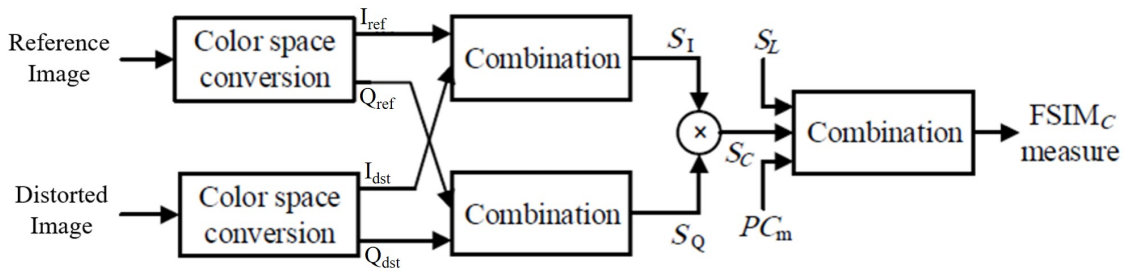


Figure 2.6: The block diagram of the $FSIM_C$ algorithm [92].

2.7 Combined Full-Reference Image Quality Metric (CQM)

In [21],K. Okarma, Chair of Signal Processing and Multimedia Engineering, Poland, proposed a new combined image quality metric which is based on three methods previously described by various researchers, MS-SSIM [41], VIF [9] and R-SVD [50]. The main advantage of the approach is the strong linear correlation with the subjective scores without additional nonlinear mapping. The values and the obtained correlation coefficients of the metric compared with some other state-of-art ones using two largest publicly available image databases, TID2008 and LIVE. The experimental results verified that the CQM has a great advantage of strong linear correlation with the subjective quality evaluation. The calculation formula of

CQM is defined as:

$$M_{CQM} = M_{MS-SSIM}^{f} \times M_{VIF}^{g} \times M_{R-SVD}^{h} \tag{2.33}$$

where $M_{MS-SSIM}$, M_{VIF} and M_{R-SVD} are objective scores of MS-SSIM [41], VIF [9] and R-SVD [50], respectively, and f, g and h are adjustable parameter values for each score. The values of parameters are set as f=7, g=0.3 and $h = -0.15$ for TID2008 database experimentally. According to the experimental results, the simple combination of the CQM led to the better performance compared to each single IQA method.

2.8 Extended Hybrid Image Similarity (EHIS)

In [22], K. Okarma proposed an Extended Hybrid Image Similarity (EHIS) to improve the prediction accuracy of the CQM in [21]. The EHIS approach based on the combination of four metrics: MS-SSIM, VIF, Riesz-based Feature Similarity (RFSIM) [100] and weighted FSIM (WFSIM) [101]. The calculation formula is defined as:

$$M_{EHIS} = M_{MS-SSIM}^{r} \times M_{VIF}^{s} \times M_{WFSIM}^{t} \times M_{RFSIM}^{u} \tag{2.34}$$

where $M_{MS-SSIM}$, M_{VIF}, M_{WFSIM} and M_{RFSIM} are objective scores of $MS-SSIM$, VIF, $WFSIM$ and $RFSIM$, respectively, and r, s, t and u are exponentiated parameter values. The values of the exponents obtained as the result of optimization conducted using TID2008 database are $r = -1.6131$, $s = 0.2037$, $t = 59.7151$ and $u = 0.1989$. Nevertheless, the combined metric, EHIS has a great advantage of strong linear correlation with the subjective quality evaluation. Moreover, it can be an interesting stimulus for further research related to the combined metric which could be more suitable for FR-IQA.

2.9 Exponentiated Combination of Two Scores for Image Quality Assessment (2SCM)

In [45], K. Isono and T. Shimamura proposed a new objective IQA method by combining two scores of objective IQA, in which one score is strong against the wide band noise and the other score is strong against the contrast change. The one score

used is LOGPSNR, which is effective for wide band noise images. In LOGPSNR, the Laplacian of Gaussian (LOG) filter is applied to the image, and then the image quality is measured by using PSNR. The other score used is VGS [44], which is effective for contrast change images. This two-scores combination method is called 2SCM in this chapter. The calculation formula of 2SCM is defined as follows:

$$M_{2SCM} = M^a_{LOGPSNR} + M^b_{VGS} \tag{2.35}$$

where $M_{LOGPSNR}$ and M_{VGS} are the objective IQA scores of LOGPSNR and VGS, respectively. a and b are exponentiated coefficients for adjusting each score. In their combination, they set the values of parameters as a=0.25 and b=1 for TID2008 database experimentally. The combination technique is able to assess distorted images correctly. However, it holds three problems, while it provides an excellent performance. First, LOGPSNR and VGS have common weak points. Both cannot assess Meanshift images in which the mean of the pixel values is changed. Second, the combination of the two exponentiated evaluation scores provides a low degree of freedom of modeling. It is desirable to obtain the final objective score utilizing a more complex function to consider linear relationship between the objective and subjective evaluations. Third, the parameters a and b in (2.33) should be decided manually according to the target images. Thus, they improved the 2SCM method as 3SCM method in [48] to provide better performance.

2.10 Optimized Three Scores Combination for Image Quality Assessment (3SCM)

In [48], K. Ishiyama, Y. Sugiura and T. Shimamura proposed a new objective IQA method adding another IQA score to the original combination technique in [45]. Adjustable parameters included in the new IQA measure are optimized with the genetic algorithm (GA). The added IQA is MSSIM and the three-scores combination method is called 3SCM in this chapter. The calculation formula of 3SCM is defined as follows:

$$M_{3SCM} = c_1 M^p_{LOGPSNR} + c_2 M^q_{VGS} + c_3 M_{MSSIM} \tag{2.36}$$

where $M_{LOGPSNR}$, M_{VGS} and M_{MSSIM} are objective scores of LOGPSNR, VGS and optimized MSSIM, respectively, and c_1, c_2, c_3, p and q are adjustable parame-

ters. In the three-scores combination, the values of parameters are set as p=5.28, q=10.44, c_1=56.60, c_2=16.00 and c_3=0.10 for TID2008 database experimentally. In the above equation, M_{MSSIM} is not exponentiated because $MS-SIM$ has been already exponentiated as shown in (2.9).

Experiments have demonstrated that the 3SCM method provided an excellent performance against Meanshift. However, it is a slight improvement relative to the two scores combination.

2.11 Particle Swarm Optimization (PSO) Algorithm

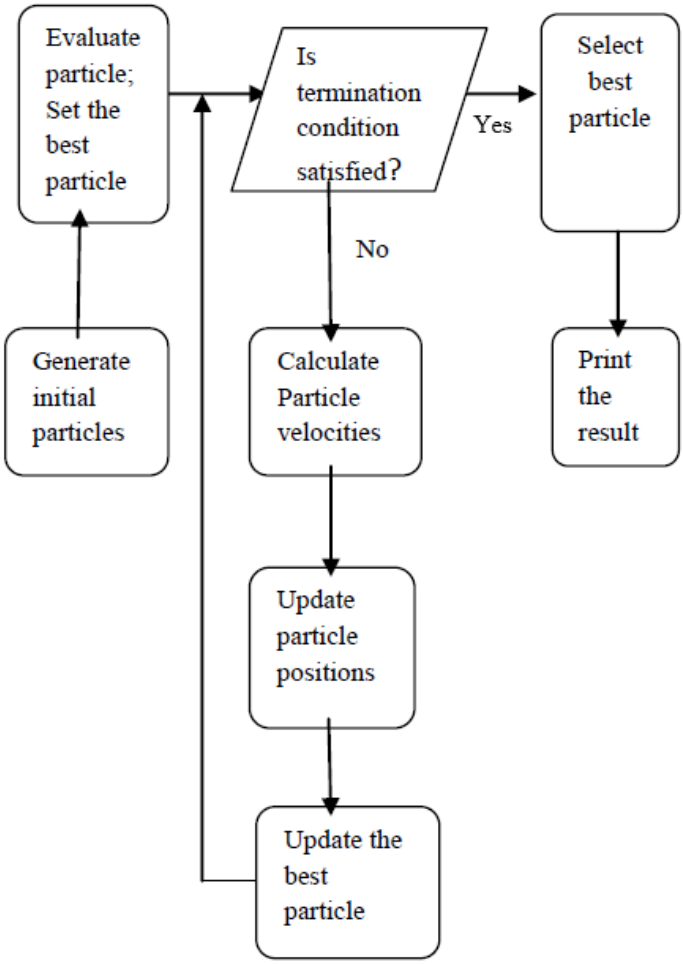


Figure 2.7: The basic structure of PSO algorithm.

PSO is a population-based search algorithm which is initialized with a population of random solutions, called particles [96]. As against the other evolutionary

computation techniques, each particle in this algorithm, called PSO is also associated with a velocity. Particles fly through the search space with velocities that are dynamically adjusted as per their historical behaviors. The particles, therefore, have the tendency to fly towards the better and better search area all over the course of the process of search. In PSO , a number of simple entities — the particles — are placed in the search space of some problem or function, and each one of these evaluates the objective function at its current location. Thereafter, each particle then determines its movement through the search space by combining some aspect of the history of its own current and best (best-fitness) locations with those of one or more members of the swarm, with some random perturbations. The next iteration takes place after all particles have moved. Eventually the swarm as a whole, like a flock of birds collectively foraging for food, is likely to move close to an optimum of the fitness function. The particle swarm is actually more than just a collection of particles. A particle by itself has almost does not solve any problem; progress takes place only when they i.e. the particles interact. Populations are organized according to some sort of communication structure or topology. This is often thought of as a social network. The topology typically consists of bidirectional edges connecting pairs of particles. It is like the alphabet j appearing in i's neighborhood, and likewise, i in j's neighbour. Each particle communicates with other particles and is affected by the best point found by any member of its topological neighborhood [97]. Each individual in the particle swarm is composed of three D dimensional vectors, where D is the dimensionality of the search space. These are the current position x_i, the previous best position p_i and the velocity v_i [97]. The i^{th} particle is represented as $X_i = (x_{i1}, x_{i2}, ..., x_{iD})$. At each generation, each particle is updated by the following two 'best' values. The first one is the best previous location (the position giving the best fitness value). This value is called $pBest$. The $pBest$ of the i^{th} particle is represented as $P_i = (p_{i1}, p_{i2}, ..., p_{iD})$. At each iteration, the P vector of the particle with the best fitness in the neighborhood, designated l or g, and the P vector of the current particle are combined to adjust the velocity along each dimension, and that velocity is then used to compute a new position for the particle. The portion of the adjustment to the velocity influenced by the individual's previous best position (P) is considered as the cognition component, and the portion influenced by the best in the neighborhood is the social component. With the addition of the inertia factor ω, [98] (brought in for balancing the global and the local search), velocity and position update equations

are:

$$v_i = \omega \times v_i + \eta_1 \times rand() \times (p_i - x_i) + \eta_2 \times rand() \times (p_g - x_i) \tag{2.37}$$

$$x_i = x_i + v_i \tag{2.38}$$

where $rand()$ and $rand()$ are two random numbers independently generated within the range [0,1] and η_1 and η_2 are two learning factors which control the influence of the social and cognitive components. In (2.31), if the sum on the right side exceeds a constant value, then the velocity on that dimension is assigned to be $\pm \mathrm{V}_{max}$. Thus, particles' velocities are clamped to the range $[-V_{max}, V_{max}]$ which serves as a constraint to control the global exploration ability of particle swarm. Thus, the likelihood of particles leaving the search space is reduced. The basic scheme of PSO algorithm is shown in Figure 2.6 [99].

2.12 Proposed Method

In this section, we derive new combination metrics by adding $MS-SSIM$ and $FSIM$ to MAD . In the metrics, we combine linearly and non-linearly the three scores which have dissimilar features by employing the PSO algorithm for parameter optimization. According to the experimental results, non-linear combination have better performance than linear combination. In this chapter, we describe both combination methods and their experimental results for clear understanding.

In the linear combination of the three objective scores, we employ the exponentiated coefficients and weighted constant values which are free parameter values for each database. The calculation formula for the three scores linear combination called $3LC$ in this chapter is defined as

$$M_{3LC} = k_1 M_{ma}^{\varphi} + k_2 M_{ms}^{\upsilon} + k_3 M_{fs}^{\rho} \tag{2.39}$$

where M_{ma}, M_{ms} and M_{fs} are objective scores of MAD , $MS-SSIM$ and $FSIM$, respectively. k_1, k_2 and k_3 are wighting factors, φ, υ and ρ are exponentiated coefficients. These parameters are optimized by employing the PSO algorithm.

In the non-linear combination of the three objective scores, we employ the exponentiated coefficients and weighted constant values which are adjustable parameter values for each database. The calculation formula for the three scores non-linear

combination called $3NC$ in this chapter is defined as

$$M_{3NC} = k_1 M_{ma}^{\varphi} + k_2 M_{ms}^{\upsilon} + M_{fs}^{\rho} + C \tag{2.40}$$

where M_{ma}, M_{ms} and M_{fs} are objective scores of MAD , $MS-SSIM$ and $FSIM$, respectively. k_1 and k_2 are wighting factors, φ, υ and ρ are exponentiated coefficients and C is a constant value. These parameters are optimized by employing the PSO algorithm.

Particle Swarm Optimization (PSO) Algorithm

PSO is an artificial intelligence (AI) technique that can be used to find approximate solutions to extremely difficult or impossible numeric maximization and minimization problems. In recent years, PSO has become a better-developed optimization algorithm. It searches the optimal solution through continuous iteration, and it finally employs the size of the value of objective function, or the function to be optimized (also known as the fitness function in the particle swarm), in order to evaluate the quality of the solution. Indeed, stochastic search techniques contain randomness process so their performances can change from problem to problem. Many parameters can affect the algorithms' performances such as problem size, number of constraint functions and constraint function type. No free lunch theorems in [63–66] logically proves that no one can propose an algorithm for solving all optimization problems. This means that the success of an algorithm in solving a specific set of problems does not guarantee solving all optimization problems with different type and nature. Thus, although there are many state-of-the-art Swarm Intelligence (SI) based approaches for parameter optimization, we have to do an experiment first to be able to decide the most suitable optimization algorithm for our specific problem. After extensive experiments, we select the PSO algorithm for parameter optimization in our non-linear combination approach because the PSO could provide faster convergence and could find better solutions. Furthermore, the main advantage of the PSO algorithm is its easiness in implementation since there are no crossover, decoding and encoding. In addition, PSO is simple both in theory and numerical implementation and the computational time is inexpensive as compared to other optimal algorithms. The steps of the PSO to be implemented are the followings:

1) Create the initial particles, and assign them initial velocities random-uniformly. As using 10 particles for each parameter is large enough to get good results for

most of the cases, the population size used for optimization in our method is 60 (6 parameters x 10 particles).

2) Evaluate the objective function at each particle location, and determines the best (lowest) function value and the best location.

3) Update the particle locations (the new location is the old one plus the velocity), velocities, and neighbors.

4) Select particles randomly based on the fitness of the particles. In our proposed method, the particles which have the higher fitness values have the higher probability to be selected.

5) Check whether the swarm of particles is converged or not.

In our work, the maximum number of iterations is set to 1000. If the maximum iteration number or minimum error criterion is attained, the PSO is terminated. Otherwise, the velocity and position are updated to create a new swarm.

2.13 Experimental Results and Discussion

In this section, we compare the performance of the proposed methods, $3LC$ and $3NC$ with MAD [8], $MS-SSIM$ [2] and $FSIM$ [9] based on six image quality databases (A57, CSIQ, TID2008, TID2013, LIVE, and IVC). For detailed performance evaluation, we compared the performance of the whole database and the performance of each distortion type based on each database.

2.13.1 Image Databases

Figure 2.8: Original Images from A57

Figure 2.9: Some Distorted Images from A57

The A57 Database [54] has 3 reference images as shwon in Figure 2.8, and 54 distorted images, including six distortion types - Flat allocation (FLT); equal distortion contrast @ all scales, JPEG compression (JPEG), JPEG2000 compression (JPEG2), JPEG2000 compression with the Dynamic Contrast-Based Quantization (DCQ), Gaussian blurring (GB), and additive Gaussian white noise (WGN) - at 3 different levels. Some of the distorted images of A57 database are shown in Figure 2.9. The subjective quality scores used for this database are DOMS, ranging from 0 to 1.

The Categorical Image Quality (CSIQ) Database [55] contains 30 reference images as shown in Figure 2.10, and 866 distorted images. Each image is distorted using 6 types of distortions - JPEG compression (JPEG), JPEG2000 compression (JPEG2), global contrast decrements (Contrast), additive Gaussian white noise (WGN), additive Gaussian pink noise (PGN), and Gaussian blurring (BLUR) - at 4 to 5 different levels. Some of distorted images of CSIQ database are shown in Figure 2.11. The score ratings (0 to 1) are reported in the form of DMOS.

