A ROBUST QUALITY METRIC FOR IMAGE SUPER RESOLUTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

YİĞİT KİPMAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

FEBRUARY 2015

Approval of the thesis:

A ROBUST QUALITY METRIC FOR IMAGE SUPER RESOLUTION

submitted by **YİĞİT KİPMAN** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Gönül Turhan Sayan
Head of Department, **Electrical and Electronics Eng.** _____

Prof. Dr. Gözde Bozdağı Akar
Supervisor, **Elec. and Electronics Eng. Dept., METU** _____

Examining Committee Members:

Prof. Dr. A. Aydın Alatan
Electrical and Electronics Engineering Dept., METU _____

Prof. Dr. Gözde Bozdağı Akar
Electrical and Electronics Engineering Dept., METU _____

Assist. Prof. Dr. Fatih Kamışlı
Electrical and Electronics Engineering Dept., METU _____

Assist. Prof. Dr. Sevinç Figen Öktem
Electrical and Electronics Engineering Dept., METU _____

Assoc. Prof. Dr. Alptekin Temizel
Informatics Institute, METU _____

Date: 06/02/2015

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: YİĞİT KİPMAN

Signature :

ABSTRACT

A ROBUST QUALITY METRIC FOR IMAGE SUPER RESOLUTION

Kipman, Yiğit
M.S., Department of Electrical and Electronics Engineering
Supervisor : Prof. Dr. Gözde Bozdağı Akar

February 2015, 70 pages

Superresolution have become an active topic in image processing in the last decade. Various superresolution algorithms have been developed; however these superresolution algorithms may introduce defects such as blurring, aliasing, added noise and ringing. Evaluating the performance of these superresolution algorithms is an important problem; because the original high resolution image is not available while quantifying the quality of superresolution image. Subjective tests can be made to quantify the perceived image quality; but they are time-consuming and expensive. Only a few objective quality assessment algorithms are proposed that evaluate the quality of superresoluted image from its low-resolution (LR) pair; but these do not correlate well with the subjective tests. In this thesis, a quality assessment algorithm for image superresolution that follows the philosophy of natural scene statistics (NSS) is analyzed and an improvement is proposed. A statistical model of frequency energy falloff characteristics of high resolution (HR) images is developed and a quality measure is calculated from the departures from HR image statistics. A no-reference spatial image quality assesment measure that also follows the philosophy of NSS is incorporated in the proposed algorithm to improve the robustness of the metric against noise. It is shown that the proposed approach is robust against noise and correlates well with the human visual system.

Keywords: Image Quality Assessment, Image Superresolution, Natural Scene Statistics, No-reference Quality Metric

ÖZ

RESİM SÜPERÇÖZÜNÜRLÜĞÜ İÇİN GÜRBÜZ BİR KALİTE METRİĞİ

Kipman, Yiğit
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Gözde Bozdağı Akar

Şubat 2015 , 70 sayfa

Süperçözünürlük konusu görüntü işleme alanında son yıllarda çok aktif bir konu haline gelmiştir. Birçok süperçözünürlük algoritması geliştirilmiştir; ancak bu süperçözünürlük algoritmaları görüntüye bulanıklaşma, örtüşme, eklenmiş gürültü ve salınım gibi çeşitli bozulmalar getirebilmektedir. Bu algoritmaların performansının değerlendirilmesi önemli bir problemdir; çünkü süperçözünürlük ile oluşturulan yüksek çözünürlüklü görüntüyü karşılaştıracak orjinal yüksek çözünürlüklü görüntü elde yoktur. Algılanan görüntü kalitesinin ölçümü için öznel testler yapılabilir; ancak bu testler zaman tüketici ve maliyetlidir. Süperçözünürlük kalitesini hesaplayan yalnızca birkaç nesnel kalite değerlendirme algoritması önerilmiştir; ancak bu algoritmalar öznel testler ile arasında uygunluk sağlamamaktadır. Bu tezde, görüntü süperçözünürlüğü için doğal görüntü istatistikleri felsefesini izleyen bir kalite değerlendirme algoritması analiz edilmiştir ve algoritmaya bir iyileştirme önerilmiştir. Yüksek çözünürlüklü görüntülerin frekans enerji düşüş karakteristiklerine dayanan istatistiksel bir model inşa edilmiş, modelden sapmalar üzerinden kalite hesaplanmıştır. Doğal görüntü istatistikleri felsefesini izleyen referanssız, uzamsal bir görüntü kalite değerlendirme ölçütü önerilen algoritmaya eklenerek algoritma gürültüye karşı gürbüz hale getirilmiştir. Sonuç olarak önerilen algoritmanın gürültüye karşı gürbüz olduğu ve insan görsel sistemi ile uygunluk sağladığı gösterilmiştir.

Anahtar Kelimeler: Görüntü Kalite Ölçümü, Görüntü Süper Çözünürlüğü, Doğal Görüntü İstatistikleri, Referanssız Kalite Metriği

to my beloved family

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Prof. Gözde Bozdağı Akar for her endless support and guidance throughout this period.

I am very grateful to Aselsan Inc. for providing me the opportunity to pursue my research. I would also like to thank all my colleagues for their motivation on me.

Finally, I would like to thank my family and my dearest love for their endless encouragement and support through my whole life.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

NSS	Natural Scene Statistics
NR	No-Reference
IQA	Image Quality Assessment
QoE	Quality of Experience
LR	Low Resolution
HR	High Resolution
SR	Super-resolution
GGD	Generalized Gaussian Distribution
SSIM	Structural Similarity
PSNR	Peak Signal to Noise Ratio
EDAT	Edge Adaptive Interpolation Using Total Variation Decomposition

CHAPTER 1

INTRODUCTION

1.1 Introduction

In image acquisition, a real image is projected on sensor arrays and the image is discretely captured with limited number of pixels. Image super-resolution (SR) algorithms increase the spatial resolution of the images beyond the imaging device resolution and play an important role in many modern applications such as medical imaging, media streaming and video surveillance applications. As the number of SR algorithms dramatically increased in the last decade, evaluating the performance of the SR algorithms have become an active research area.

There exists straightforward approaches to the problem such as subjective evaluation, where subjects are asked to evaluate multiple resolution enhanced images by different SR algorithms and score each image within a finite scale. The mean opinion scores (MOS) of the subjects are evaluated which gives out the performance metrics of the tested algorithms. Subjective tests provide a reliable ground truth when comparing the performances of different SR algorithms; because human eyes are the last receivers and ultimate evaluators of images. Although subjective tests provide reliable results to the problem, conducting subjective experiments is not practical as it is extremely time consuming and expensive. Evaluating the performance of SR algorithms objectively remains an open problem and there is a strong need for an objective quality assessment metric which correlates well with the human perception.

Objective image quality assessment (IQA) algorithms have been developed such

as peak-signal-to-noise-ratio (PSNR) and structural-similarity-index (SSIM) [2]. However, these algorithms are not applicable to the problem of IQA for image SR; because they require the original image to be available as a reference to compare the SR image with. There also exist some reduced reference algorithms which require some feature information about the original HR image to be available. Again, since the original image is not available in the SR IQA problem, the features can not be computed for original image and these algorithms are not applicable. SR IQA can only be estimated by a no-reference (NR) IQA algorithm. There are some NR IQA algorithms available that quantify specific distortion types on images. These algorithms may quantify specific distortions that are added to the HR image created by the SR methods but they fall short on quantifying the high frequency detail reconstruction properties of the created HR images.

1.2 Scope of the Thesis

In this thesis, a SR IQA algorithm [1] which is one of the initial attempts to develop an objective image SR IQA metric is analyzed in detail. The algorithm is tested for different image sets with different spatial details and degradation types. The aim is to find out the shortcomings of the algorithm and come up with a new method to improve the performance of the analyzed algorithm.

We have searched the literature for all possible IQA algorithms that are designed to evaluate the performance of image SR. In addition, other no reference (NR) IQA algorithms are investigated for possible applications in superresolution image quality assessment problem.

We have implemented and tested the chosen SR IQA and NR IQA algorithms for detailed evaluation. Based on detailed theoretical and experimental analysis on these algorithms, we proposed a multi domain IQA algorithm for image SR. The proposed algorithm is tested for different SR algorithms that are available in the literature.

1.3 Outline of the Thesis

In Chapter 2, literature review about available IQA algorithms is given. Natural scene statistics approach to the IQA problem is clarified.

In Chapter 3, a natural scene statistics (NSS) based SR IQA algorithm which uses statistical models both in frequency and spatial domains is explained in detail. In addition, a no reference spatial domain IQA method which may be applicable to the superresolution IQA problem is explained in detail.

In Chapter 4, some basic and some advanced image SR algorithms are analyzed, which are used in SR IQA performance evaluation tests.

In Chapter 5, possible weaknesses and shortcomings of analyzed SR IQA algorithms are described. The proposed NSS based SR IQA algorithm is explained and formulated in detail.

In Chapter 6, the proposed algorithm is compared against other available SR IQA algorithms. Subjective evaluations are compared with the algorithm results in order to create a ground truth for the comparison.

In Chapter 7, conclusions of this study are drawn and any future works are discussed.

CHAPTER 2

LITERATURE RESEARCH

2.1 Image Observation Model

Image observation model relates the observed low resolution image to the original high resolution image. Image acquisition systems can not capture the images perfectly due to the limitations of the discrete elements used inside. Non perfect sensor elements introduce sensor blur, non-zero aperture time introduces motion blur and finite aperture size introduces optical blur to the image acquisition system. Sensor, motion and optical blur can be modelled by a function called point spread function (PSF). In addition, the resolution of the acquired image is limited by the sensor array. A diagram of image observation model is given in Figure 2.1.

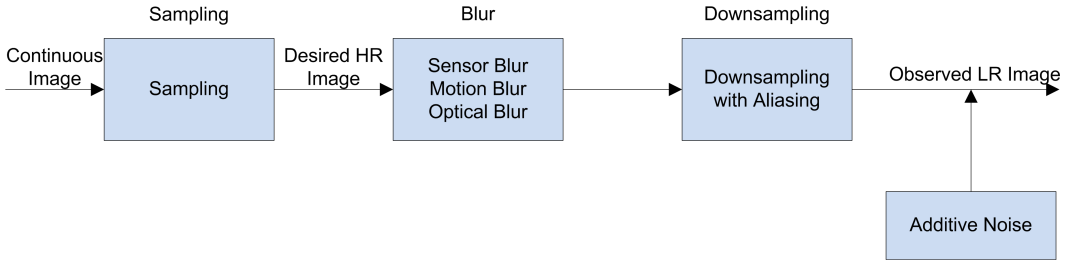


Figure 2.1: Image Observation Model

The image observation model can then be modelled by

$$\boldsymbol{L} = DB\boldsymbol{H} + N \tag{2.1}$$

where H is the high resolution image, L is the low resolution image, D is the downsampling operator, B is the point spread function that models the blurring

effect and N is the additive noise. The equation can also be written in a single matrix form, $\boldsymbol{L} = M\boldsymbol{H} + N$ where M is the degradation matrix.

2.2 Superresolution Image Enhancement

Superresolution image enhancement refers to the task of recovering the original high resolution image from one or several of its low resolution pairs. Most super-resolution algorithms assume that the low resolution input image was obtained by down-sampling the high-resolution image with a blur kernel which is usually assumed to be the Point Spread Function (PSF) of the image acquisition device. In many cases PSF is not known and the blur kernel is assumed to be some standard low-pass filter (LPF) like a Gaussian or a bicubic kernel.

There are mainly four main approaches to the image superresolution problem. These are spatial domain approaches, transform domain approaches, statistical approaches and example based approaches.

Spatial domain approaches only rely on the available pixel information apparent in a low resolution image. The high resolution pixels are interpolated only by the available pixels of low resolution image.

Transform domain approaches are used to simplify the deconvolution and 2D filtering operations used in superresolution as these operations become simple multiplication operations in the transform domain.

Statistical approaches rely on stochastic models to optimally reconstruct the high resolution image from its low resolution pair.

Example based approaches use dictionaries that are built either from one image or a set of many images. Existing information in the dictionaries are extracted and used for estimating the high resolution pixels.

2.3 Image Quality Assessment Approaches

Image quality assessment approaches can be categorized into two main groups, namely subjective approaches and objective approaches.

Subjective methods are focused on psychological experiments in which human observers take part. Different techniques are used in subjective experiments such as single stimulus, double stimulus or pair wise comparison among others. A standard for measuring subjective image quality is also described in ITU-R BT.500-11 [3]. However, the involvement of real people within subjective image quality assessment methods require all the factors that influence human perception to be taken into account during the tests and strict protocols have to be adopted. The results of the subjective methods appeared to correlate well with the human visual system; however the process of conducting these subjective tests are extremely time consuming and expensive. Therefore the efficiency of the subjective methods are low compared to the objective methods.

In objective methods, different metrics are computed directly from the digitally available image. Objective image quality assessment methods can be classified according to the availability of the original high resolution image, namely Full Reference (FR) Metrics, Reduced Reference (RR) Metrics and No Reference (NR) Metrics. Within the scope of this thesis, only the objective image quality assessment will be studied.

In the following sections, a summary of well known FR, RR and NR metrics is given.

2.3.1 Full Reference Metrics

In full reference methods, direct comparison in between the original image and the image under test is performed. There is a limitation to the applicability of FR methods as the original image is required to perform the computation. The simplest FR metrics that are widely used are Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR). These metrics are widely used in ap-

plications; however the results of these algorithms do not correlate well with the subjective results and human visual system.