The Tampere Image Database (TID2008) [56] includes 25 reference images as shown in Figure 2.12 and 1700 distorted images. Each reference image is distorted using 17 types of distortions as shown in Table 2.1, with 4 different levels. The subjective quality scores provided for this database are MOS , ranging from 0 to

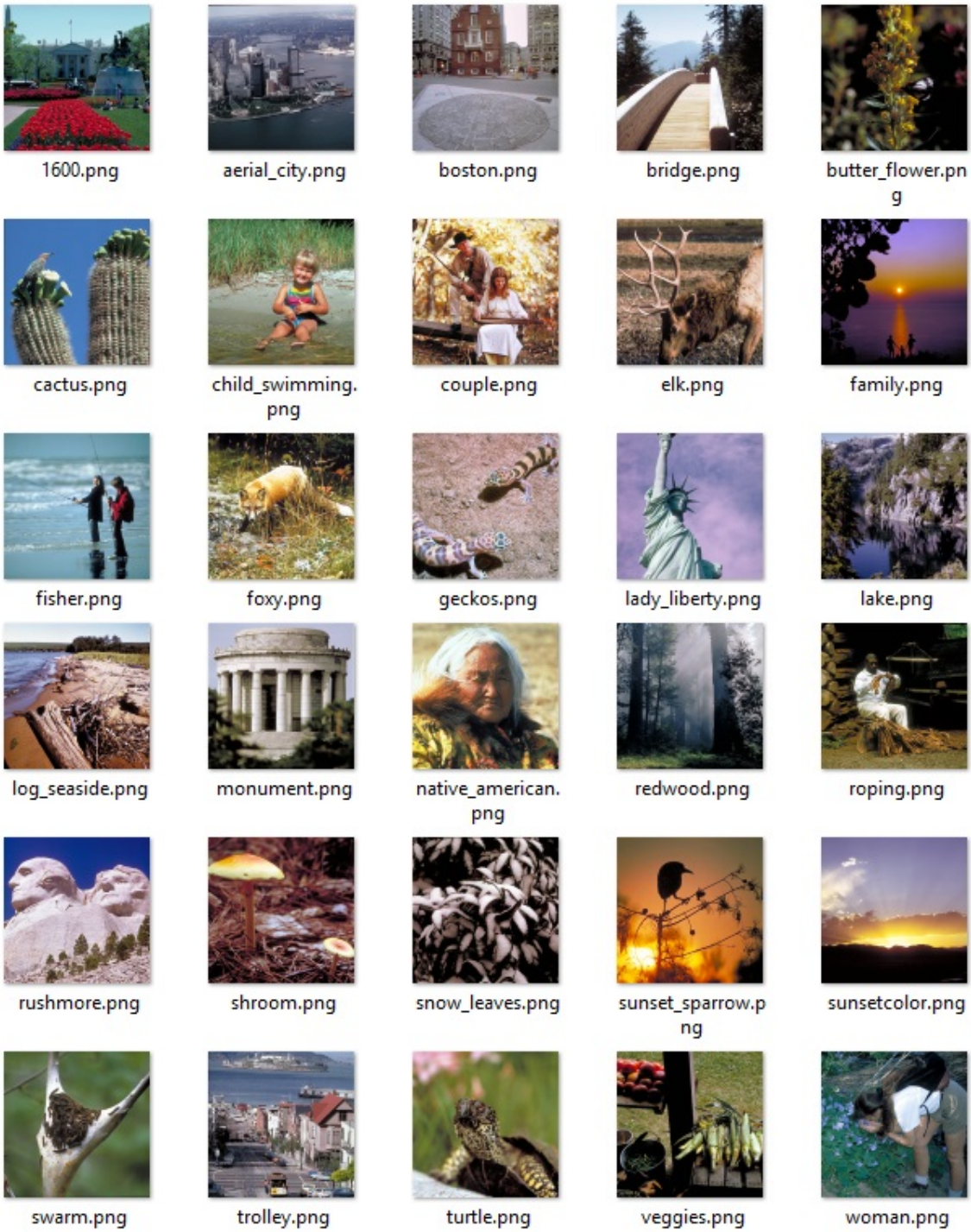


Figure 2.10: Original Images from CSIQ

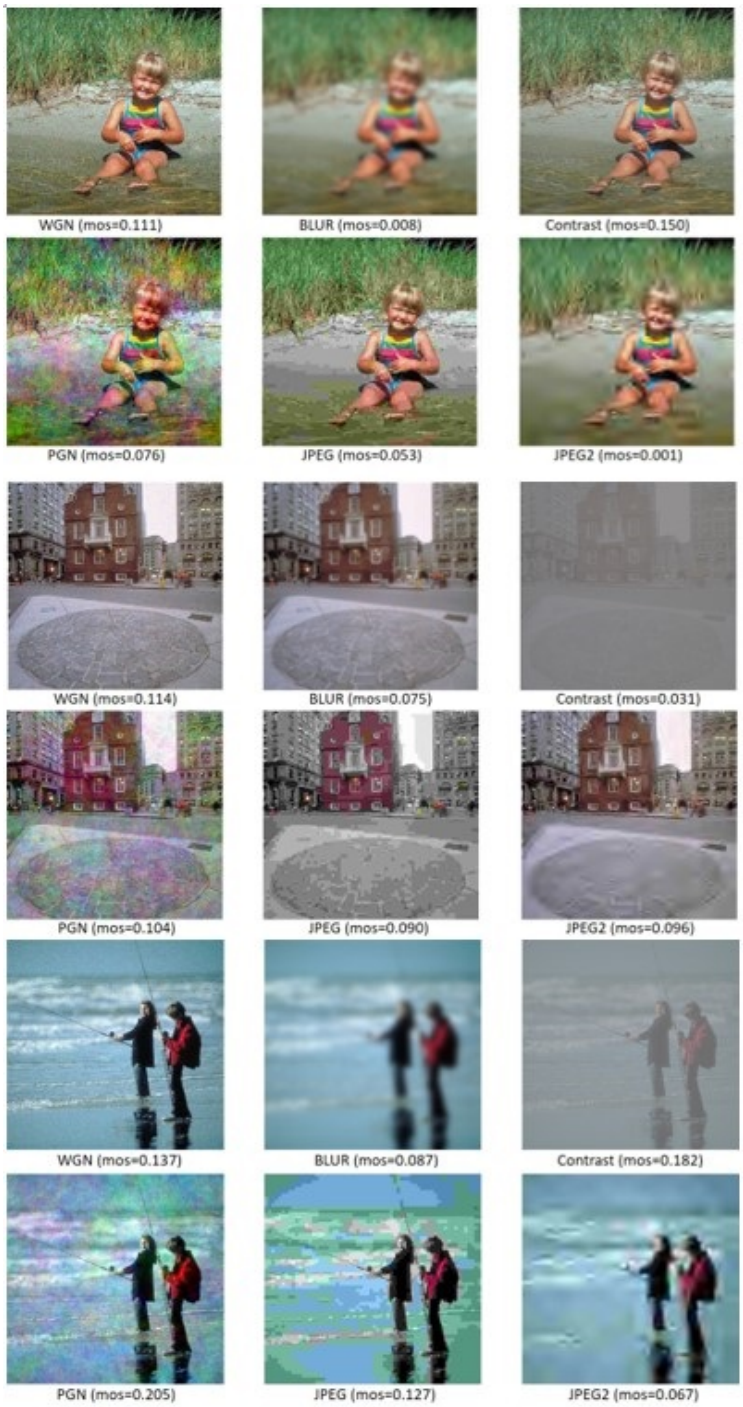


Figure 2.11: Some of Distorted Images from CSIQ

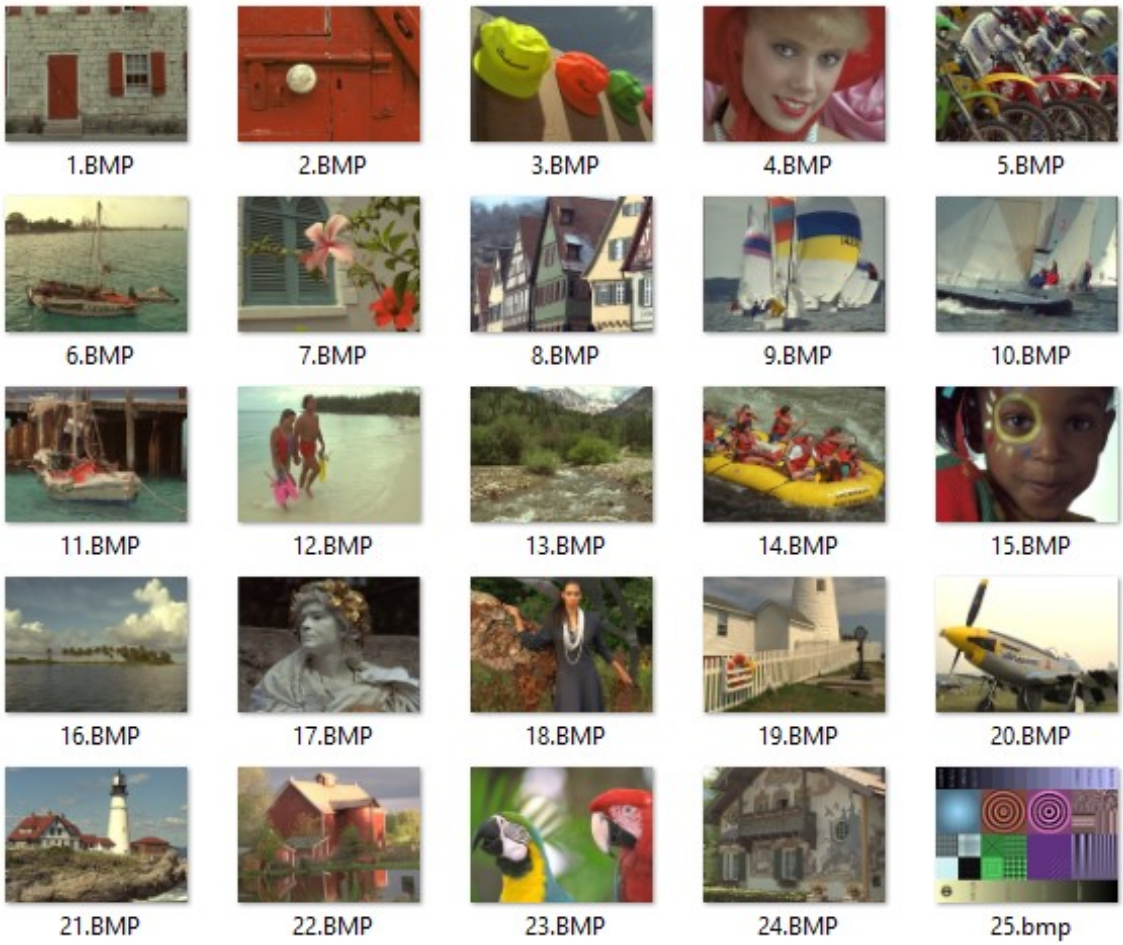


Figure 2.12: Original Images from TID2008

9. Examples of distorted images for TID2008 database are shown in Figure 2.13.

Figure 2.13: Some Distorted Images (17 types) from TID2008

Table 2.1: Types of Distortions in TID2008

No	**Types**
1	Additive Gaussian noise
2	Additive noise in color component
3	Spatially correlated noise
4	Masked noise
5	High frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change

The Tampere Image Database (TID2013) [57] is the extended version of TID2008 described in above and it includes 25 reference images, 24 types of distortions for each reference image as shown in Table 2.2, and 5 different levels for each type of distortion. The whole database contains 3000 distorted images. Examples of distorted images for TID2013 database is shown in Figure 2.14.

Table 2.2: Types of Distortions in TID2013

No	**Types**
1	Additive Gaussian noise
2	Additive noise in color component
3	Spatially correlated noise
4	Masked noise
5	High frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change
18	Change of color saturation
19	Multiplicative Gaussian noise
20	Comfort noise
21	Lossy compression of noisy images
22	Image color quantization with dither
23	Chromatic aberrations
24	Sparse sampling and reconstruction

Figure 2.14: Some Distorted Images from TID2013

The LIVE Image Quality Database [58] has 29 reference images as shown in Figure 2.15 and 779 distorted images, including five distortion types - JPEG2000 (JPEG2), JPEG, white noise in the RGB components (WN), Gaussian blur (GB), and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model (Fast-fading). Some distorted images of LIVE are shown in Figure 2.16.

Figure 2.15: Original Images from LIVE

Figure 2.16: Some Distorted Images from LIVE

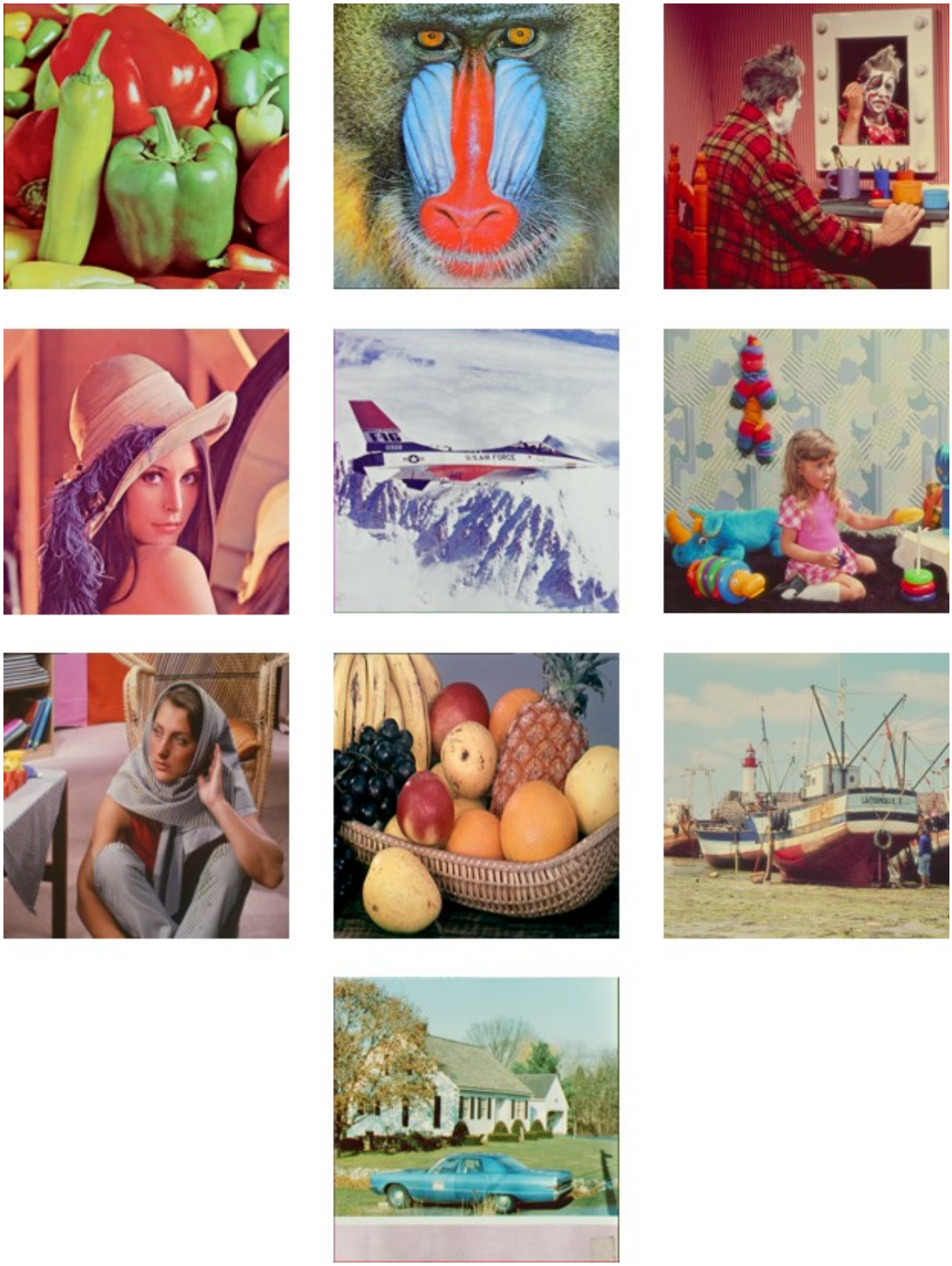


Figure 2.17: Original Images from IVC

The IVC Database [59] has 10 reference images as shown in Figure 2.17, and 185 distorted images, including 4 types of distortions blurring (BLUR), JPEG2000

(JPEG2), JPEG, and locally adaptive resolution (LAR) coding. Some distorted images of IVC can be seen in Figure 2.18.

Figure 2.18: Some Distorted Images from IVC

2.13.2 Cross Validation

We use all images from one distortion group in one database for the training, and test the images from the same distortion group in the other remaining databases. For example, we build a trained dataset (for JPEG compression) from LIVE

database. Then, we use this trained dataset to do the testing on A57, CSIQ, TID2008, TID2013, and IVC.

2.13.3 Parameter Optimization

We optimized the adjustable parameters in (2.39) and (2.40) utilizing the PSO algorithm we explained in Section 2.12. For the optimization, we used references images in each database. For example, we used 25 reference images in TID2013 (total 3025 images).

First, we utilized the original exponentiated values of $MS-SSIM$ in (2.9) as 1 described in Section 2.3.

Second, we optimized (2.39) and (2.40) evaluating CC between the subjective scores of each database and the corresponding objective scores. Here, the objective function of PSO is finding the maximum value of PCC between MOS and related objective scores for each database. The resulting parameter values of linear combination $3LC$ and non-linear combination $3NC$ for each database are summarized in Tables 2.3 and 2.4.

Table 2.3: Optimized Parameter Values for each database by PSO algorithm for $3LC$

Parameter	**A57**	**CSIQ**	**TID2008**	**TID2013**	**LIVE**	**IVC**
k1	1.122	63.000	0.010	0.278	83.675	0.001
k2	0.854	0.001	0.040	0.029	54.167	0.001
k3	1.172	0.001	1.055	0.997	40.174	0.001
φ	0.001	0.972	0.001	0.001	0.094	0.001
υ	0.001	0.001	9.190	6.342	13.502	0.001
ρ	0.001	0.001	8.580	9.269	100	19.707

Table 2.4: Optimized Parameter Values for each database by PSO algorithm for $3NC$

Parameter	**A57**	**CSIQ**	**TID2008**	**TID2013**	**LIVE**	**IVC**
k1	1.122	63.000	0.010	0.278	83.675	0.001
k2	0.854	0.001	0.040	0.029	54.167	0.001
φ	0.001	0.972	0.001	0.001	0.094	0.001
υ	0.001	0.001	9.190	6.342	13.502	0.001
ρ	0.001	0.001	8.580	9.269	100	19.707
C	61.031	0.071	7.980	7.056	22.199	63.880

2.13.4 Performance Evaluation

The performances of the methods between MOS and objective scores are calculated based on the Pearson's Correlation Coefficient (PCC) and Spearman Rank-Order Correlation Coefficient (SROCC). PCC and SROCC are defined as

$$\text{PCC} = \frac{\sum a_j b_j - J\bar{a}\bar{b}}{(J-1)s_a s_b} \tag{2.41}$$

where J is the data size, a_j and b_j are the single subjective and objective scores with index j, respectively. (s_a), s_b denote the standard deviations of subjective and objective scores, respectively and $\bar{a}$, $\bar{b}$ denote the mean values of subjective and objective scores, respectively.

$$\text{SROCC} = 1 - \frac{6\sum d^2}{J^3 - J} \tag{2.42}$$

where d is the difference between the subjective and objective score ranks.

2.13.5 Experimental Results

In Tables 2.5-2.10, we compare the PCC and SROCC values of twenty-one IQA methods based on six databases. The experimental results show that the proposed method $3NC$ is more effective and well correlated with HVS in comparison to other state−of−the−art not only single methods, Global Local Distortion with PFT (GLD-PFT) [60], Locally Adaptive Fusion (LAF) [61], Difference of Gaussian (DOG)-SSIM ($DOGSSIM$) [62] and Edge Similarity ($ESIM$) [102] but also combination methods, Two Scores Combination Method (2SCM) [45], Three Scores Combination Method (3SCM) [48], Composed Quality Measure (CQM) [21], Extended Hybrid Image Similarity (EHIS) [22] and even linear combination $3LC$.

For detailed performance, we compare the PCC and SROCC values of our proposed methods, $3LC$ and $3NC$ with the values of MAD , $MS-SSIM$ and $FSIM$ for each distortion type of each database. The results are shown in Tables 2.11-2.22 in Appendix of this chapter. According to these experimental results, although our proposed method $3NC$ has lower values for some distortion types, it is still powerful in almost all distortion types of every database.

Table 2.5: PCC and SROCC Performance Comparison for A57 Database

IQA Index	**PCC**	**SROCC**
SSIM	0.8019	0.8067
MS-SSIM	0.5596	0.6673
VSNR	0.9502	0.9359
PSNR	0.6347	0.6189
PSNR-HVS	0.8832	0.8502
IFC	0.4549	0.3187
NQM	0.8027	0.7978
IW-SSIM	0.9024	0.8713
MAD	0.5009	0.8499
VIF	0.6160	0.6223
FSIM	0.5566	0.4962
2SCM [45]	-	-
3SCM [48]	-	-
CQM [21]	-	-
EHIS [22]	-	-
GLD-PFT [60]	-	-
LAF [61]	-	-
DOG-SSIM [62]	-	-
ESIM [102]	-	-
3LC (proposed)	0.9120	0.9350
3NC (proposed)	**0.9585**	**0.9423**

Table 2.6: PCC and SROCC Performance Comparison for CSIQ Database

IQA Index	**PCC**	**SROCC**
SSIM	0.8594	0.8755
MS-SSIM	0.7837	0.8734
VSNR	0.8005	0.8108
PSNR	0.8001	0.8057
PSNR-HVS	0.8231	0.8294
IFC	0.8358	0.7671
NQM	0.7422	0.7411
IW-SSIM	0.9025	0.9212
MAD	0.9348	0.9281
VIF	0.9253	0.9194
FSIM	0.7999	0.9192
2SCM [45]	-	-
3SCM [48]	-	-
CQM [21]	-	-
EHIS [22]	-	0.9498
GLD-PFT [60]	-	0.9549
LAF [61]	-	0.9630
DOG-SSIM [62]	-	0.9204
ESIM [102]	-	0.9620
3LC (proposed)	0.9009	0.9506
3NC (proposed)	**0.9359**	**0.9686**

Table 2.7: PCC and SROCC Performance Comparison for TID2008 Database

IQA Index	**PCC**	**SROCC**
SSIM	0.7715	0.7749
MS-SSIM	0.8389	0.8528
VSNR	0.6820	0.7046
PSNR	0.5355	0.5245
PSNR-HVS	0.5977	0.5943
IFC	0.7186	0.5707
NQM	0.6103	0.6243
IW-SSIM	0.8488	0.8559
MAD	0.8306	0.8340
VIF	0.8055	0.7496
FSIM	0.8710	0.8805
2SCM [45]	0.8670	0.8920
3SCM [48]	0.8923	0.8891
CQM [21]	-	0.8720
EHIS [22]	-	0.9098
GLD-PFT [60]	-	0.8849
LAF [61]	-	0.8100
DOG-SSIM [62]	-	0.9259
ESIM [102]	-	0.9026
3LC (proposed)	0.8956	0.9000
3NC (proposed)	**0.9007**	**0.9298**

Table 2.8: PCC and SROCC Performance Comparison for TID2013 Database

IQA Index	**PCC**	**SROCC**
SSIM	0.6504	0.6274
MS-SSIM	0.7290	0.7859
VSNR	-	-
PSNR	0.6390	0.6394
PSNR-HVS	-	-
IFC	-	-
NQM	-	-
IW-SSIM	-	-
MAD	0.8070	0.8380
VIF	0.7720	0.6769
FSIM	0.8210	0.8022
2SCM [45]	0.7997	0.7957
3SCM [48]	0.8050	0.7994
CQM [21]	-	-
EHIS [22]	-	-
GLD-PFT [60]	-	-
LAF [61]	-	-
DOG-SSIM [62]	-	0.8942
ESIM [102]	-	0.8804
3LC (proposed)	0.8744	0.8547
3NC (proposed)	**0.8779**	**0.8944**

Table 2.9: PCC and SROCC Performance Comparison for LIVE Database

IQA Index	**PCC**	**SROCC**
SSIM	0.9384	0.9479
MS-SSIM	0.9402	0.9521
VSNR	0.9235	0.9279
PSNR	0.8701	0.8756
PSNR-HVS	0.9134	0.9186
IFC	0.9261	0.9259
NQM	0.9128	0.9093
IW-SSIM	0.9425	0.9567
MAD	0.9672	0.9669
VIF	0.9597	0.9636
FSIM	0.9540	0.9634
2SCM [45]	-	-
3SCM [48]	-	-
CQM [21]	-	-
EHIS [22]	-	0.9622
GLD-PFT [60]	-	0.9631
LAF [61]	-	0.9570
DOG-SSIM [62]	-	0.9423
ESIM [102]	-	0.9420
3LC (proposed)	0.9601	0.9600
3NC (proposed)	**0.9772**	**0.9806**

Table 2.10: PCC and SROCC Performance Comparison for IVC Database

IQA Index	**PCC**	**SROCC**
SSIM	0.9117	0.9018
MS-SSIM	0.8435	0.8997
VSNR	0.8027	0.7993
PSNR	0.7192	0.6885
PSNR-HVS	0.8648	0.8590
IFC	0.9093	0.8993
NQM	0.8489	0.8343
IW-SSIM	0.9228	0.9125
MAD	0.8986	0.9082
VIF	0.9026	0.8964
FSIM	0.9376	0.9262
2SCM [45]	-	-
3SCM [48]	-	-
CQM [21]	-	-
EHIS [22]	-	-
GLD-PFT [60]	-	-
LAF [61]	-	-
DOG-SSIM [62]	-	-
ESIM [102]	-	-
3LC (proposed)	0.9499	0.9364
3NC (proposed)	**0.9534**	**0.9368**

Therefore, according to the experimental results, our proposed method $3NC$ outperforms consistently well almost all types of distortions in six databases and is more robust than other previous combination methods including linear combination $3LC$.

2.14 Conclusion

As far as we know, there is no perfect single method which can give the best performance for all distortion types on every database. Similarly, there is no perfect single combination method which is very robust in the prediction of the quality of images of all databases. In this paper, therefore, we have proposed new linear and nonlinear combination methods which combines MAD , $MS-SSIM$ and $FSIM$ scores using particle swarm optimization algorithm. Our experimental results have demonstrated that non-linear combination method $3NC$ is a superior one in comparison to other not only conventional single methods but also previous combination methods.

Table 2.11: PCC performance evaluations for each distortion type based on A57 Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Flat allocation	0.5460	0.6187	0.6226	0.7132	**0.7945**
2	JPEG compression	0.4732	0.5201	0.5591	0.7701	**0.8667**
3	JPEG-2000 compression	0.4733	0.4898	0.4838	0.7217	**0.9333**
4	JPEG-2000+DCQ compression	0.5238	0.4556	0.5445	0.8603	**0.9333**
5	Gaussian blur	0.5672	0.4848	0.5759	0.8032	**0.8541**
6	Gaussian white noise	0.8011	0.6825	**0.8425**	0.7096	0.7119

Table 2.12: SROCC performance evaluations for each distortion type based on A57 Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Flat allocation	0.6557	0.6724	0.7057	**0.7124**	0.4391
2	JPEG compression	0.5391	0.3724	0.6724	0.7282	**0.7667**
3	JPEG-2000 compression	0.5557	0.4224	0.6224	0.8129	**0.8833**
4	JPEG-2000+DCQ compression	0.4224	0.5391	0.3724	0.8333	**0.9633**
5	Gaussian blur	0.4224	0.4891	0.3891	0.8663	**0.8767**
6	Gaussian white noise	0.3891	**0.5891**	0.5224	0.4792	0.4891

Table 2.13: PCC performance evaluations for each distortion type based on CSIQ Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Additive Gaussian White noise	0.9486	0.9018	0.7643	**0.9496**	**0.9496**
2	Gaussian blurring	0.9713	0.8367	0.8822	**0.9723**	**0.9723**
3	Global contrast decrements	0.7632	0.7470	0.6514	**0.7667**	**0.7667**
4	Additive Gaussian Pink noise	0.8663	0.7269	0.7254	**0.8665**	**0.8665**
5	JPEG compression	**0.9595**	0.8991	0.8493	0.9585	0.9585
6	JPEG-2000 compression	0.9808	0.8663	0.9073	**0.9813**	**0.9813**

Table 2.14: SROCC performance evaluations for each distortion type based on CSIQ Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Additive Gaussian White noise	0.9540	0.9317	0.9258	0.9533	**0.9551**
2	Gaussian blurring	0.9680	0.9174	**0.9721**	0.9679	0.9690
3	Global contrast decrements	0.8044	**0.8149**	0.8057	0.7950	0.7950
4	Additive Gaussian Pink noise	**0.8191**	0.7636	0.7921	**0.8191**	**0.8191**
5	JPEG compression	0.9581	0.9459	0.9544	0.9580	**0.9590**
6	JPEG-2000 compression	0.9752	0.9129	0.9684	0.9750	**0.9754**

Table 2.15: PCC performance evaluations for each distortion type based on LIVE Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Bit errors in JPEG2000 bit stream	0.7359	0.8366	0.8947	0.9243	**0.9859**
2	Gaussian blur	0.9545	0.8739	**0.9810**	0.6192	0.6198
3	JPEG2000	0.4309	0.5557	0.6997	0.8991	**0.9102**
4	JPEG compresses	0.4191	0.5269	0.6586	0.7086	**0.7822**
5	White noise	0.9418	**0.9663**	0.8494	0.7011	0.7177

Table 2.16: SROCC performance evaluations for each distortion type based on LIVE Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Bit errors in JPEG2000 bit stream	0.8628	0.7964	0.8463	0.8740	**0.9531**
2	Gaussian blur	0.9621	0.8521	**0.9717**	0.8274	0.8396
3	JPEG2000	0.7295	0.8567	0.7103	0.8662	**0.8908**
4	JPEG compresses	0.6046	0.7665	0.6440	0.8533	**0.8661**
5	White noise	0.9501	**0.9524**	0.9448	0.8021	0.8033

Table 2.17: PCC performance evaluations for each distortion type based on IVC Database

No.	Distortions	MAD	MS-SSIM	FSIM	3LC	3NC
1	Blurring	0.9714	0.8815	0.8904	0.9863	**0.9864**
2	JPEG2000 compression	0.9160	0.9145	0.9296	0.9650	**0.9657**
3	JPEG compression	0.9401	0.9270	0.9601	0.9923	**0.9923**
4	Locally adaptive resolution (LAR)	**0.9558**	0.9414	0.9352	0.9330	0.9384

Table 2.18: SROCC performance evaluations for each distortion type based on IVC Database

No.	**Distortions**	**MAD**	**MS-SSIM**	**FSIM**	**3LC**	**3NC**
1	Blurring	0.9630	0.8526	0.9625	0.9639	**0.9640**
2	JPEG2000 compression	0.9098	0.9383	0.9620	0.9620	**0.9624**
3	JPEG compression	0.8962	0.9143	**0.9805**	**0.9805**	**0.9805**
4	Locally adaptive resolution (LAR)	**0.9489**	0.9474	0.8857	0.8857	0.8857

Table 2.19: PCC performance evaluation for each distortion type on TID2008 Database

Distor-tions	**MAD**	**MS-SSIM**	**FSIM**	**3LC**	**3NC**
1	0.8165	0.7743	0.7828	0.8193	**0.8582**
2	0.8267	0.7982	0.8198	0.8359	**0.8654**
3	0.8598	0.7908	0.7938	0.8504	**0.8783**
4	0.7566	0.8086	0.7689	0.8051	**0.8104**
5	0.8931	0.8522	0.8383	0.8917	**0.9369**
6	0.0417	0.7231	0.6740	0.7216	**0.7558**
7	0.7981	0.7641	0.7831	0.8395	**0.8803**
8	0.9227	0.9075	0.9079	0.9427	**0.9597**
9	0.9612	0.8974	0.9320	0.9674	**0.9878**
10	0.9487	0.8999	0.9253	0.9685	**0.9959**
11	0.9733	0.8498	0.9554	**0.9738**	0.9732
12	0.8556	0.8140	0.8418	0.8803	**0.9002**
13	0.8295	0.8089	0.7877	0.8494	**0.8566**
14	**0.8242**	0.6661	0.7264	0.7345	0.7659
15	0.8007	**0.8902**	0.8410	0.8528	0.8757
16	0.5709	0.6832	0.6704	0.6980	**0.7176**
17	0.2573	0.5890	0.7285	0.7131	**0.7294**

Table 2.20: SROCC performance evaluation for each distortion type on TID2008 Database

Distortions	**MAD**	**MS-SSIM**	**FSIM**	**3LC**	**3NC**
1	0.8389	0.8027	0.8577	0.8578	**0.8777**
2	0.8257	0.8129	0.8517	0.8513	**0.8805**
3	0.8671	0.8280	0.8479	0.8478	**0.8678**
4	0.7339	0.8140	0.8022	0.8117	**0.8241**
5	0.8865	0.8451	0.9095	0.9074	**0.9295**
6	0.0644	0.7543	0.7455	0.7446	**0.7655**
7	0.8158	0.7963	0.8550	0.8550	**0.8750**
8	0.9195	0.9396	0.9470	0.9494	**0.9671**
9	0.9433	0.9229	0.9604	0.9611	**0.9802**
10	0.9276	0.8973	0.9283	0.9284	**0.9485**
11	0.9708	0.8830	0.9776	0.9779	**0.9976**
12	0.8657	0.8203	0.8705	0.8715	**0.8905**
13	0.8400	0.8393	0.8551	0.8554	**0.8751**
14	**0.8287**	0.6921	0.7497	0.7451	0.7695
15	0.7959	**0.8963**	0.8478	0.8488	0.8678
16	0.5170	**0.7192**	0.6700	0.7043	0.6940
17	0.2722	0.5535	0.6482	0.6471	**0.6687**