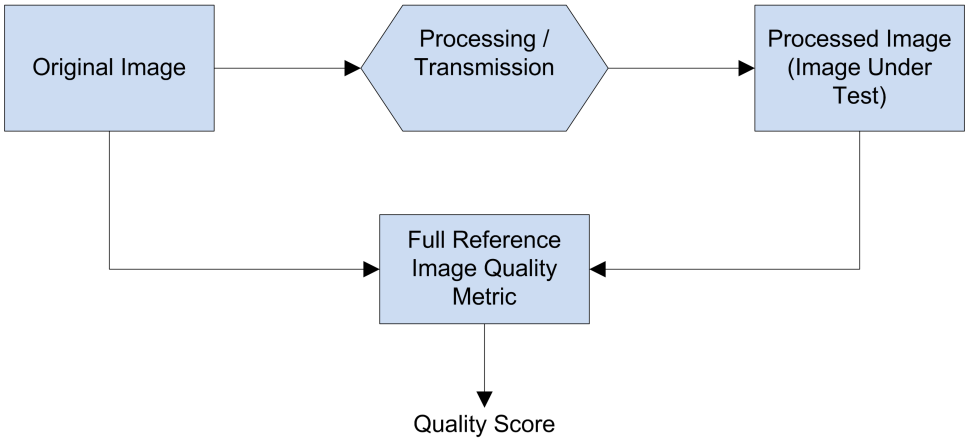


Figure 2.2: Full Reference Image Quality Assessment Model

A measurement of structural similarity, the Structural Similarity Measure (SSIM) [2] is based on the assumption that the natural images are highly structured and a measure of distortion or dissimilarity provides a good approximation to the perceived image quality.

Visual Information Fidelity Index proposed by Sheikh and Bovik [4] is another FR approach that uses the natural scene statistics and quantifies loss of information due to distortions present in the image.

While there have been other well-established FR methods [5], [6], [7] which correlate well with human perception of quality, there is still considerable room for the improvement of objective no-reference image quality assessment (NR-IQA) methods.

2.3.2 Reduced Reference Metrics

In reduced reference metrics, only partial information about the original image is available while quantifying the quality of the image under test. Therefore RR metrics lie in between the no reference and full reference metrics in terms of available information about the original image. RR methods just extract some features from both the original and the processed image and quantifies the qual-

ity of the image corresponding to these features, which are the representatives of all the information in the images. Extracted features mostly describe the image content or distortion based properties. Compared to FR and NR metrics, only a few RR metrics [8], [9], [10] are available in the literature which correlate well with human perception of quality.

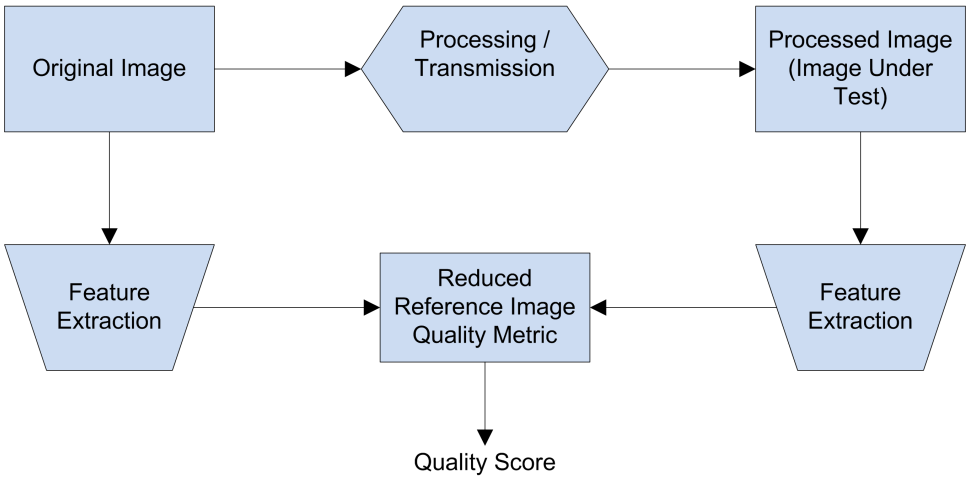


Figure 2.3: Reduced Reference Image Quality Assessment Model

2.3.3 No Reference Metrics

No reference metrics are also known as blind metrics as there is no way of direct comparison in between the original image and the image under test; because the original image is not available. In image processing applications such as super-resolution, the original image is not available for direct comparison. Therefore no-reference metrics are highly applicable in the subject of superresolution image quality assessment. Most of the NR IQA algorithms do not exploit natural image modeling but they assume that distortion type affecting the image quality is known. Some of them estimate image blur [11], [12] or JPEG/JPEG2000 compression artifacts by investigating the features of the artifact in spatial and frequency domain [13], [14], [15], [16], [17]. These NR metrics are designed to distinguish specific image degradation types and quantify their presence from specific properties of the characterized artifacts. Therefore most NR methods can be classified as distortion-aware since they can handle only one or few specific degradation types.

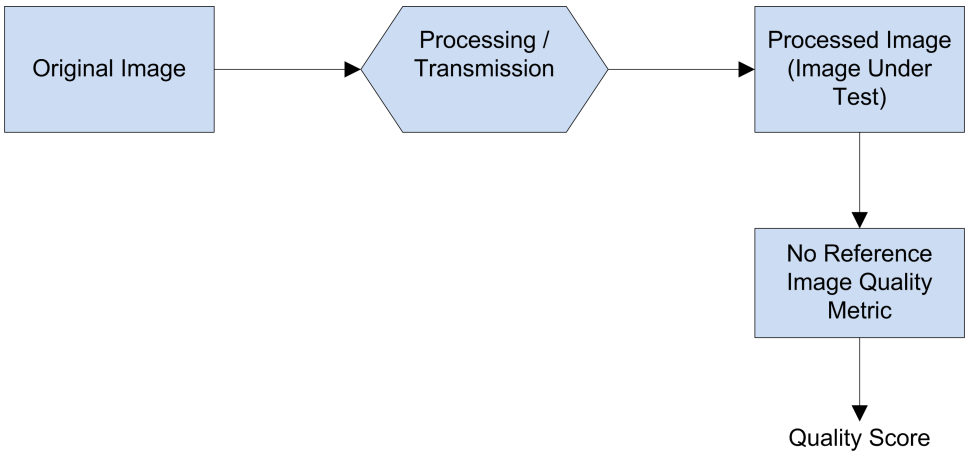


Figure 2.4: No Reference Image Quality Assessment Model

Another set of NR metrics are developed by exploiting the relationships of natural scene statistics (NSS) with the human perceived image quality [18], [19], [20], [21], [22], [23], [24]. Natural scene statistical models seek ways to capture the statistical characteristics of natural images which hold across different image contents. These NSS based NR IQA methods which can be applied to the SR IQA problem are explained in detail in the next chapter.

2.3.3.1 Natural Scene Statistics Approach

Everyone can easily distinguish images from the natural world from synthetic or man-made images. Natural images are easily distinguishable as they contain some typical types of structural properties.

The images that we consider natural comprise just a tiny subset of all possible images. Even though images are expressed as elements of a large vector space, namely the image space, the subset of natural images that we are interested in are rather restricted. So we can create an image manifold which is an isolated subset of images and learn a probability model such as a Gaussian like model on it.

It has long been discovered that natural scenes possess scale invariance [25]. The marginal distributions of natural image statistics remain unchanged even if the images are scaled. There is a rich literature focusing on NSS models [25], [26],

[27] but only a small proportion of this work was exploited in RR and NR image quality assessment algorithms.

NSS based approaches to NR IQA problem assume that natural scenes possess certain statistical characteristics, and presence of any distortion affects these characteristics. Various properties of statistical characteristics of natural images are studied, such as intensity, color, spatial correlation and higher order statistics. Current NR IQA algorithms which exploit NSS based features are [18], [19], [20], [21], [22], [23]. Some NSS based NR methods are computed in the transform (frequency) domain to make use the common statistical characteristics of the power spectra of natural images [24].