Table 2.21: PCC performance evaluation for each distortion type on TID2013 Database

Distortions	**MAD**	**MS-SSIM**	**FSIM**	**3LC**	**3NC**
1	0.8704	0.8492	0.8282	0.8722	**0.8792**
2	**0.8232**	0.8016	0.8172	0.6764	0.6564
3	0.8813	0.8257	0.8088	0.8830	**0.8900**
4	0.7974	0.8150	**0.8163**	0.4685	0.4242
5	0.9067	0.8878	0.8468	0.9024	**0.9107**
6	0.2676	0.7602	0.6881	0.7618	**0.7707**
7	0.8552	0.8791	0.8377	0.8954	**0.9011**
8	0.9397	0.9259	0.8919	0.9471	**0.9491**
9	0.9595	0.9364	0.9388	0.9744	**0.9717**
10	0.9518	0.9420	0.9278	0.9706	**0.9717**
11	0.9674	0.8860	0.9336	**0.9699**	0.9614
12	0.8661	0.8512	0.8719	0.9015	**0.9034**
13	0.8676	0.8286	0.8029	**0.8906**	0.8901
14	**0.8545**	0.7839	0.8221	0.8265	0.8264
15	0.3042	**0.6075**	0.5114	0.5212	0.5200
16	0.6882	**0.7317**	0.6986	0.2762	0.2519
17	0.2612	0.5433	**0.6796**	0.6421	0.6288
18	0.0581	**0.476**	0.3814	0.0544	0.0549
19	**0.8436**	0.7866	0.7900	0.8304	0.8373
20	0.9279	0.8564	0.924	0.9450	**0.9452**
21	0.9513	0.9231	0.9392	0.9569	**0.9570**
22	0.8691	0.7817	0.8149	0.8934	**0.9015**
23	0.9549	0.9444	0.9409	0.9624	**0.9646**
24	0.9653	0.8870	0.9441	**0.9728**	0.9608

Table 2.22: SROCC performance evaluation for each distortion type on TID2013 Database

Distor-tions	**MAD**	**MS-SSIM**	**FSIM**	**3LC**	**3NC**
1	0.8812	0.8708	0.9002	0.8996	**0.9402**
2	0.7896	0.7836	**0.8196**	0.7241	0.7573
3	0.8931	0.8774	0.8813	0.8815	**0.9213**
4	0.7521	0.8041	**0.8168**	0.6201	0.6601
5	0.8805	0.8557	0.8959	0.8960	**0.9387**
6	0.2664	0.8081	0.8018	0.8023	**0.8418**
7	0.8726	0.9134	0.9060	0.9070	**0.9460**
8	0.9408	0.9673	0.9590	0.9607	**0.9792**
9	0.9242	0.9442	0.9497	0.9508	**0.9897**
10	0.9215	0.9330	0.9330	0.9346	**0.9739**
11	0.9543	0.9043	0.9579	0.9590	**0.9785**
12	0.8393	0.8215	0.8453	0.8522	**0.8914**
13	0.8794	0.8724	0.8886	0.8897	**0.9286**
14	0.8322	0.7972	0.8139	0.8140	**0.8539**
15	0.2590	**0.6281**	0.5355	0.5379	0.5755
16	0.6591	**0.7813**	0.7606	0.4198	0.4494
17	0.1997	0.3859	0.4648	0.4638	**0.5048**
18	0.0091	**0.4436**	0.3950	0.3625	0.3940
19	0.8430	0.8021	0.8543	0.8524	**0.8946**
20	0.9024	0.8596	0.9165	0.9143	**0.9526**
21	0.9439	0.9295	0.9564	0.9561	**0.9664**
22	0.8792	0.7993	0.8947	0.8928	**0.9344**
23	0.8375	0.8881	0.8779	0.8867	**0.9241**
24	0.9564	0.9072	**0.9646**	**0.9646**	**0.9646**

Chapter 3

Convolutional Neural Network for Blind Image Quality Assessment

Blind image quality assessment methods ($BIQA$) are able to measure the quality of distorted images even without referencing the original images. This property is peerless in image processing field because reference images are normally not available in practice. Unlike the existing trained models, in our work, the training process is constructed as an end-to-end learning mechanism that minimizes the loss between the predicted score and the ground truth score of Human Vision System (HVS). Moreover, the convolutional neural network (CNN) takes distorted images as inputs and outputs the related score for each image. In this section, we evaluate the proposed method on the publicly available six benchmarks and cross-database validation performance on LIVE, CSIQ and TID2013 databases. The experimental results show that our proposed method outperforms other state-of-the-art ones.

3.1 Related Work

Nowadays, huge amount of images are daily produced for several purposes, for example, forecasting weather, finding diseases and monitoring criminals. For these reasons, it is very importance to keep the quality of such images at an acceptable visual level at the end-users after the production and transmission. Furthermore, accurate measurement of the image quality is an important step in many image-based applications. To achieve this goal, effective image quality assessment (IQA) algorithms are necessary and have recently become a very hot research topic.

As human observers are the superlative users in most of the multimedia applications, the most accurate and reliable way of assessing the quality of images is through subjective evaluation. Moreover, human beings can efficiently assess the quality of images without using any reference image. However, subjective evaluations are expensive and time-consuming to apply in real-world applications. Therefore, objective IQA designs mathematical models that can predict the quality of images accurately and automatically as human observers.

Generally, based on the availability of a reference image, there are three types of objective IQA measures. The first one is full-reference IQA ($FR-IQA$) where the information of original image is fully available. The second one is reduced-reference IQA (RR-IQA) where the information of original image is partly available. The third one is no-reference IQA ($NR-IQA$) or blind IQA ($BIQA$) where the information of original image is unavailable. State-of-the-art $FR-IQA$ measures, such as $PSNR$, $SSIM$ [6], $MS-SSIM$ [41], MAD [20], VIF [9] and $FSIM$ [12], achieve a very high correlation with human perception. Nevertheless, the degree of improvement is limited by their insufficient consideration of HVS properties. Besides, since any information of distorted images are normally not available in reality, $BIQA$ methods are becoming more considerable importance in image processing field than $FR-IQA$ and RR-IQA methods.

Mostly, successful $BIQA$ approaches use Natural Scene Statistics (NSS) based features. In this chapter, we will explain about recently state-of-the-art $BIQA$ methods.

3.1.1 Blind or Referenceless Image Spatial Quality Evaluator (BRISQUE)

A. Mittal et al. proposed a Blind or Referenceless Image Spatial Quality Evaluator ($BRISQUE$) [30] which exploits an NSS model framework of locally normalized luminance coefficients and measures the quantity of "naturalness" using the parameters of the model. Figure. 3.1 shows the steps involved in calculating BRISQUE. The steps needed for $BRISQUE$ algorithm used for $BIQA$ are the followings.

Step 1: Extract Natural Scene Statistics (NSS)

First, locally normalized luminances are computed via Mean Subtracted Contrast Normalization (MSCN). To calculate the MSCN Coefficients, the image intensity $I(i,j)$ at pixel (i,j) is transformed to the luminance $\widehat{I}(i,j)$.

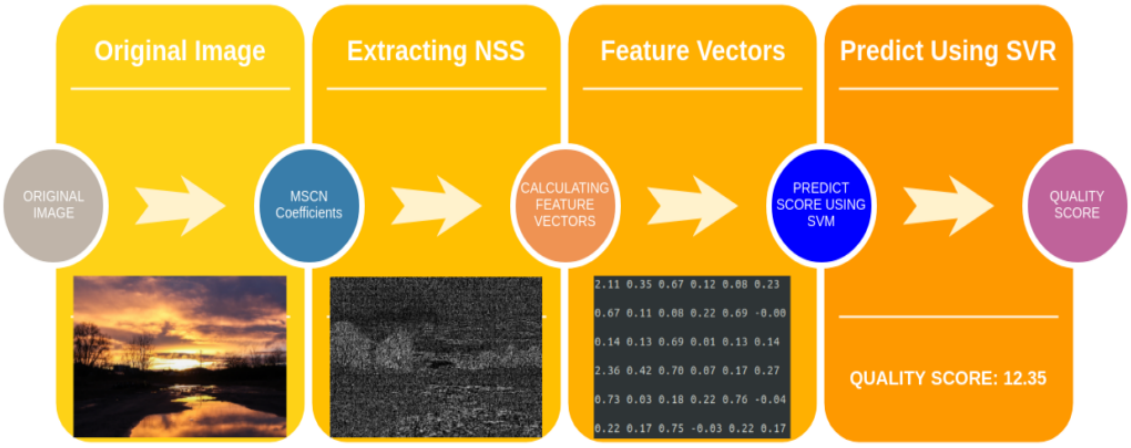


Figure 3.1: Steps to calculate image quality score using BRISQUE Model [30]

$$\widehat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + C} \tag{3.1}$$

Where $i \in 1, 2, \cdots M, j \in 1, 2, \cdots N$ (M and N are height and width respectively). Functions $\mu(i,j)$ and $\sigma(i,j)$ are local mean field and local variance field respectively. Figure 3.2 shows how to calculate MSCN coefficients by visualization.

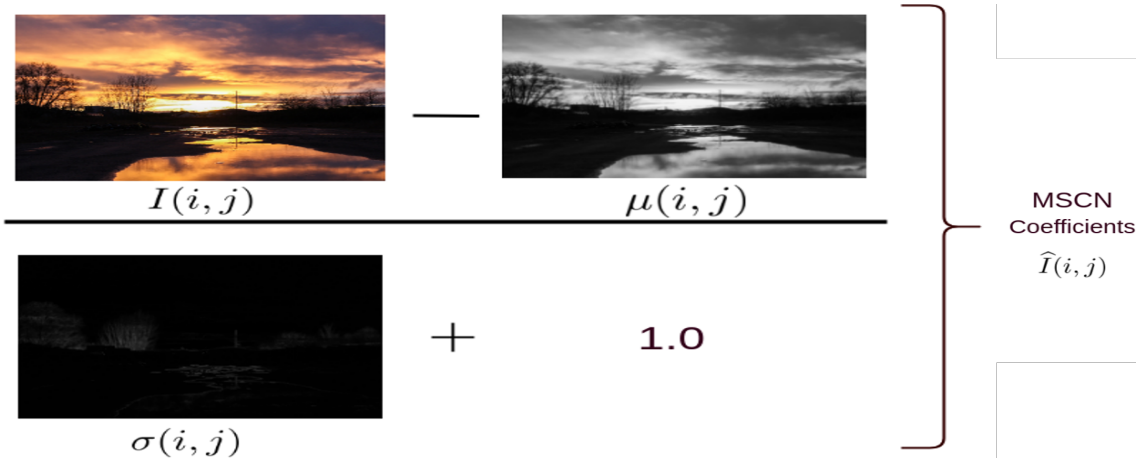


Figure 3.2: Calculation of MSCN coefficients [30]

Step 2 : Calculate Feature Vectors

In this step, the images derived from the original image 1 MSCN image and 4 pairwise product images to capture neighbor relationships (Horizontal, Vertical, Left Diagonal, Right Diagonal) are utilized to calculate feature vectors.

Step 3: Prediction of Image Quality Scores

In this step, images are first converted to feature vectors. Then the feature vectors and quality scores of all images in the training dataset are fed to a learning algorithm, Support Vector Machine (SVM).

3.1.2 Codebook Representation for No-Reference Image Assessment (CORNIA)

Recent development in *BIQA* methods, *CORNIA* [34], promotes extracting features from the spatial domain, which leads to a significant reduction in computation time. In addition, they utilize raw-image-patches extracted from a set of unlabeled images to learn a dictionary in an unsupervised manner. There are four steps in the *CORNIA* approach.

Step 1 : Use raw-image-patches local descriptors in the learning framework instead of hand-crafted features.

Step 2 : Use a codebook based approach which allows to learn highly effective features automatically as shown in Figure 3.3. Key components of the learning framework include (1) Local feature extraction, (2) Codebook construction, (3) Local feature encoding and (4) Feature pooling. The details of these components are described in [34].

Step 3 : Use soft-assignment coding with max pooling for encoding.

Step 4 : Use linear-Support Vector Machine (SVM) for obtaining objective quality scores.

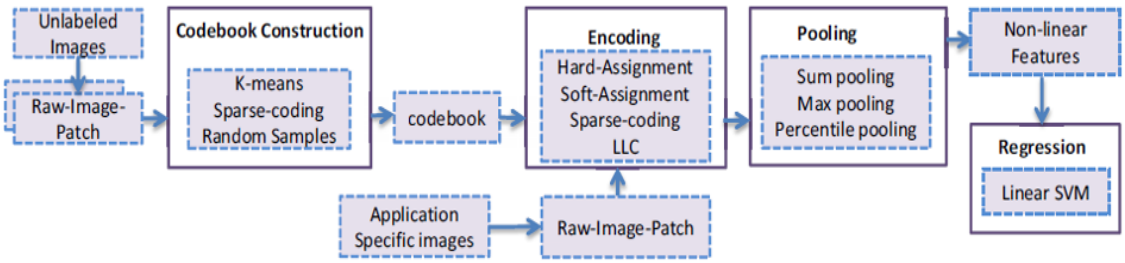


Figure 3.3: Codebook based framework [34]. Note that locality-constrained linear coding (LLC) [24].

3.1.3 High Order Statistics Aggregation (HOSA)

A more recent technique, High Order Statistics Aggregation (HOSA) [76], employs the k-means clustering of normalized image patches and describes them using the low and high order statistics to obtain a small codebook. In HOSA, the soft assignment is used to build image representation and SVR is used for the mapping of features into the subjective scores. HOSA consists of three steps.

First, local normalized image patches are extracted as local features through a regular grid and a codebook containing 100 codewords is constructed by K-means clustering. In addition to the mean of each cluster, the diagonal covariance

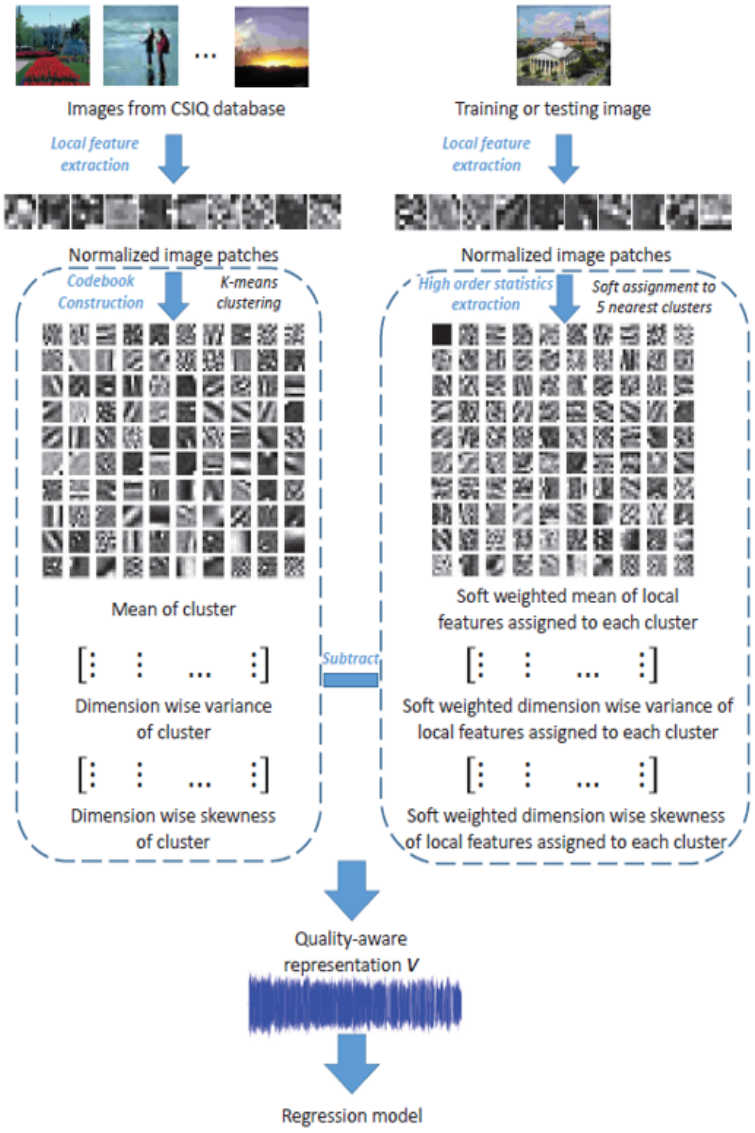


Figure 3.4: Pipeline of the HOSA model [76].

and coskewness (i.e., dimension wise variance and skewness) of clusters are also calculated.

Second, each local feature is softly assigned to several nearest clusters and the differences of high order statistics (mean, variance and skewness) between local features and corresponding clusters are softly aggregated to build the global quality aware image representation.

Third, support vector regression (SVR) is adopted to learn the mapping between perceptual features and subjective opinion scores.

The key components in the HOSA BIQA framework are the extraction of local features, the construction of comprehensive codebook, high order statistics aggregation, and regression. Figure 3.4 illustrates the pipeline of the HOSA model.

3.1.4 Learning from Rankings for NR-IQA and Fine-tuning (RankIQA+FT)

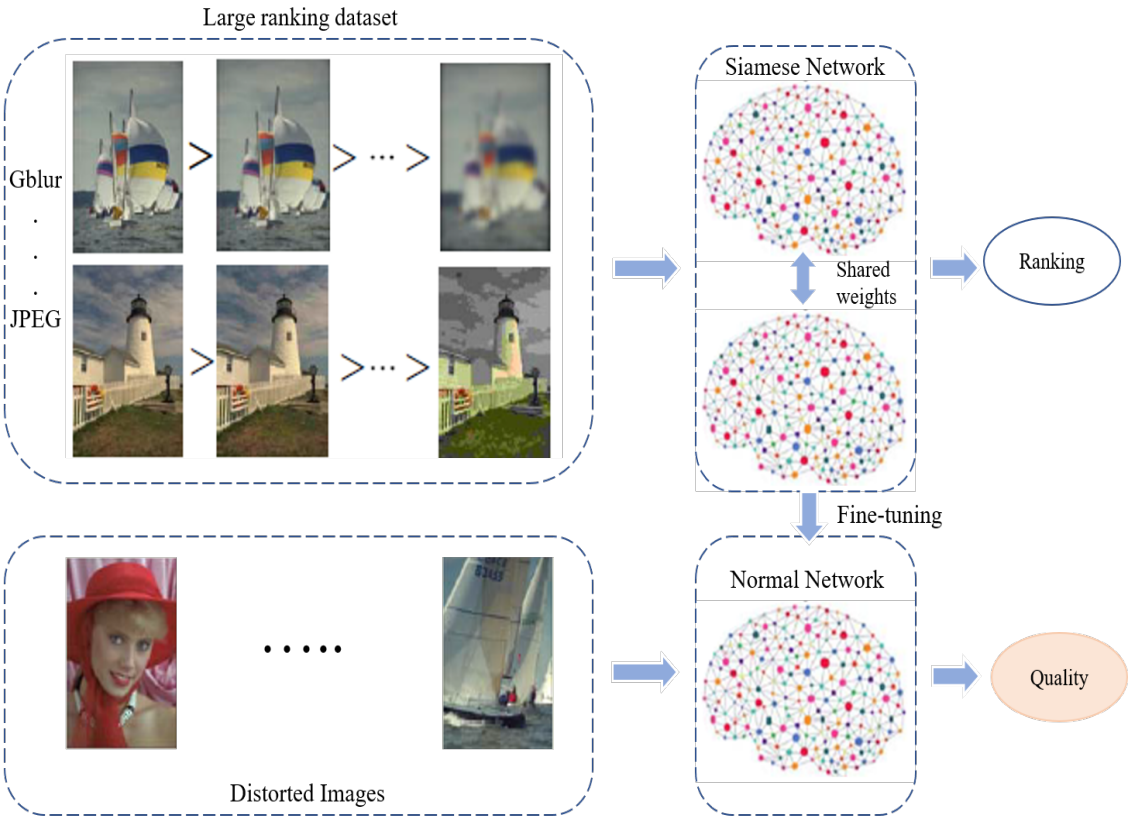


Figure 3.5: Network diagram for RankIQA+FT [79].

In 2017, X. Liu, et al. proposed a no-reference image quality assessment (NR-IQA) approach that learns from rankings (RankIQA) [79]. In that approach, the authors train a Siamese Network to rank images in terms of image quality by using synthetically generated distortions for which relative image quality is known. Then, that ranked image sets is automatically generated without laborious human labeling. Then, they use fine-tuning to transfer the knowledge represented in the trained Siamese Network to a traditional CNN that estimates absolute image quality from single images as shown in Figure 3.5.

Currently, Convolutional Neural Network (CNN) has gained researchers' attention and achieved great success on various computer vision tasks, because this technique has shown superior performance on many standard object recognition benchmarks. However, all existing $BIQA$ methods have still restrictions to get the highest performance.

In this section, therefore, we construct an end-to-end learning mechanism using CNN to overcome such restrictions. One of CNN s advantages is that it can take raw images as input and incorporate feature learning into the training process. Thus, in our work, we take distorted images labelled with Mean Opinion Score (MOS) as inputs and output the related score for each image. Through the experimental results, it is shown that our proposed method outperforms other state -of-the-art IQA methods.

The remainder of this chapter is organized as follows. Section II presents the proposed method in detail. In Section III, we evaluate the performance of the proposed method by experiments. Finally, Section IV concludes the section.

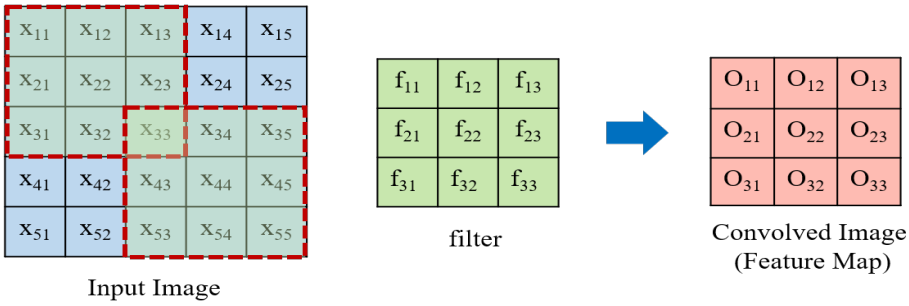
3.2 Convolutional Neural Network

A convolutional neural network (CNN) is a type of artificial neural network used in image recognition and processing that is specifically designed to process pixel data. CNNs are powerful image processing, artificial intelligence (AI) that use deep learning to perform both generative and descriptive tasks, often using machine vison (MV) that includes image and video recognition, along with recommender systems and natural language processing (NLP).

A neural network is a system of hardware or software patterned after the operation of neurons in the human brain. Traditional neural networks are not ideal for image processing and must be fed images in reduced-resolution pieces. CNNs

have their neurons arranged more like those of the frontal lobe, the area responsible for processing visual stimuli in humans and other animals. The layers of neurons are arranged in such a way as to cover the entire visual field avoiding the piecemeal image processing problem of traditional neural networks. A CNN uses a system much like a multilayer perceptron that has been designed for reduced processing requirements. The main advantage of CNN compared to its predecessors is that it automatically detects the important features of inputs without any human supervision. The layers of a CNN consist of an input layer, an output layer and a hidden layer that includes multiple convolutional layers, pooling layers, normalization layers and fully connected layers.

3.2.1 Convolution



$$O_{11} = x_{11}f_{11} + x_{12}f_{12} + x_{13}f_{13} + x_{21}f_{21} + x_{22}f_{22} + x_{23}f_{23} + x_{31}f_{31} + x_{32}f_{32} + x_{33}f_{33}$$
$$O_{12} = x_{12}f_{11} + x_{13}f_{12} + x_{14}f_{13} + x_{22}f_{21} + x_{23}f_{22} + x_{24}f_{23} + x_{32}f_{31} + x_{33}f_{32} + x_{34}f_{33}$$
$$O_{13} = x_{13}f_{11} + x_{14}f_{12} + x_{15}f_{13} + x_{23}f_{21} + x_{24}f_{22} + x_{25}f_{23} + x_{33}f_{31} + x_{34}f_{32} + x_{35}f_{33}$$

$$O_{21} = x_{21}f_{11} + x_{22}f_{12} + x_{23}f_{13} + x_{31}f_{21} + x_{32}f_{22} + x_{33}f_{23} + x_{41}f_{31} + x_{42}f_{32} + x_{43}f_{33}$$

$$\vdots$$

$$O_{33} = x_{33}f_{11} + x_{34}f_{12} + x_{35}f_{13} + x_{43}f_{21} + x_{44}f_{22} + x_{45}f_{23} + x_{53}f_{31} + x_{54}f_{32} + x_{55}f_{33}$$

Figure 3.6: Convolution layer

The main layer of CNN is the convolution layer. Convolution is a mathematical operation to merge two sets of information. In our model, the convolution is applied on the input data using a convolution filter to produce a feature map. Figure 3.1 is an example of convolution operation in 2D using a 3x3 filter.

3.2.2 Pooling

After a convolution operation, we usually perform pooling to reduce the dimensionality. This enables us to reduce the number of parameters, which both shortens the training time and lessen overfitting. Pooling layers downsample each feature map independently. The most common type of pooling is maxpooling which takes the maximum value in the pooling window. Contrary to the convolution operation, pooling has no parameters. It slides a window over its input, and simply takes the maximum value in the window. Similar to a convolution, we specify the window size and stride. Figure 3.2 is a simple example of how maxpooling works with a size of 2x2 and stride 2. Each color denotes a different window. Since both the window size and stride are 2, the windows are not overlapping. Note that this window and stride configuration halves the size of the feature map. This is the main use case of pooling, downsampling the feature map while keeping the important information.

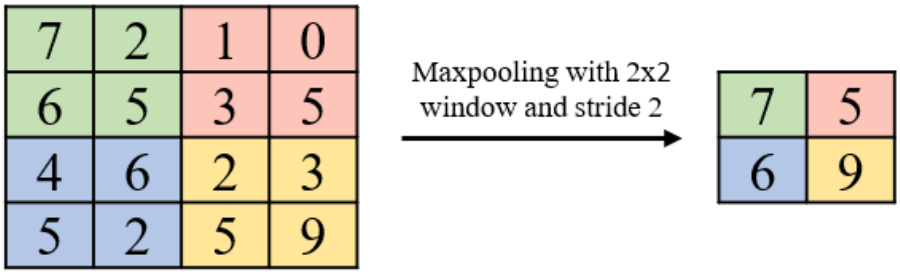


Figure 3.7: Maxpooling

3.2.3 Batch Normalization

To increase the stability of a neural network, batch normalization normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. However, after this scale of activation outputs by some randomly initialized parameters, the weights in the next layer are no longer optimal. Stochastic Gradient Descent (SGD) undoes this normalization if its a way for it to minimize the loss function. Consequently, batch normalization adds two trainable parameters to each layer, so the normalized output is multiplied by a standard deviation parameter (gamma) and add a mean parameter (beta). In other words, batch normalization lets SGD do the denormalization by changing

only these two weights for each activation, instead of losing the stability of the network by changing all the weights.

Batch normalization improves the learning rates, performance, and stability of CNN because it makes sure that theres no activation that has gone really high or really low. Besides, it reduces overfitting because it has a slight regularization effects.

3.2.4 Activation functions

Rectified Linear Unit (ReLU)

ReLU is the most commonly used activation function in neural networks, especially in CNNs. The mathematical formula of ReLU is y = max(0, x) and it is linear for all positive values, and zero for all negative values. Visually, it looks like the following:

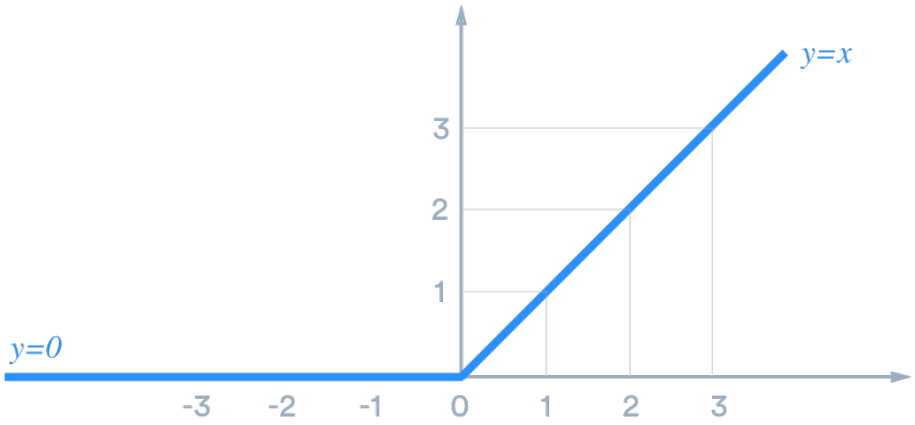


Figure 3.8: ReLU

Parametric ReLU (PReLU)

Unlike ReLU, which always outputs 0 for inputs less than 0, PReLU multiplies inputs less than 0 with a constant value to output results. The value of a, which is a gradient less than 0, is obtained from training. The mathematical formula of PReLU is y = max(0, x) + a min(0, x), where x is the input.

Fully Connected (FC)

At the end of a CNN , the output of the last Pooling Layer acts as input to the so called FC Layer. There can be one or more of these layers and fully connected

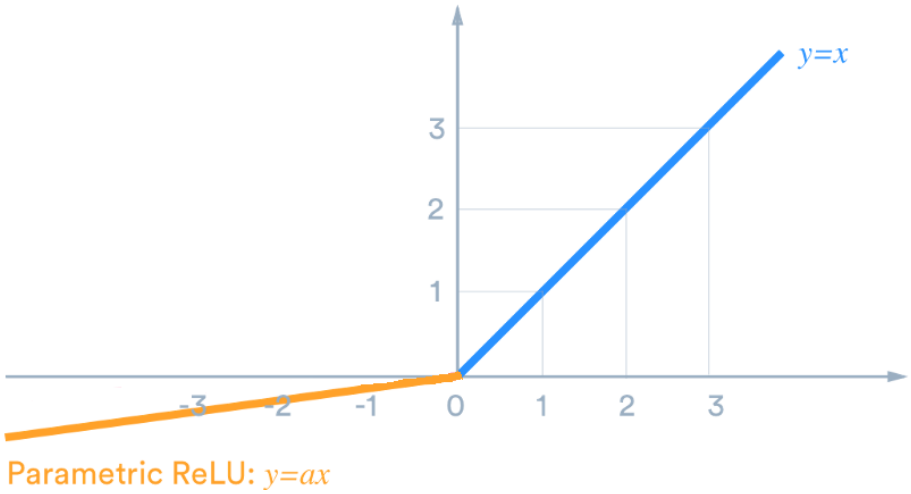


Figure 3.9: PReLU

means that every node in the first layer is connected to every node in the second layer. Moreover, fully Connected layers perform classification based on the features extracted by the previous layers. In our case, this layer contains a SquaredError loss function, which outputs a continuous value for each of the input images.

3.3 Proposed Method

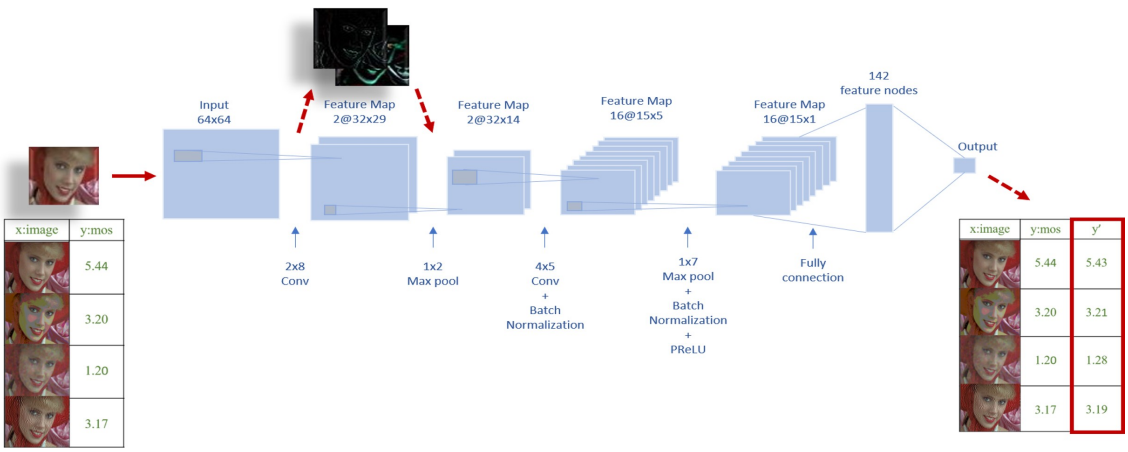


Figure 3.10: Network architecture of the proposed method

The architecture of the proposed method is shown in Figure 3.5. It is a form of end-to-end learning mechanism which takes the whole image as an input and directly outputs the continuous score and then let the network to be able to accept

any arbitrary size of input images. In our case, we utilize 64x64-size distorted images as inputs. Firstly, images are labelled with their related MOS of HVS and then these labelled images are input to the CNN for supervised training. The usage of CNN is motivated by the fact that it is able to learn relevant features from an image at different levels similar to a human brain. As the target scores of input images are assigned with MOS scores of human beings, the trained CNN can realize the reliable objective scores that correspond with subjective scores of HVS . In our CNN , two-dimensional convolution is utilized and the features of the input images are convolved with 2 kernels with a size of 2x8 and each kernel generates a feature map as shown in Figure 3.6.