Basically, NSS based NR IQA algorithms exploit the relationships of natural scene statistics with the human perceived image quality. These NSS features which are related to human perceived image quality also differ for SR images which are created by different SR algorithms as they have different reconstruction and distortion characteristics on images. Therefore the performance of various SR algorithms can be evaluated objectively by the departures of computed NSS features from the constructed NSS model which relate well with human perceived image quality.

In the next chapter, some NSS based NR IQA methods which can be applied to the SR IQA problem are explained in detail. Based on the explained methods, we will develop a statistical model to characterize the degradations in image SR process and quantify the performance of different SR algorithms. Details of the constructed NSS model for SR IQA problem will be described later in Chapter 5.

CHAPTER 3

NO REFERENCE SUPERRESOLUTION QUALITY ASSESSMENT ALGORITHMS

In this Chapter, two NSS based no reference IQA algorithms which are applicable to the problem of superresolution IQA are explained in detail. We have focused our study on these two algorithms; because the first algorithm exploits frequency falloff characteristics of natural images which is a good representative of the high frequency reconstruction properties of SR methods, whereas the second algorithm computes spatial NSS features which are good representatives of the artifacts which may be present in the SR image, no matter which specific type of distortion it is affected by.

3.1 Quality Assessment Metric for Image Super Resolution by Natural Scene Statistics [1]

This algorithm uses the natural scene statistics (NSS) approach to characterize the image degradations of a SR image by only using the available LR image. The computation is made in both the frequency domain and spatial domain.

3.1.1 Frequency Energy Falloff Statistics

Amplitude spectrum of natural images falls with the spatial frequency proportional to $1/f^p$, where f is the spatial frequency and p is is an exponent that varies over a small range across natural images [1]. A statistical model in fre-

quency domain is built from this fact. The model is built by computing the total energy of LR and HR images in different dyadic scales computed by steerable pyramid transform. Steerable pyramid transform is used because it provides a tight frame and preserves the spatial domain energy in the frequency domain. The calculated frequency energy falloff curves of both LR and HR images appear to be parallel in log-log scale which can be used to predict the frequency falloff curves beyond the finest scale for the LR image. The error between the prediction and the frequency energy falloff curve of SR image in the finest scale is used to quantify the SR quality score in the frequency domain. Therefore the

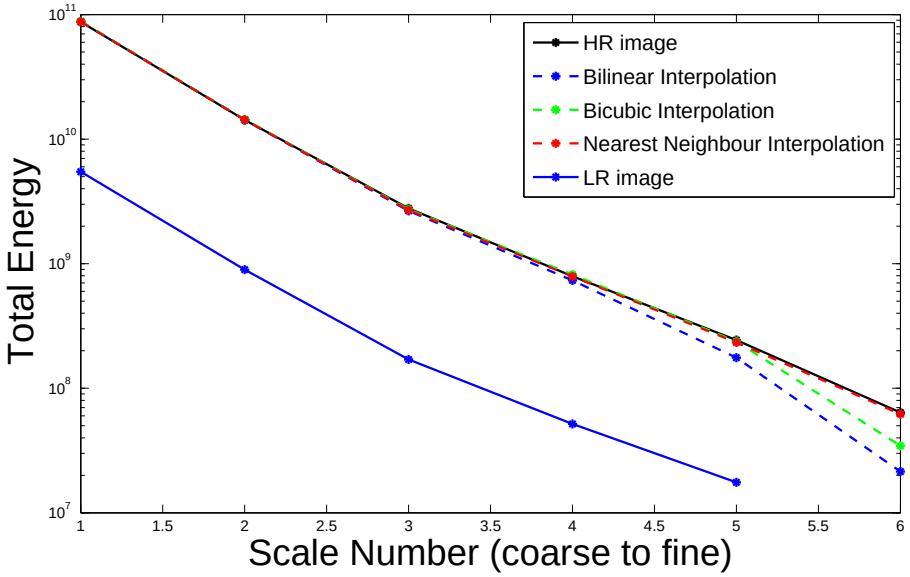


Figure 3.1: Frequency energy falloffs of original HR, SR and LR images

framework of this algorithm begins by computing the frequency energy falloff curves of LR and SR images by steerable pyramid transform. The transform coefficients are then squared and summed to compute the total energy in each scale. The computation is made for 6 different scales for SR image and 5 different scales for LR image. Then the slopes of frequency falloff curves of HR image

is predicted as

$$\begin{aligned}
\hat{s}_1^H &= s_1^L \\
\hat{s}_2^H &= s_2^L \\
\hat{s}_3^H &= a_0 + a_1 s_3^L \\
\hat{s}_4^H &= b_0 + b_1 s_4^L \\
\hat{s}_5^H &= c_0 + c_1 \hat{s}_3^H + c_2 \hat{s}_4^H
\end{aligned} \tag{3.1}$$

where s_i^L is the falloff curve slope of LR image in i-th scale, $\hat{s}_i^H$ is the predicted falloff curve slope of HR image in i-th scale and $a_0, a_1, b_0, b_1, c_0, c_1$ and c_2 are obtained by simple least square regression using the statistical model built by real LR and HR images. The predicted coefficients from the regression are $a_0 = 0.07, a_1 = 1.00, b_0 = 0.89, b_1 = 1.06, c_0 = -3.38, c_1 = -0.10$ and $c_2 = 0.89$.

The normalized error in frequency energy falloff slopes in the finest scale in between the predicted HR image (from LR) and the SR image is computed as

$$e_f = \frac{\hat{s}_5^H - s_5^H}{\hat{s}_5^H} \tag{3.2}$$

Here e_f is close to zero when the SR image is similar to the original HR image. Then e_f is inserted to the generalized Gaussian density function which is fitted by testing with 1400 natural HR images

$$p_{e_f}(e_f) = \frac{1}{Z_f} \exp\left[-\left(\frac{|e_f - \mu_f|}{\alpha_f} \right)^{\beta_f} \right] \tag{3.3}$$

where $Z_f = \dfrac{\beta_f}{2\alpha_f \Gamma(1/\beta_f)}$ is a normalization factor. The maximum likelihood based fitting result used in the algorithm is $\mu_f = 0.029, \alpha_f = 0.0608$ and $\beta_f = 0.6124$.

3.1.2 Spatial Continuity Based Statistics

The spatial domain statistics are calculated with a continuity based statistic model developed to measure the naturalness of the SR image.

In the algorithm, a straightforward method is developed to compute the signal continuity. Let $f(i)$ for $i = 0, ..., N-1$ be one row or column of pixels from the

SR image.

$$g(i) = |f(i+1) - f(i)| \tag{3.4}$$

for $0 <= i <= N-2$ where N is the number of pixels in a row.