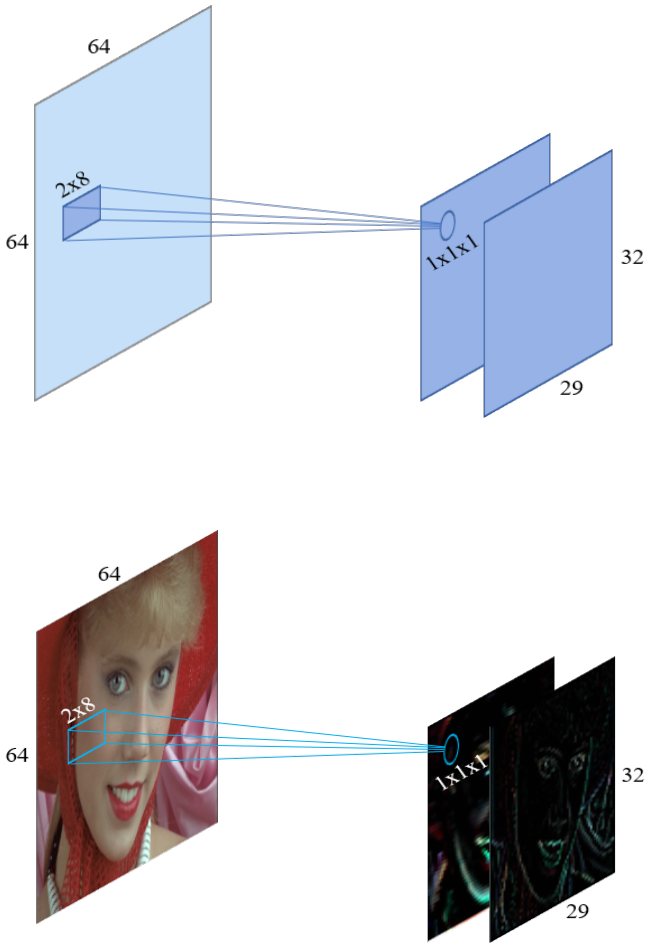


Figure 3.11: The process of first convolution layer of the proposed method

We perform multiple convolutions on an input, each using a different filter and resulting in a distinct feature map. We then stack all these feature maps together and that becomes the final output of the convolution layer.

After convolution operation, we use maxpooling to be able to cover the entire image as quickly as possible exponentially. The pooling size is 1x2 with a stride of 2 and it reduces the spatial dimensions on a CNN by combining the outputs of neuron clusters at one layer into a single neuron in the next layer. Therefore, the computational performance of the CNN becomes high and the overfitting problem becomes less because of less spatial information. In our case, maxpooling down-samples the features that have been extracted by convolution layer.

Then, we apply a second convolution layer with 16 kernels of size 4x5 and then BatchNormalization layer is inserted in our CNN . Batch normalization allows each layer of a network to learn by itself more independently of other layers. In this layer, batch normalization makes sure that there is no activation that has gone really high or really low. Furthermore, it reduces overfitting because it has a slight regularization effect. Similar to dropout, it adds some noise to each hidden layer′s activations. To increase the stability of a neural network, batch normalization normalizes the output of a previous activation layer by subtracting the batch mean and dividing by the batch standard deviation. Thus, inserting this layer after convolution improves accuracy and accelerating convergence of the network model.

The second pooling layer is 1x7, with a stride of 16 and then the second Batch-Normalization layer is applied again to improve the accuracy of the model.

In practice, images are naturally non-linear. When we look at any image, we can find that it contains a lot of non-linear features, for example, the transition between pixels, the borders, the colors, etc. The purpose of applying the rectifier function is to increase the non-linearity in such images. In our network, therefore, we use Rectified Linear Unit (ReLU) in the fully connected layers instead of traditional tanh or sigmoid function. The tanh function is mainly used for classification between two classes. Both tanh and logistic sigmoid activation functions are used in feed-forward networks. ReLU enables the network to train several times faster compared to using traditional activation functions [77]. Note that ReLU only allows nonnegative values to pass. Thus, we use Parametric Rectified Linear Unit (PReLU) nonlinearity in pooling layer instead of ReLU because there are many negative features after normalization. Unlike ReLU which always outputs 0 for inputs less than 0, PReLU multiplies inputs less than 0 with a constant value to output results. Figure 3.7 shows some examples of the estimated scores of the proposed method for each labelled input image.

After the second pooling, the resulting feature maps are fully connected with

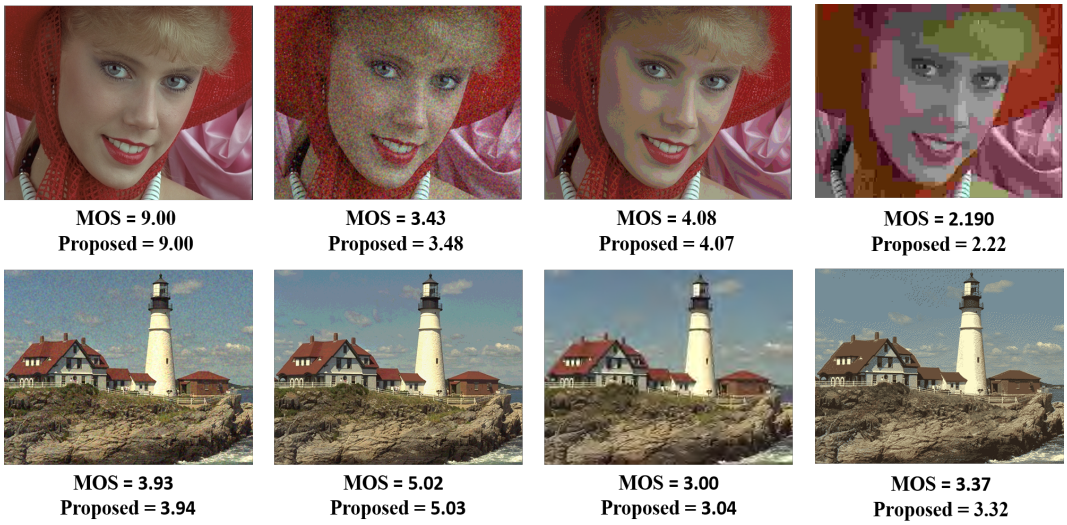


Figure 3.12: Examples of the estimated scores of the proposed method

one layer which consists of 142 feature nodes. All the nodes are fully connected and resulting in one node to produce the output.

Finally, SquaredError loss function is utilized to output the continuous values for each input image. The loss is minimized by using stochastic gradient descent with the standard backpropagation [78].

Other CNNs may contain larger or smaller numbers of feature nodes in the fully connected layer, and greater or fewer fully connected layers. Engineers often experiment to figure out the configuration that produces the best results for their model. In our network, we set the layer structure to get the high accuracy performance.

As a brief statement, the point of Figure. 3.5 is that the proposed system is a very simple idea but it saves computation time and cost and, produces incredible performance for all image quality assessment databases than other previous state-of-the-art ones.

3.4 Implementation

We implement the proposed method using Neural Network Console (NNabla).

3.4.1 NNABLA

Neural Network Libraries by the Tokyo-based Sony is an open-sourced deep learning framework that is intended to be used for research, development as well as production. Neural Network Libraries enable the users to speed-up the computation on GPUs. For examples, from basic to state-of-the-art working examples of Neural Network Libraries are provided as a Windows GUI app. The core libraries can carry out neural network learning and execution and allow deep learning supported technique developments with lower iteration time and highest speed. Neural Network Console also provides a layer of interface functions in Python which allows high-efficiency development as well as easy prototyping for deep learning development. Therefore, developers can develop deep learning technologies creating neural networks more efficiently, in less time, and at a lower cost. Figure 3.8 is the Neural Network Console diagram and Figure 3.9 is the Python source code for the proposed method.

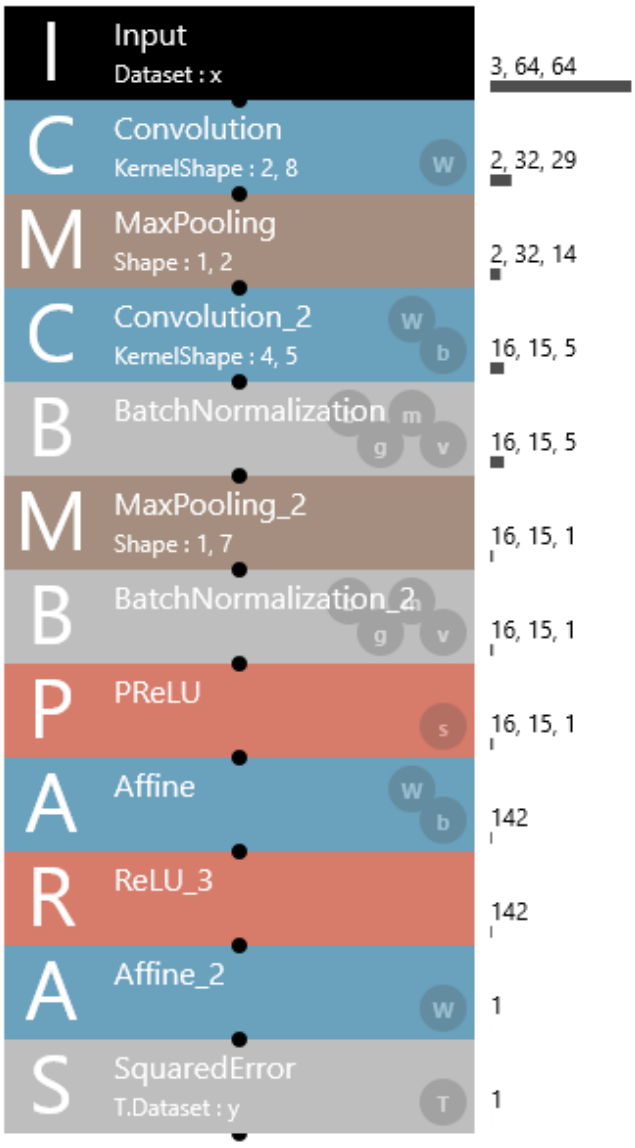


Figure 3.13: Neural Network Console

```python
import nnabla as nn
import nnabla.functions as F
import nnabla.parametric_functions as PF

def network(x, y, test=False):
    # Input:x -> 3,64,64
    # Convolution -> 2,32,29
    h = PF.convolution(x, 2, (2,8), (0,0), (2,2), name='Convolution')
    # MaxPooling -> 2,32,14
    h = F.max_pooling(h, (1,2), (1,2))
    # Convolution_2 -> 16,15,5
    h = PF.convolution(h, 16, (4,5), (0,0), (2,2), name='Convolution_2')
    # BatchNormalization
    h = PF.batch_normalization(h, (1,), 0.9, 0.0001, not test, name='BatchNormalization')
    # MaxPooling_2 -> 16,15,1
    h = F.max_pooling(h, (1,7), (1,7))
    # BatchNormalization_2
    h = PF.batch_normalization(h, (1,), 0.9, 0.0001, not test, name='BatchNormalization_2')
    # PReLU
    h = PF.prelu(h, 1, False, name='PReLU')
    # Affine -> 142
    h = PF.affine(h, (142,), name='Affine')
    # ReLU_3
    h = F.relu(h, True)
    # Affine_2 -> 1
    h = PF.affine(h, (1,), name='Affine_2')
    # SquaredError
    h = F.squared_error(h, y)
    return h
```

Figure 3.14: Python code of the proposed method

3.5 Experimental Results and Discussion

In this section, we compare the overall performance of our proposed method with that of other state-of-the-art not only $FR-IQA$ methods but also $BIQA$ methods. As $FR-IQA$ methods, we utilize $SSIM$ [6], $MS-SSIM$ [41], $PSNR$, MAD [20], VIF [9], $FSIM$ [12]. As $BIQA$ methods, we utilize BLIINDS-II [29], $BRISQUE$ [30], $CORNIA$ [34], HOSA [76], RankIQA+FT [79], NIQE [31], IL-NIQE [75], DIQaM-NR [80], WaDIQaM-NR [80] and $DIIVINE$ [28]. Six Standard benchmarks that we use are A57, TID2008, TID2013, CSIQ, LIVE, and IVC. For detailed performance evaluation, we compare the performance of each distortion type based on the six databases.

3.5.1 Image Databases

The A57 Database [54] has 3 reference images, and 54 distorted images, including six distortion types - Flat allocation; equal distortion contrast at all scales (FA), JPEG compression (JPEG), JPEG2000 compression (JPEG2K), JPEG2000

compression with the Dynamic Contrast-Based Quantization (JPEG2K+DCQ), Gaussian blurring (GB), and additive Gaussian white noise (WGN) - at 5 different levels. The subjective quality scores used for this database are DMOS, ranging from 0 to 1.

The Tampere Image Database (TID2008) [56] includes 25 reference images and 1700 distorted images. Each reference image is distorted using 17 types of distortions as shown in Table 3.1, with 4 different levels. The subjective quality scores provided for this database are MOS , ranging from 0 to 9.

Table 3.1: Types of distortions in TID2008

No	**Types**
1	Additive Gaussian noise
2	Additive noise in color component
3	Spatially correlated noise
4	Masked noise
5	High frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change

Tampere Image Database (TID2013) [57] includes 25 reference images, 24 types of distortions for each reference image as shown in Table 3.2, and 5 different levels for each type of distortion. The whole database contains 3000 distorted images. MOS is provided in this database, and the scores range from 0 to 9.

Categorical Image Quality (CSIQ) [55] contains 900 distorted images. Each image is distorted by 6 types of distortions - JPEG compression (JPEG), JPEG2000 compression (JPEG2K), global contract decrements (CTD), additive Gaussian white noise (WGN), additive pink Gaussian noise (PGN), and Gaussian blurring

Table 3.2: Types of distortions in TID2013

No	Types
1	Additive Gaussian noise
2	Additive noise in color component
3	Spatially correlated noise
4	Masked noise
5	High frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change
18	Change of color saturation
19	Multiplicative Gaussian noise
20	Comfort noise
21	Lossy compression of noisy images
22	Image color quantization with dither
23	Chromatic aberrations
24	Sparse sampling and reconstruction

(GB) from 4 to 5 different levels. The score ratings with 0-to-1 scale are reported in the form of DMOS.

The LIVE Image Quality Database [58] has 29 reference images and 779 distorted images, including five distortion types - JPEG2000 compression (JPEG2K), JPEG compression (JPEG), white noise in the RGB components (WGN), Gaussian blur (GB), and transmission errors in the JPEG2000 bit stream using a fast-fading Rayleigh channel model (FF). The subjective quality scores provided in this database are DMOS, ranging from 0 to 100.

The IVC Database [59] has 10 reference images, and 185 distorted images, including 4 types of distortions Gaussian blurring (GB), JPEG2000 compression

(JPEG2K), JPEG compression (JPEG), and locally adaptive resolution (LAR) coding. The subjective quality scores provided in this database are MOS , ranging from 1 to 5.

3.5.2 Training and Testing

For training and testing, the models which are trained on the eighty percent images of each database are tested on the totally different twenty percent of each one.

For cross-database evaluation, a model trained on full LIVE database is evaluated on subsets of TID2013 and CSIQ databases which contain only four types of distortions –White Noise, Gaussian Blur, JPEG and JPEG2000 shared between the three databases.

3.5.3 Performance Evaluation

To fairly compare the performance, we employ the widely used correlation coefficients, which are the Pearson Linear Correlation Coefficient (PLCC) and the Spearman's Rank Ordered Correlation Coefficient (SROCC). They measure the correlation between the predicted scores and the ground truth scores of images. PLCC is defined as

$$\text{PLCC} = \frac{1}{N-1}\sum_{n=1}^{N}\left(\frac{x_n-\bar{x}}{d_x}\right)\left(\frac{y_n-\bar{y}}{d_y}\right) \tag{3.2}$$

where N is the data size, x_n and y_n are the single subjective and objective scores with index n, respectively. d_x, d_y denote the standard deviations of subjective and objective scores, respectively and $\bar{x}$, $\bar{y}$ denote the mean values of subjective and objective scores, respectively. SROCC is defined as

$$\text{SROCC} = 1 - \frac{6\sum D^2}{N(N^2-1)} \tag{3.3}$$

where D is the difference between the subjective and objective score ranks.