In the case of superresolution by a factor of 2, even and odd samples in $f(i)$ will have different levels of continuities. Such discontinuities should not be observed in $g(i)$ computed from natural HR images. Therefore the spatial continuity measure is quantified as

$$e_s = \frac{1}{M} \sum_{i=0}^{M-1} [g(2i) - g(2i+1)] \tag{3.5}$$

where $M = N/2$. Spatial continuity measure e_s is computed for every row and column in the SR image and averaged over all rows and columns. As in the frequency falloff statistics model, computed e_s is inserted to the generalized Gaussian density function which is fitted by computing the spatial continuities with 1400 natural HR images

$$p_{e_s}(e_s) = \frac{1}{Z_s} \exp\left[-\left(\frac{|e_s - \mu_s|}{\alpha_s} \right)^{\beta_s} \right] \tag{3.6}$$

where $Z_s = \dfrac{\beta_s}{2\alpha_s \Gamma(1/\beta_s)}$ is a normalization factor. The maximum likelihood based fitting result used in the algorithm is $\mu_s = 0.007, \alpha_s = 0.0751$ and $\beta_s = 0.8679$.

3.1.3 Multidomain Quality Assessment Metric

In the algorithm, useful measures of image naturalness is constructed by probability model p_{e_f} in the frequency domain and p_{e_s} in the spatial domain. A normalized joint probability model for image naturalness measure is defined as

$$p_n = \frac{1}{K} p_{e_f} p_{e_s} \tag{3.7}$$

where $K = \dfrac{Z_f}{Z_s}$ is the normalization factor so that p_n is upper-bounded by 1. To convert this probability based metrics to a distortion metric

$$D_n = \left(\frac{|e_f - \mu_f|}{\alpha_f} \right)^{\beta_f} + \left(\frac{|e_s - \mu_s|}{\alpha_s} \right)^{\beta_s} \tag{3.8}$$

Finally we arrive at D_n, which quantifies the departures of statistics of SR images from natural HR image statistics computed by frequency energy falloffs in the frequency domain and spatial continuities in the spatial domain.

3.2 Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE)

The Blind/Referenceless Image Spatial QUality Evaluator algorithm [23] which is also known as BRISQUE utilizes a natural scene statistics model framework from locally normalized luminance coefficients and quantifies the naturalness of the images by the departures from this model. The model developed in BRISQUE is in spatial domain and it computes the quality based on the statistics of pairwise products of locally normalized neighbouring pixel luminance values.

3.2.1 Spatial Domain MSCN Statistics

Ruderman observed that applying a local non-linear operation to log contrast luminances to remove local mean displacements from zero log-contrast and to normalize the local variance of the log contrast has a decorrelating effect. Ruderman also observed that these normalized luminance values strongly tend towards a unit normal Gaussian characteristic for natural images. [26] Therefore the algorithm begins with an operation to compute mean subtracted contrast normalized (MSCN) coefficients

$$\hat{I(i,j)} = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1} \tag{3.9}$$

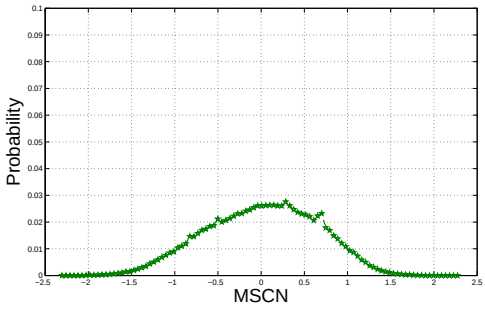
where $i = 1, ..., M$ $j = 1, ..., N$ are indices of image, M, N are image row and column size and

$$\mu(i,j) = \sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} I_{k,l}(i,j)$$

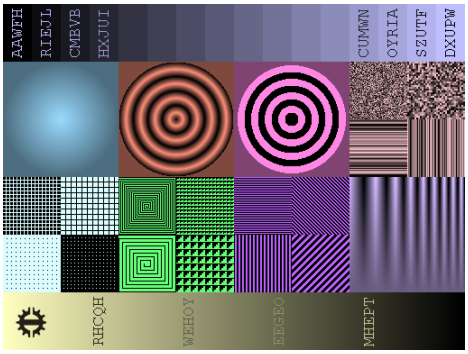
$$\sigma(i,j) = \sqrt{\sum_{k=-K}^{K} \sum_{l=-L}^{L} w_{k,l} \left(I_{k,l}(i,j) - \mu(i,j)\right)^2}$$

where $w = w_{k,l} | k = -K, ..., K, l = -L, ..., L$ is a 2D circularly symmetric Gaussian weighting function. The algorithm uses $K = L = 3$ as the window size.

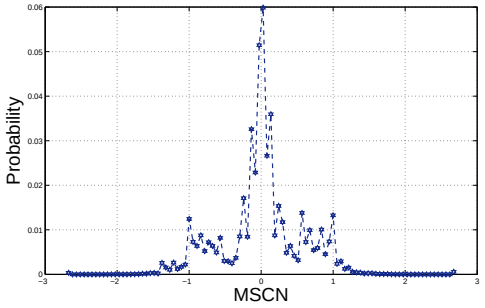
(a)



(b)



(c)



(d)

Figure 3.2: Gaussianity of (a-b) natural images and (c-d) synthetic images.

In Figure 3.2, histogram of MSCN coefficients for a natural image and a man-made (artificial) image is given. The histogram clearly shows that MSCN coefficients have characteristic statistical properties that vary for natural images. As the histogram of natural image MSCN coefficients tend to follow a Gaussian like characteristic, histogram of artificial image MSCN coefficients divert from the Gaussian shape.

MSCN coefficients computed in the algorithm have characteristic statistical properties that are also changed by the presence of distortions. By quantifying these changes, the algorithm can predict the perceptual quality scores of the images. Figure 3.3 visualizes the computed MSCN coefficient distributions in a histogram for a natural HR image and distorted versions of it. It is easily noticed that the natural scene image exhibits a Gaussian like appearance where the GGD with zero mean is given as

$$f(x;\alpha,\sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^{\alpha}\right) \tag{3.10}$$

where $\beta = \sigma\sqrt{\frac{\Gamma(a/\alpha)}{\Gamma(3/\alpha)}}$ and $\Gamma(.)$ is the gamma function.

Therefore by estimating the Gaussian shape parameters (α, σ^2) of natural images and SR images, we quantify the departures of SR image parameters from the natural image parameters which quantify the perceptual quality of image.

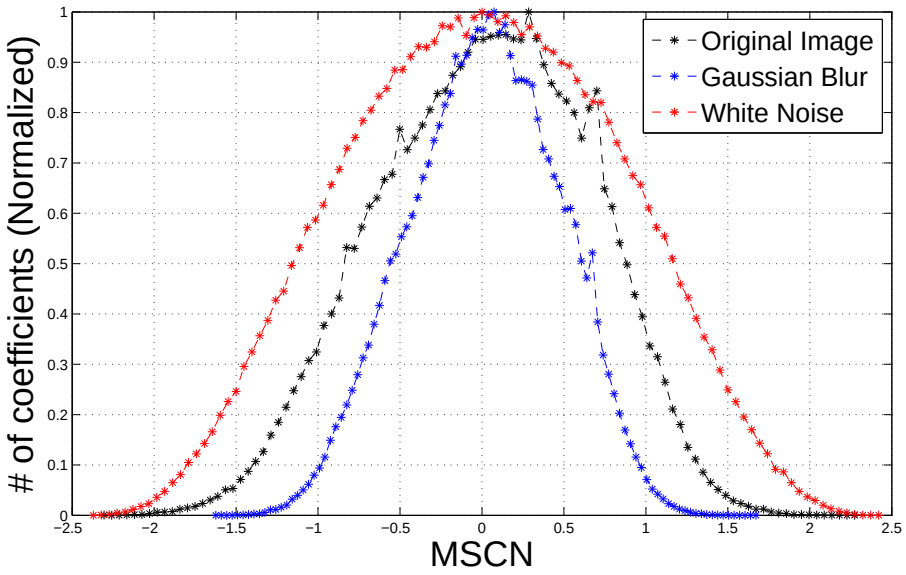


Figure 3.3: Histogram of MSCN coefficients for a natural undistorted image and its various distorted versions.

3.2.2 Neighbouring Pixel Statistics

In the algorithm, statistical relationships in between neighbouring pixels are computed for 4 orientations namely horizontal (H), vertical (V), main-diagonal $(D1)$ and secondary diagonal $(D2)$.

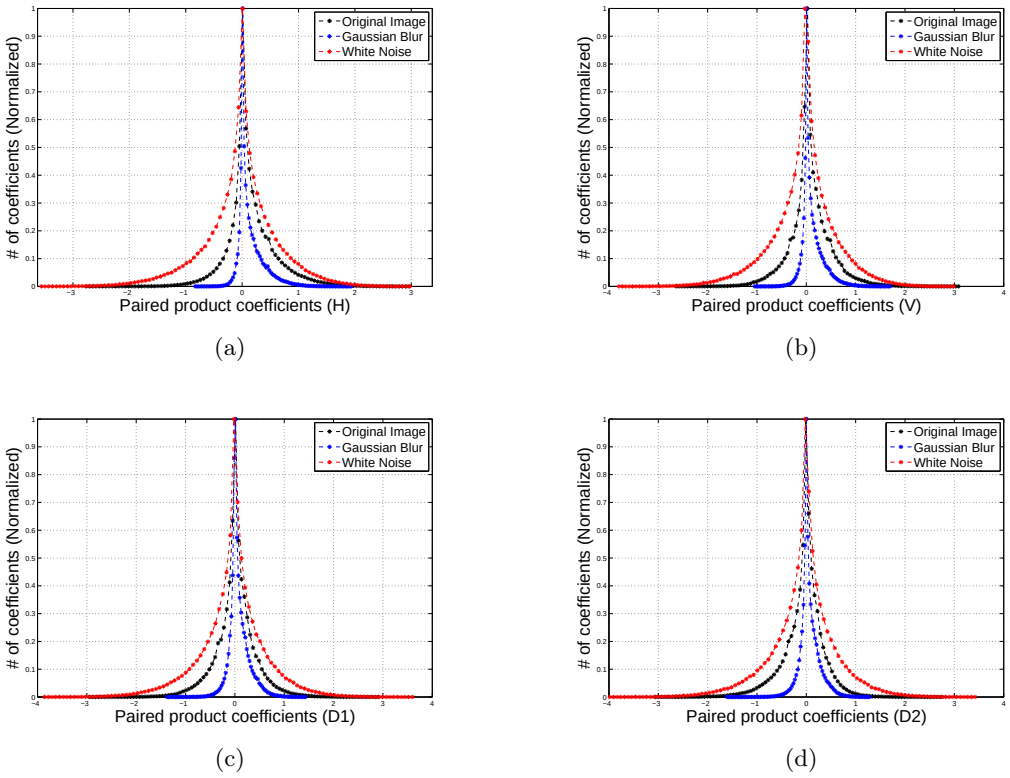
The following general asymmetric generalized Gaussian distribution (AGGD) model has been adopted for the histograms of paired products along each of four orientations

$$f(x; v, \sigma_l^2, \sigma_r^2) = \frac{v}{(\beta_l + \beta_r)\Gamma(1/v)} \exp\left(-\left(\frac{|x|}{\beta_l}\right)^v\right) \tag{3.11}$$

where

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}}, \quad \beta_r = \sigma_r \sqrt{\frac{\Gamma(1/v)}{\Gamma(3/v)}} \quad and \quad \zeta = (\beta_r - \beta_l)\frac{\Gamma(2/v)}{\Gamma(1/v)}.$$

Therefore by estimating the asymmetric Gaussian shape parameters $(\zeta, v, \sigma_l^2, \sigma_r^2)$ of natural images and SR images, we quantify the deviation of SR parameters from the natural image parameters which quantifies the perceptual quality of image.



(a) (b) (c) (d)

Figure 3.4: Histograms of paired products of MSCN coefficients of a natural undistorted image and various distorted versions of it. (a) Horizontal. (b) Vertical. (c) Main Diagonal. (d) Secondary Diagonal.

Extracting 2 features, (α, σ^2) from MSCN coefficients and 16 features, $(\zeta, v, \sigma_l^2, \sigma_r^2)$ from neighbouring pair products for each four orientations, 18 features are extracted. The feature extraction is made in 2 scales to better identify the distortions in multiple scales. From 36 features, a mapping to quality scores is learned by using a support vector machine (SVM) framework.

CHAPTER 4

EVALUATED SUPERRESOLUTION ALGORITHMS

In this Chapter, SR algorithms which are chosen to test the proposed SR IQA algorithm is explained in detail. The algorithms are explained from the simplest one to the most advanced and best performing one. We first consider nearest neighbour interpolation, bilinear interpolation and bicubic interpolation methods; because they are simple techniques for image SR and will provide a reference frame for SR performance comparison. We also consider EDAT [28] and example based SR by Kim [29] as they perform well in reconstructing high frequency details in the structure component of the image.

4.1 Nearest Neighbour Interpolation

Nearest neighbour (NN) interpolation is one of the simplest known techniques to interpolate an image. In this technique, the unknown pixel is replaced by a pixel which is in the closest neighbourhood. The output image from nearest neighbour interpolation preserves the sharp edges in the image; however it involves undesired jaggedness and increases aliasing in the image, where the diagonal lines and curves appear pixelated.

4.2 Bilinear Interpolation

Bilinear interpolation is a simple interpolation technique which can be applied in the image SR problem. Interpolation considers nearest 2x2 neighbour pixel

values surrounding the unknown pixel. After computing the weighting factors for these 4 coefficients, a weighted average is computed and a final interpolated pixel value is acquired. This method blurs the sharp edges, therefore there is loss of high frequency details after SR by bilinear interpolation.

On a unit square, consider the four data points where f is known are (0,0), (0,1), (1,0) and (1,1). Therefore in matrix operations, the bilinear interpolation is formulated as

$$f(x,y) \approx \begin{bmatrix} 1-x & x \end{bmatrix} \begin{bmatrix} f(0,0) & f(0,1) \\ f(1,0) & f(1,1) \end{bmatrix} \begin{bmatrix} 1-y \\ y \end{bmatrix} \tag{4.1}$$

4.3 Bicubic Interpolation

Being a simple interpolation technique as bilinear interpolation, bicubic interpolation goes one step further by considering nearest 4x4 neighbour pixel values surrounding the unknown pixel. Weighting is done by the distances of known pixels from the unknown pixel. Closer pixels are given higher weighting and as a result, the interpolated surface is a smoother surface than the one from bilinear interpolation. The interpolated surface can be formulated by

$$p(x,y) = \sum_{i=0}^{3} \sum_{j=0}^{3} a_{ij} x^i y^j \tag{4.2}$$

where the 16 weighting values, a_{ij}, are determined by function values, directional derivatives and cross derivatives of known pixels.