3.5.4 Experimental Results

In Tables 3.3-3.8, we compare the PLCC and SROCC values of the proposed method with those of other sixteen state-of-the-art IQA methods based on six databases for the whole databases. Through the experimental results, we see that

Table 3.3: PLCC and SROCC performance comparison on A57 database. $FR-IQA$ methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.802	0.807
MS-SSIM	0.560	0.667
PSNR	0.635	0.619
MAD	0.501	0.850
VIF	0.616	0.622
FSIM	0.557	0.496
BLIINDS-II	-	-
BRISQUE	-	-
CORNIA	-	-
HOSA	-	-
RankIQA+FT	-	-
NIQE	-	-
IL-NIQE	-	-
DIQaM-NR	-	-
WaDIQaM-NR	-	-
DIIVINE	-	-
Proposed	**0.947**	**0.923**

Table 3.4: PLCC and SROCC performance comparison on TID2008 database. FR-IQA methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.772	-
MS-SSIM	0.839	-
PSNR	0.536	-
MAD	0.831	-
VIF	0.806	-
FSIM	0.871	-
BLIINDS-II	-	-
BRISQUE	-	-
CORNIA	0.837	0.813
HOSA	-	-
RankIQA+FT	-	-
NIQE	-	-
IL-NIQE	-	-
DIQaM-NR	-	-
WaDIQaM-NR	-	-
DIIVINE	-	-
Proposed	**0.987**	**0.986**

Table 3.5: PLCC and SROCC performance comparison on TID2013 database. FR-IQA methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.650	0.627
MS-SSIM	0.729	0.786
PSNR	0.639	0.639
MAD	0.807	0.838
VIF	0.772	0.677
FSIM	0.821	0.802
BLIINDS-II	0.628	0.536
BRISQUE	0.651	0.562
CORNIA	0.613	0.549
HOSA	-	0.728
RankIQA+FT	-	0.780
NIQE	0.426	0.317
IL-NIQE	-	-
DIQaM-NR	0.855	0.835
WaDIQaM-NR	0.787	0.761
DIIVINE	0.654	0.549
Proposed	**0.981**	**0.977**

Table 3.6: PLCC and SROCC performance comparison on CSIQ database. FR-IQA methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.859	0.876
MS-SSIM	0.784	0.873
PSNR	0.800	0.806
MAD	0.935	0.928
VIF	0.925	0.919
FSIM	0.800	0.919
BLIINDS-II	0.832	0.780
BRISQUE	0.817	0.775
CORNIA	0.781	0.714
HOSA	-	-
RankIQA+FT	-	-
NIQE	0.725	0.627
IL-NIQE	0.865	0.822
DIQaM-NR	-	-
WaDIQaM-NR	-	-
DIIVINE	-	-
Proposed	**0.976**	**0.964**

Table 3.7: PLCC and SROCC performance comparison on LIVE database. FR-IQA methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.906	0.913
MS-SSIM	0.940	0.952
PSNR	0.856	0.866
MAD	0.967	0.967
VIF	0.960	0.964
FSIM	0.954	0.963
BLIINDS-II	0.927	0.924
BRISQUE	0.931	0.933
CORNIA	0.935	0.942
HOSA	-	-
RankIQA+FT	0.982	0.981
NIQE	0.908	0.908
IL-NIQE	0.906	0.902
DIQaM-NR	0.972	0.960
WaDIQaM-NR	0.963	0.954
DIIVINE	0.917	0.916
Proposed	**0.982**	**0.973**

Table 3.8: PLCC and SROCC performance comparison on IVC database. FR-IQA methods are italicized.

IQA Method	**PLCC**	**SROCC**
SSIM	0.938	0.948
MS-SSIM	0.940	0.952
PSNR	0.870	0.876
MAD	0.967	0.967
VIF	0.960	0.964
FSIM	0.954	0.963
BLIINDS-II	-	-
BRISQUE	-	-
CORNIA	-	-
HOSA	-	-
RankIQA+FT	-	-
NIQE	-	-
IL-NIQE	-	-
DIQaM-NR	-	-
WaDIQaM-NR	-	-
DIIVINE	-	-
Proposed	**0.988**	**0.981**

Table 3.9: PLCC performance evaluation of each distortion type on A57 database

Distortion	**MS-SSIM**	**MAD**	**FSIM**	**Proposed**
FA	0.619	0.546	0.623	**0.960**
JPEG	0.520	0.473	0.559	**0.963**
JPEG2K	0.490	0.473	0.484	**0.935**
JPEG2K+DCQ	0.456	0.524	0.545	**0.956**
GB	0.485	0.567	0.576	**0.861**
WGN	0.683	0.801	**0.843**	0.794

Table 3.10: SROCC performance evaluation of each distortion type on A57 database

Distortion	**MS-SSIM**	**MAD**	**FSIM**	**Proposed**
FA	0.672	0.656	0.706	**0.933**
JPEG	0.372	0.539	0.672	**0.967**
JPEG2K	0.422	0.556	0.622	**0.917**
JPEG2K+DCQ	0.539	0.422	0.372	**0.833**
GB	0.489	0.422	0.389	**0.600**
WGN	0.589	0.389	0.522	**0.833**

the proposed method significantly outperforms other state-of-the-art blind IQA methods and even full-reference IQA ones for all databases.

In Tables 3.9 and 3.10, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other state-of-the-art $FR-IQA$ methods for each distortion type of A57 database. Incredibly, the proposed method is more robust for almost all types of distortions than state-of-the-art even $FR-IQA$ methods.

Likewise, in Tables 3.11 and 3.12, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other state-of-the-art $FR-IQA$ methods for each distortion type of TID2008 database. Significantly, the proposed method outperforms all types of distortions than state-of-the-art even $FR-IQA$ methods.

In Table 3.13, we compare the SROCC values of our proposed method with those of other state-of-the-art $BIQA$ methods for each distortion type of TID2013 database to be fair performance comparison. Unfortunately, we cannot show the PLCC performance comparison of TID2013 for each distortion type because there is still no description of such PLCC values on other papers. The experimental results validated that the proposed method is more robust for all various types of

Table 3.11: PLCC performance evaluation of each distortion type on TID2008 database

Dist-ortion	**MS-SSIM**	**MAD**	**FSIM**	**Proposed**
1	0.774	0.817	0.783	**0.952**
2	0.798	0.827	0.820	**0.926**
3	0.791	0.860	0.794	**0.971**
4	0.809	0.757	0.769	**0.914**
5	0.852	0.893	0.838	**0.968**
6	0.723	0.042	0.674	**0.936**
7	0.764	0.798	0.783	**0.974**
8	0.908	0.923	0.908	**0.977**
9	0.897	0.961	0.932	**0.992**
10	0.900	0.949	0.925	**0.992**
11	0.850	0.973	0.955	**0.996**
12	0.814	0.856	0.842	**0.981**
13	0.809	0.830	0.788	**0.979**
14	0.666	0.824	0.726	**0.984**
15	0.890	0.801	0.841	**0.980**
16	0.683	0.571	0.670	**0.921**
17	0.589	0.257	0.729	**0.983**

Table 3.12: SROCC performance evaluation of each distortion type on TID2008 database

Dist-ortion	**MS-SSIM**	**MAD**	**FSIM**	**Proposed**
1	0.803	0.839	0.858	**0.955**
2	0.813	0.826	0.852	**0.922**
3	0.828	0.867	0.848	**0.971**
4	0.814	0.734	0.802	**0.898**
5	0.845	0.887	0.910	**0.955**
6	0.754	0.064	0.746	**0.937**
7	0.796	0.816	0.855	**0.972**
8	0.940	0.920	0.947	**0.981**
9	0.923	0.943	0.960	**0.989**
10	0.897	0.928	0.928	**0.976**
11	0.883	0.971	0.978	**0.992**
12	0.820	0.866	0.871	**0.982**
13	0.839	0.840	0.855	**0.979**
14	0.692	0.829	0.750	**0.974**
15	0.896	0.796	0.848	**0.978**
16	0.719	0.517	0.670	**0.869**
17	0.554	0.272	0.648	**0.967**

Table 3.13: SROCC performance evaluation of each distortion type on TID2013 database

Dist-ortion	**BLII-NDS-II**	**BRIS-QUE**	**COR-NIA**	**HO-SA**	**Rank-IQA+ FT**	**Pro-posed**
1	0.714	0.630	0.341	0.853	0.667	**0.976**
2	0.728	0.424	0.196	0.625	0.620	**0.967**
3	0.825	0.727	0.689	0.782	0.821	**0.989**
4	0.358	0.321	0.184	0.368	0.365	**0.915**
5	0.852	0.775	0.607	0.905	0.760	**0.969**
6	0.664	0.669	0.014	0.775	0.736	**0.974**
7	0.780	0.592	0.673	0.810	0.783	**0.993**
8	0.852	0.845	0.896	0.892	0.809	**0.987**
9	0.754	0.553	0.787	0.870	0.767	**0.991**
10	0.808	0.742	0.875	0.893	0.866	**0.986**
11	0.862	0.799	0.911	0.932	0.878	**0.994**
12	0.251	0.301	0.310	0.747	0.704	**0.983**
13	0.755	0.672	0.625	0.701	0.810	**0.983**
14	0.081	0.175	0.161	0.199	0.512	**0.981**
15	0.371	0.184	0.096	0.327	0.622	**0.960**
16	0.159	0.155	0.008	0.233	0.268	**0.916**
17	0.082	0.125	0.423	0.294	0.613	**0.983**
18	0.109	0.032	0.055	0.119	0.662	**0.988**
19	0.699	0.560	0.259	0.782	0.619	**0.980**
20	0.222	0.282	0.606	0.532	0.644	**0.983**
21	0.451	0.680	0.555	0.835	0.800	**0.992**
22	0.815	0.804	0.592	0.855	0.779	**0.990**
23	0.568	0.715	0.759	0.801	0.629	**0.982**
24	0.856	0.800	0.903	0.905	0.859	**0.994**

Table 3.14: SROCC performance evaluation of each distortion type on CSIQ database

Dist-tion	**BLIINDS-II**	**BRI-SQUE**	**COR-NIA**	**Pro-posed**
WGN	0.801	**0.925**	0.746	0.817
GB	0.892	0.903	0.917	**0.963**
CTD	0.012	0.024	0.302	**0.972**
PGN	0.379	0.253	0.420	**0.948**
JPEG	0.900	0.909	0.908	**0.978**
JPEG2K	0.895	0.867	0.914	**0.982**

Table 3.15: PLCC performance evaluation of each distortion type on LIVE database

Dist-ortion	**BLII-NDS-II**	**BRI-SQUE**	**COR-NIA**	**Rank-IQA +FT**	**DII-VINE**	**Pro-posed**
JPEG2K	0.935	0.923	0.951	0.975	0.922	**0.984**
JPEG	0.968	0.973	0.965	**0.986**	0.921	0.983
WGN	0.980	0.985	0.987	**0.994**	0.988	0.986
GB	0.938	0.951	0.968	0.988	0.923	**0.989**
FF	0.896	0.903	0.917	0.960	0.888	**0.994**

Table 3.16: SROCC performance evaluation of each distortion type on LIVE database

Dist-ortion	**BLII-NDS-II**	**BRI-SQUE**	**COR-NIA**	**Rank-IQA +FT**	**DII-VINE**	**Pro-posed**
JPEG2K	0.929	0.914	0.943	**0.970**	0.913	0.966
JPEG	0.942	0.965	0.955	0.978	0.910	**0.981**
WGN	0.969	0.979	0.976	**0.991**	0.984	0.976
GB	0.923	0.951	0.969	**0.988**	0.921	0.910
FF	0.889	0.887	0.906	0.954	0.863	**0.966**

Table 3.17: PLCC performance evaluation of each distortion type on IVC database

Dist-ortion	**MS-SSIM**	**MAD**	**FSIM**	**Pro-posed**
GB	0.853	0.971	0.890	**0.979**
JPEG2K	0.882	0.916	0.930	**0.993**
JPEG	0.915	0.940	0.960	**0.993**
LAR	0.927	0.956	0.935	**0.980**

Table 3.18: SROCC performance evaluation of each distortion type on IVC database

Dist-ortion	**MS-SSIM**	**MAD**	**FSIM**	**Pro-posed**
GB	0.853	0.963	0.963	**0.988**
JPEG2K	0.938	0.910	0.962	**0.973**
JPEG	0.914	0.896	**0.981**	0.979
LAR	0.947	0.949	0.886	**0.961**

distortions than other state-of-the-art $BIQA$ methods.

In Table 3.14, we could compare the SROCC values of our proposed method with those of other three $BIQA$ methods for each distortion type on CSIQ database because there is still the SROCC values of only the three $BIQA$ methods for performance comparison of each distortion type on CSIQ database. Through the experimental results, we can observe that our proposed method outperforms consistently well almost all types of distortions on CSIQ database than other previous excellent $BIQA$ methods.

Similarly, in Tables 3.15 and 3.16, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other $BIQA$ methods for each distortion type of LIVE database. Considerably, the proposed method is robust for some types of distortions than other $BIQA$ methods.

In Tables 3.17 and 3.18, we compare the PLCC and SROCC values, respectively, of our proposed method with those of other $FR-IQA$ methods for each distortion type of IVC database. Obviously, the proposed method exceeds almost all types of distortions than $FR-IQA$ methods.

3.5.5 Cross validation

In Table 3.19, we extend the cross-database validation and compare the SROCC performance of the proposed method with other state-of-the-art $BIQA$ methods. The model trained on the full LIVE database is evaluated on the subsets of CSIQ and TID2013 databases based on the same types of distortions (JPEG, JPEG2K, GB and WGN) for three databases.

Table 3.19: SROCC performance in Cross-database evaluation. All models are trained on full LIVE database and evaluated on subsets of CSIQ and TID2013 databases based on the four types of distortions shared with LIVE database

Method	**CSIQ**	**TID2013**
BRISQUE	0.899	0.882
CORNIA	0.899	0.892
DIQaM-NR	**0.908**	0.867
WaDIQaM-FR	0.866	0.872
Proposed	0.781	**0.969**

In the experiments, the maximum number of epochs for CNN is 950, the batch size is 16 and the learning rate is 0.1. Experimental results demonstrated that

the proposed method is slightly weak for CSIQ subset but it shows superior performance compared to all other methods on the TID2013 subset for independent cross-database validation.

3.6 Conclusion

In this chapter, we developed an end-to-end convolutional neural network that can take the input images of any arbitrary size and can directly output their predicted quality scores for blind image quality assessment. Experimental results verified that the proposed method achieves superior performance not only for the whole database but also for several types of distortions compared to other state-of-the-art $BIQA$ methods and even $FR-IQA$ methods.

Chapter 4

Conclusion and Future Work

This chapter concludes the thesis with a summary of our study and the future research is also described in this Chapter.

4.0.1 Summary of the Research

In this study, we have proposed a combination method which has dissimilar feature-scores for $FR-IQA$ employing PSO algorithm and an end-to-end learning mechanism by using CNN for $BIQA$.

As there is no perfect single method or perfect combination method which is very robust in the prediction of the quality of images of all databases, we introduce a very simple and robust combination method. Among state-of-the-art $FR-IQA$ methods up to now, we select the most suitable three IQA indices utilizing the BIRD algorithm that is used to select the most appropriate method for combination. Then, we optimize the parameter values by using the PSO algorithm which is based on the swarm intelligence. The experimental results have demonstrated that the proposed method is a superior one in comparison to other not only conventional single methods but also previous combination methods. Indeed, in our combination, we decide MAD , $MS-SSIM$ and $FSIM$ as the most suitable methods which has dissimilar features for the combination among the best recognized $FR-IQA$ measures until now. However, further extension of the proposed method could involve using other best recognized $FR-IQA$ measures for combination in the future.

On the other hand, it is very difficult to get the original images in practice. In the situation, we need to measure the quality of distorted images without referencing the original images and it is the most difficult IQA measure in image

processing field. Thus, we consider to construct an end-to-end learning mechanism which takes the whole image as an input and directly outputs the continuous score by using the CNN for $BIQA$ as our second proposed method. First, we take distorted images labelled with MOS scores as inputs and output the related score for each input image. Experimental results verified that our proposed method outperforms other state -of-the-art $BIQA$ methods and even $FR-IQA$ ones.

4.0.2 Future Work

In the future, many biometric methods are based on images, including images of faces, fingerprints, palmprints, hand shapes, and handwritings. In practice, the acquisition process of these images may not be perfect, and thus the biometric systems may have to work under the conditions of noisy, distorted, or partially impaired images. In these application scenarios, it would be useful to know the level of quality degradations of these images and what recognition accuracy can be expected under such quality degradations. In this study, therefore, we have proposed robust IQA measures which are different from traditional quality evaluation of IQA measures to benefit a wide variety of real applications. However, further research is needed to improve IQA measures based on the current challenges and for future challenges.

Bibliography

[1] A. McAndrew, "An introduction to digital image processing with matlab notes for SCM2511 image processing", *School of Computer Science and Mathematics, Victoria University of Technology*, pp.1-264, 2004.

[2] A. C. Bovik, "The essential guide to image processing", *Academic Press*, 2009.

[3] A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity", *IEEE Trans.*, vol.21, no.4, pp.1500-1512, 2012.

[4] N. Damera-Venkata et al., "Image quality assessment based on a degradation model", *IEEE Trans. and Image Proc.*, vol.9, no.4, pp.636-650, 2000.

[5] Z. Wang and A. C. Bovik, "A universal image quality index", *IEEE Signal Proc. Lett.*, vol.9, no.3, pp.81-84, 2002.

[6] Z. Wang et al., "Image quality assessment: From error visibility to structural similarity", *IEEE Trans. Image Proc.*, vol.13, no.4, pp.600-612, 2004.