This method produces sharper images than bilinear interpolation, and has a good combination of processing time and SR image quality. Therefore, bicubic interpolation is also chosen to be used while testing the SR IQA algorithm proposed in this study.

4.4 EDAT

Edge Adaptive Interpolation Using Total Variation Decomposition (EDAT) [28] is a framework that uses total variation decomposition to separate the structured component of the image from textured parts. After decomposition, the textured component is interpolated with simple bicubic interpolation whereas the structure component is substituted to edge adaptive linear interpolation and filtered with a special filter that enhances the outline of the image, "the shock filter". EDAT is analyzed in detail as it outperforms most of the other available superresolution algorithms in subjective tests. There are no studies available that quantify the quality of EDAT algorithm in an objective manner; therefore the proposed algorithm in Chapter 5 will be tested with EDAT.

EDAT framework, which is created by fusing Total Variation Decomposition [30] and Edge Adaptive Interpolation [31] is given in Figure 4.1.

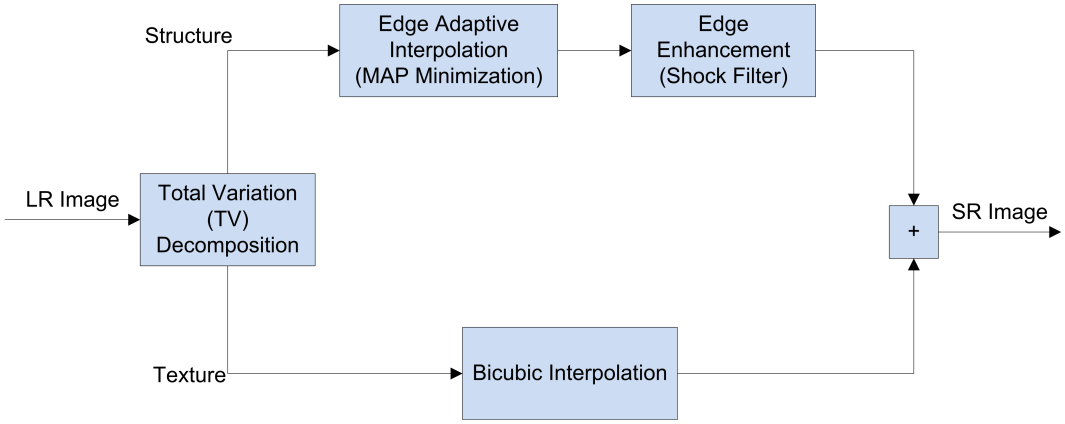


Figure 4.1: EDAT Framework

4.4.1 Total Variation Decomposition

Total Variation (TV) decomposition [30] decomposes the input image into two components, namely high frequency (texture) and low frequency (structure) bands. TV decomposition is carried out as given in Eq 4.3

$$p_{i,j}^{(n+1)} = \frac{p_{i,j}^{(n)} + \tau \left\{ \nabla (divp^n - Y/\lambda) \right\}_{i,j}}{1 + \tau |\nabla (divp^n - Y/\lambda)_{i,j}|} \tag{4.3}$$

where p denotes a dual vector which is used within the iterations, Y denotes the input image, div is the divergence operation and τ is a coefficient. After p converges, the texture component v and the structure component u is found as in Eq. 4.4 and Eq. 4.5.

$$v = \lambda divp \tag{4.4}$$

$$u = Y - v \tag{4.5}$$

4.4.2 Edge Adaptive Interpolation Using Bilateral Filter

Edge adaptive interpolation used in EDAT preserves the edge orientations. The method estimates the edge directions with bilateral filter coefficients. The method used in EDAT is similar to the method used by Siu et al in [31]. Range distance values for bilateral filter is computed by Maximum a Posteriori estimation.

4.4.3 Shock Filter

Shock filter is a non-linear filter that enhances the edges of an image. A drawback of shock filter is that it degrades the SNR of images. Therefore in EDAT, edge enhancement operation by shock filter is only applied to the low frequency (structure) component of the image to minimize its degradations on the total image.

Shock filter is described in [32] by the following formulas:

$$Y^{(n+1)} = Y^{(n)} - sign\left(\triangle Y^{(n)}\right) ||\nabla Y^{(n)}|| dt \tag{4.6}$$

$$\triangle Y = \partial_x^+ \partial_x^- Y . \left(\partial_x Y\right)^2 + \partial_y^+ \partial_y^- Y . \left(\partial_y Y\right)^2 + \left(\partial_x^- \partial_y^- Y + \partial_x^+ \partial_y^+ Y\right) . \partial_x Y . \partial_y Y \tag{4.7}$$

$$||\nabla Y|| = \sqrt{\left(\partial_x Y\right)^2 + \left(\partial_y Y\right)^2} \tag{4.8}$$

$$\partial_x Y = m\left(d_x^+ Y, d_x^- Y\right) \tag{4.9}$$

$$\partial_x^{\pm} Y = \pm \left(Y\left(x \pm 1, y\right) - Y(x, y)\right) \tag{4.10}$$

where Y is the input image, dt is the step size and $m(x, y)$ is defined as below

$$m(x,y) = \begin{cases} sign(x).min(|x|,|y|) & \text{if } xy > 0 \\ 0 & \text{if } xy \leq 0 \end{cases} \tag{4.11}$$

4.4.4 Results of EDAT

To better understand the characteristics of Edge Adaptive Interpolation Using Total Variation Decomposition (EDAT), its results are compared with bilinear interpolation and bicubic interpolation methods that are widely used in super-resolution applications. Full reference results such as peak signal to noise ratio (PSNR) can not be computed since the original high resolution images are not available for comparison with superresolution image.

Evaluation of the results are made for the superresolution images created from low resolution images and low resolution images with different degradations that is likely to be introduced with regard to the image observation model, explained in Chapter 2.1.

The SR methods are carried out on standard test images which are available for public use.

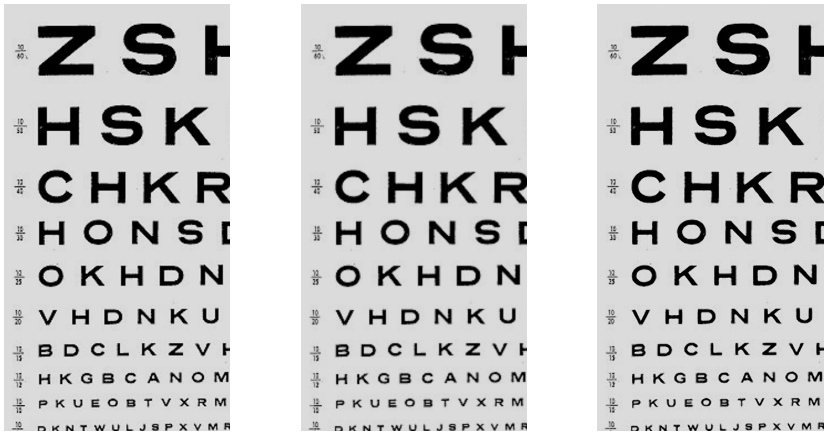


Figure 4.2: Comparison of results for Bicubic Interpolation (left), Bilinear Interpolation (middle) and EDAT (right).