[7] Z. Wang, E. P. Simoncelli and A. C.Bovik, "Multi-scale structural similarity for image quality assessment", *In: Proc. IEEE Int. Conf. on Signals, Systems, and Computers, (ASILOMAR)*, 2003, pp.1398-1402.

[8] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment", *IEEE Trans. and Image Proc.*, vol.20, no.5, pp.1185-1198, 2011.

[9] H. R. Sheikh and C. Bovik, "Image information and visual quality", *IEEE Trans. and Image Proc.*, vol.15, no.2, pp.430-444, 2006.

[10] H. R. Sheikh, A. C. Bovik and G. Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Trans. and Image Proc.*, vol.14, no.12, pp.2117-2128, 2005.

[11] L. Zhang and X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms", *In: Proc. IEEE Int. Conf. on Image Processing (ICIP)*, 2010.

[12] L. Zhang, X. Mou, D. Zhang, "FSIM: A feature similarity index for image quality assessment", *IEEE Trans. and Image Proc.*, vol.20, no.8, pp.2378-2386, 2011.

[13] L. Zhang and H. Li *SR-SIM: A fast and high performance IQA index based on spectral residual*, In: Process. IEEE Int. Conf. on Image Processing (ICIP), 2012.

[14] L. Zhang, Y. Shen and H. Li, "VSI: A visual saliency-induced index for perceptual image quality assessment", *IEEE Trans. Image Proc.*, vol.23, no.10, pp.4270-4281, 2014.

[15] A. Liu, W. Lin and M. Narwaria, "Image quality assessment based on gradient similarity", *IEEE Trans. and Image Proc.*, vol.21, no.4, pp.1500-1512, 2012.

[16] J. Wu, W. Lin and G. Shi, "Image quality assessment with degradation on spatial structure", *IEEE Signal Proc. Lett.*, vol.21, no.4, pp.437-440, 2014.

[17] F. Wang et al., "An object-distortion based image quality similarity", *IEEE Signal Proc. Lett.*, vol.22, no.10, pp.1534-1537, 2015.

[18] F. Zhou et al., "Image quality assessment based on inter-patch and intra-patch similarity", *PLoS ONE*, vol.10, no.3, 2015.

[19] M. Liu and X. Yang, "A new image quality approach based on decision fusion", *In: Proc. Int. Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2009, pp.10-14.

[20] E. C. Larson and D. M. Chandler, "Most apparent distortion: Full-reference image quality assessment and the role of strategy", *J Electron Imaging*, vol.19, no.1, 2010.

[21] K. Okarma, "Combined full-reference image quality metric linearly correlated with subjective assessment", *In: Artificial Intelligence and Soft Computing, Springer*, pp.539-546, 2010.

[22] K. Okarma, "Extended hybrid image similaritycombined full-reference image quality metric linearly correlated with subjective scores", *Elektronika ir Elektrotechnika*, vol.19, no.10, pp.129-132, 2013.

[23] M. Oszust, "Decision fusion for image quality assessment using an optimization approach", *IEEE Signal Proc. Lett.*, vol.23, no.1, pp.65-69, 2016.

[24] A. Lahouhou, E. Viennet and A. Beghdadi, "Selecting low-level features for image quality assessment by statistical methods", *CIT.*, vol.18, no.2, 2010.

[25] P. Peng and Z. N. Li, "A mixture of experts approach to multi-strategy image quality assessment", *Image Analysis and Recognition, vol.7324 of Lecture Notes in Computer Science, Springer*, pp.123-130, 2012.

[26] P. Peng and Z. N. Li, "Regularization of the structural similarity index based on preservation of edge direction", *IEEE International Conf.*, 2012, pp.2127-2132.

[27] I. C. Trelea, "The particle swarm optimization algorithm convergence analysis and parameter selection", *Information Proc. Lett.*, vol.85, pp.317-325, 2003.

[28] A. K. Moorthy and A. C. Bovik, "Blind image quality assessment: from natural scene statistics to perceptual quality", *IEEE Trans. on Image Proc.*, vol.20, no.12, pp.3350-3364, 2011.

[29] M. A. Saad, A. C. Bovik and C. Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain", *IEEE Trans. on Image Proc.*, vol.21, no.8, pp.3339-3352, 2012.

[30] A. Mittal, A. K. Moorthy and A. C. Bovik, "No-reference image quality assessment in the spatial domain", *IEEE Trans. on Image Proc.*, vol.21, no.12, pp.4695-4708, 2012.

[31] A. Mittal, R. Soundararajan and A. C. Bovik, " Making a , 'completely blind' image quality analyzer", *IEEE Signal Proc. Lett.*, vol.20, no.3, pp.209-212, 2013.

[32] D. Ghadiyaram and A. Bovik, "LIVE in the wild image quality challenge database", 2015.

[33] D. Ghadiyaram and A. C. Bovik, "Massive online crowdsourced study of subjective and objective picture quality", *IEEE Trans. on Image Proc.*, vol.25, no.1, pp.372-387, 2016.

[34] P. Ye et al., "Unsupervised feature learning framework for no-reference image quality assessment", *IEEE Conf. on Computer Vision and Pattern Recognition*, 2012, pp.1098-1105.

[35] P. Zhang et al., "SOM: semantic obviousness metric for image quality assessment", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp.2394-2402.

[36] L. Zhang et al., "Training quality-aware filters for no-reference image quality assessment", *IEEE MultiMedia*, vol.21, no.4, pp.67-75, 2014.

[37] L. Kang et al., "Convolutional neural networks for no-reference image quality assessment", *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp.1733-1740.

[38] J. Kim and S. Lee, "Fully deep blind image quality predictor", *IEEE Journal of Selected Topics in Signal Proc.*, vol.11, no.1, pp.206-220, 2017.

[39] S. Bosse et al., "A deep neural network for image quality assessment", *IEEE International Conf. on Image Proc. (ICIP)*, 2016, pp.3773-3777.

[40] S. Bosse et al., "Neural network-based full-reference image quality assessment", *Picture Coding Symposium (PCS)*, 2016, pp.1-5.

[41] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment", *Proc. Asilomar Conf. on Signals, Systems and Computers*, 2004, vol.2, pp.1398-1402.

[42] G. H. Chen, C. L. Yang and S. L. Xie, "Gradient-based structural similarity for image quality assessment", *Proc. IEEE International Conf. on Image Proc.*, 2006, pp.2929-2932.

[43] D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images", *IEEE Trans. Image Proc.*, vol.16, no.9, pp.2284-2298, 2007.

[44] J. Zhu and N. Wang, "Image quality assessment by visual gradient similarity", *IEEE Trans. Image Proc.*, vol.21, no.3, pp.919-933, 2012.

[45] K. Isono and T. Shimamura, "Exponentiated combination of two scores for image quality assessment", *Proc. IEEE Region 10 Conf.*, 2013, pp.1-4.

[46] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics", *IEEE Trans. Image Proc.*, vol.14, no.12, pp.2117-2128, 2005.

[47] K. Egiazarian et al., "A new full-reference quality metrics based on HVS", *Proc. 2nd International Workshop Video Proc. Quality Metrics.*, 2006, pp.1-4.

[48] K. Ishiyama, Y. Sugiura and T. Shimamura, "Optimized three scores combination for image quality assessment", *Proc. IEEE Asia Pacific Conf.*, 2016.

[49] T. Liu, W. Lin and C. J. Kuo, "Image quality assessment using multi-method fusion", *IEEE Trans. Image Proc.*, vol.22, no.5, pp.1793-1807, 2013.

[50] M. Esmaeilpour, A. Mansouri and A. Mahmoudi-Aznaveh, "A new SVD-based image quality assessment", *Machine Vision and Image Proc.*, 2013.

[51] M. Oszust, "Full-reference quality assessment with linear combination of genetically selected quality measures", *PLos ONE.*, 2016.

[52] I. C. Trelea, "The particle swarm optimization algorithm: convergence analysis and parameter selection", *Information Proc. Lett.*, vol.85, pp.317-325, 2003.

[53] R. Hassan et al., "A comparison of particle swarm optimization and the genetic algorithm", *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conf.*, 2005.

[54] A57 Database, [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html.

[55] Categorical Image Quality (CSIQ) Database, [Online]. Available: http://vision.okstate.edu/csiq.

[56] Tampere Image Database, [Online]. Available: http://www.ponomarenko.info/tid2008.htm.

[57] N. Ponomarenko et al. *Image database TID2013: peculiarities, results and perspectives*, Image Communication., Vol. 30, pp. 57-77, 2015.

[58] LIVE Image Quality Assessment Database, [Online]. Available: http://live.ece.utexas.edu/research/quality/subjective.htm.

[59] L. Callet, et al., "Subjective quality assessment IRCCyN/IVC database", 2005. [Online]. Available: http://www.irccyn.ec-nantes.fr/ivcdb.

[60] A. Saha and Q. M. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures", *Signal Proc.*, vol.128, pp.186-197, 2016.

[61] A. Barri, et al., "A locally adaptive system for the fusion of objective quality measures", *IEEE Trans. Image Proc.*, vol.23, no.6, pp.2446-2458, 2014.

[62] S. C. Pei and L. H. Chen, "Image quality assessment using human visual DOG model fused with random forest", *IEEE Trans. Image Proc.*, vol.24, no.11, pp.3282-3292, 2015.

[63] H. David and Wolpert, "The supervised learning no-free-lunch theorems", *Proc. 6th Online World Conf. on Soft Computing in Industrial Applications*, 2001.

[64] H. David, Wolpert and W. G. Macready, "No free lunch theorems for search", *Technical Report SFI-TR-95-02-010. Sante Fe, NM, USA: Santa Fe Institute*, 1995.

[65] H. David, Wolpert and W. G. Macready, "No free lunch theorems for optimization", *IEEE Trans. on Evolutionary Computation*, vol.1, no.1, pp.67-82, 1997.

[66] H. David, Wolpert and W. G. Macready, "Coevolutionary free lunches", *IEEE Trans. on Evolutionary Computation*, vol.9, no.6, pp.721-735, 2005.

[67] Z. Wang and A. C. Bovik, "Modern image quality assessment", *Synthesis Lectures on Image, Video Multimedia Proc.*, 2006.

[68] H. R. Sheikh, A. C. Bovik and L. Cormack, "No-reference quality assessment using nature scene statistics: JPEG2000", *IEEE Trans. Image Proc.*, vol.14, pp.1918-1927, 2005.

[69] L. Liang et al., "No-reference perceptual image quality metric using gradient profiles for JPEG2000", *Signal Proc. and Image Communication*, vol.25, pp.502-516, 2010.

[70] T. Brando and M. P. Queluz, "No-reference image quality assessment based on DCT domain statistics", *Signal Proc.*, vol.88, pp.822-833, 2008.

[71] Z. Wang, H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images", *IEEE Int. Conf. on Image Proc.*, 2002.

[72] R. Ferzli and L. J. Karam, "A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB)", *IEEE Trans. Image Proc.*, vol.18, pp.717-728, 2009.

[73] Z. Wang and A. C. Bovik, "Reduced-and no-reference image quality assessment: the natural scene statistic model approach", *IEEE Signal Proc. Magazine*, vol.28, pp.29-40, 2011.

[74] M.A. Saad, A.C. Bovik and C. Charrier, "A DCT statistics-based blind image quality index", *IEEE Signal Proc. Lett.*, vol.17, no.6, pp.583-586, 2010.

[75] L. Zhang et al., "A feature-enriched completely blind image quality evaluator", *IEEE Trans. Image Proc.*, vol.24, no.8, pp.2579-2591, 2015.

[76] J. Xu et al., "Blind image quality assessment based on high order statistics aggregation", *TIP*, vol.25, no.9, pp.4444-4457, 2016.

[77] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks", *NIPS*, vol.1, pp.1097-1105, 2012.

[78] L. Bengio et al., "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol.11, pp.2278-2324, 1998.

[79] X. Liu, et al., "RankIQA: learning from rankings for no-reference image quality assessment", *Computer Vision and Pattern Recognition., arXiv:1707.08347 [cs.CV]*, 2017.

[80] S. Bossey et al., "Deep neural networks for no-reference and full-reference image quality assessment", *IEEE Trans. Image Proc.*, vol.27, no.1, pp.206-219, 2018.

[81] Collobert et al., "Natural language processing (almost) from scratch", *Journal of Machine Learning Research*, pp.2493-2537, 2011.

[82] B. A. Wandell, "Foundations of Vision", *Sinauer Associates, Inc.*, 1995.

[83] D. Marr, "Vision", *MIT Press*, July 2010.

[84] D. Marr and E. Hildreth, "Theory of edge detection", *Proc. R. Soc. Lond. B*, vol.207, pp.187-217, 1980.

[85] M. C. Morrone and D. C. Burr, "Feature detection in human vision: a phase-dependent energy model", *Proc. R. Soc. Lond. B*, vol.235, pp.221-245, 1988.

[86] M. C. Morrone et al., "Mach bands are phase dependent", *Nature*, vol.324, pp.250-253, 1986.

[87] M. C. Morrone and R. A. Owens, "Feature detection from local energy", *Pattern Recognition. Lett.*, vol.6, pp.303-313, 1987.

[88] P. Kovesi, "Image features from phase congruency", *Videre: J. Comp. Vis. Res.*, vol.1, pp.1-26, 1999.

[89] L. Henriksson, A. Hyvarinen and S. Vanni, "Representation of cross-frequency spatial phase relationships in human visual cortex', *Journal of Neurosci.*, vol.29, pp.14342-14351, 2009.

[90] J. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images", *IEEE Trans. Inf. Theory*, vol.20, pp.525-536, 1974.

[91] S. Daly, "Subroutine for the generation of a human visual contrast sensitivity function", *Eastman Kodak Tech. Report*, 1987.

[92] P. Mohammadi, A. Ebrahimi-Moghadam and S. Shirani, "Subjective and objective quality assessment of image: a survey", *Majlesi Journal of Electrical Engineering*, vol.9, no.1, 2015.

[93] R. Jain, R. Kasturi and B. G. Schunck, "Machine vision", *McGraw-Hill*, 1995.

[94] B. Jahne, H. Haussecker and P. Geissler, "Handbook of computer vision and applications", *Academic Press*, 1999.

[95] C. C. Yang and S. H. Kwok, "Efficient gamut clipping for color image processing using LHS and YIQ", *Opt. Eng.*, vol.42, pp.701-711, 2003.

[96] M.A. Senthil, M. V. C. Rao and A. Chandramohan, "Competitive approaches to PSO algorithms via new acceleration co-efficient variant with mutation operators", *In: Proceedings of the 5th International Conf. on Computational Intelligence and Multimedia Applications*, 2005.

[97] S. M. Shan and G. S. Deng, "Data clustering using hybridization of clustering based on grid and density with PSO", *In: IEEE International Conf. on Service Operations and Logistics, and Informatics*, 2006.

[98] X. Hu, Y. Shi and R. C. Eberhart, "Recent Advances in Particle Swarm", *In Proceedings of Congress on evolutionary Computation (CEC), Portland, Oregon*, pp.90-97, 2004.

[99] J. A. Hartigan, "Clustering Algorithms", *John Wiley and Sons, Inc.*, 1975.

[100] L. Zhang, L. Zhang, X. Mou, "RFSIM: A feature based image quality assessment metric using Riesz transforms, *17th IEEE International Conf. on Image Processing ICIP*, Hong Kong, China, 2010, pp.321-324.

[101] K. Okarma, "Weighted Feature Similarity - a nonlinear combination of gradient and phase congruency for full-reference image quality assessment, *Image Processing and Communications Challenges, Advances in Intelligent Systems and Computing*, vol.184. Springer, 2013, pp.187-194.

[102] Z. Ni, H. Zeng and J. Chen, "ESIM: Edge Similarity for Screen Content Image Quality Assessment", *IEEE Transactions on Image Processing*, vol.26, no.10, pp.4818-4831, 2017.

LIST OF PUBLICATIONS

Journal Articles

1. Yadanar Khaing, Yosuke Sugiura and Tetsuya Shimamura, "Convolutional Neural Network for Blind Image Quality Assessment", *Journal of Signal Processing*, 9-pages, (Accepted).
2. Yadanar Khaing, Yosuke Sugiura and Tetsuya Shimamura, "Combination of Dissimilar Feature-Scores for Image Quality Assessment Using Particle Swarm Optimization Algorithm", *Journal of Signal Processing*, 9-pages, (Accepted).

International Conferences

1. Yadanar Khaing, Yosuke Sugiura and Tetsuya Shimamura, "Combination of Image Quality Scores Based on Particle Swarm Optimization", *TENCON 2018 - 2018 IEEE Region 10 Conference*, pp. 0072 - 0075, 4-pages, Jeju Island, South Korea.
2. Yadanar Khaing, Yosuke Sugiura and Tetsuya Shimamura, "Convolutional Neural Network for Blind Image Quality Assessment" *RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing 2019*, 4-pages, Hawaii, USA.