The subjective comparisons from the interpolated images in Figure 4.2 shows

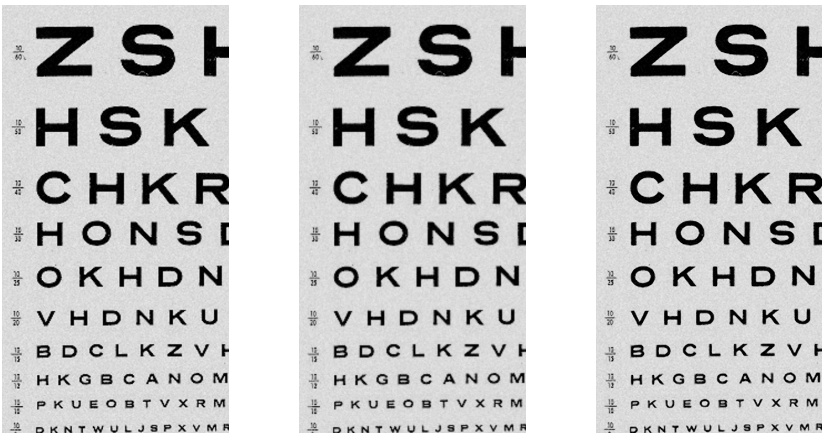


Figure 4.3: Comparison of results for Bicubic Interpolation (left), Bilinear Interpolation (middle) and EDAT (right) with Gaussian Noise.

that EDAT has evident improvement on the image than bilinear interpolation and bicubic interpolation methods. The edges appeared to be sharper in EDAT; because the structure component of the image is interpolated in an edge adaptive manner which is based on bilateral filter and a shock filter is applied after interpolation which further sharpened the edges apparent in the structure component of the image.

Subjective performance evaluations have also been conducted for the superresolution results performed on low resolution images with additive white noise, in order to see the noise amplification effects of SR methods. The superresolution results in Figure 4.3 shows that Bilinear Interpolation method suppresses noise which is evident in the LR image because of its blurring behaviour. However blurring also appeared at the sharp edge locations in the image and degraded the structure component of the image. EDAT and Bicubic Interpolation failed to suppress additive noise on the LR image; however the sharp edges were preserved in the superresolution image.

4.5 Example Based Learning for Superresolution

Example based learning for SR which is proposed in [29] uses a kernel ridge regression technique to estimate the high frequency details of the SR image. After regression, the regression results are post processed using a model of the generic image class to further resolve the artifacts introduced by SR. Experimental results, together with subjective results show that the method outperforms the existing example based image SR algorithms. The details of this algorithm is not in the scope of this thesis; however the algorithm is implemented to test its performance with the proposed SR IQA.

CHAPTER 5

PROPOSED ALGORITHM

Detailed descriptions and formulations of various no-reference image quality assessment algorithms are given in Chapter 3. In this Chapter, test results and potential weaknesses of these proposed algorithms will be given. An algorithm will be proposed to overcome the potential weaknesses of these algorithms.

5.1 Evaluation of No Reference Superresolution Quality Assessment Algorithms

A no reference image quality assessment algorithm for image superresolution using a natural scene statistics approach [1] is described in Chapter 3. The described method uses frequency falloff statistics that are computed from the total energy of the image in each scale computed by steering pyramid transform, which provides a tight frame by preserving the total energy of the image in the frequency domain.

The steerable pyramid transform [33] is defined in Fourier domain and the pyramid can be designed to be computed in any number of orientation subbands. Each of the subbands is computed by a directional derivative function. Therefore depending on the directions of the subbands in which the total energy is computed, the frequency falloff characteristic curve of the tested image might depart from the predicted frequency falloff slope coefficients in the algorithm. Therefore the algorithm might be vulnerable to the orientation of the high frequency details in the image depending on the subbands selected for the total

energy computation. The computed image quality score by the algorithm might not correlate well with the subjective scores if the total energy of the image is computed by oriented derivatives that does not align well with the high frequency details of the image. The diagram showing the spectral decomposition in angular and radial domains is given in Figure 5.1.

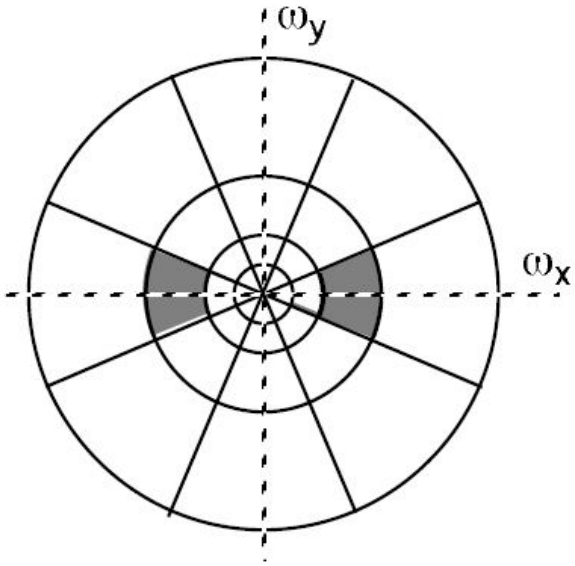


Figure 5.1: Spectral decomposition illustration of the steerable pyramid transform with 4 subbands. Shaded region corresponds to the vertically oriented subband.

The angular decomposition subbands computed in steering pyramid transform is useful in applications such as orientation decomposition, texture blending, depth from stereo and optical flow [33]. However, in its application in image quality assessment problem, the transform causes sensitivity to the chosen orientation subbands that the image energy is computed. In addition, angular decomposition of subbands is found to bring computational inefficiency to the algorithm.

A straightforward spatial continuity based statistics model has also been used in the algorithm. The spatial continuity metric computes the differences of neighbouring pixels and compares it with the statistics model computed from natural scenes. This method quantifies the continuity defects incorporated in the images after superresolution. However, the low resolution image which is processed by superresolution algorithms might contain degradations such as blurring and

aliasing as stated in Chapter 2.1. With the degradations in low resolution image, the straightforward spatial continuity computation proposed in this algorithm excessively fails.

The algorithm is tested with noise-free images and with images degraded by additive White Noise with $\sigma^2 = 0.0005$. The results are given in Figure 5.2 and 5.3.

(a) $D_s = 4.1632$ (b) $D_s = 2.7845$

Figure 5.2: Spatial continuity scores computed for noise-free LR image. (a) Bilinear interpolation vs. (b) EDAT.

(a) $D_s = 8.6522$ (b) $D_s = 2.6812$

Figure 5.3: Spatial continuity scores computed for LR image degraded with additive White Noise. (a) Bilinear interpolation vs. (b) EDAT.

The spatial continuity measure Ds is tested for different superresolution algorithms, namely bilinear interpolation and EDAT. With noise-free images, spatial continuity score of EDAT image outperformed the bilinear interpolated image which also complies with the subjective interpretations. However, when the same superresolution algorithms are performed on LR image with additive White Noise, spatial continuity score of EDAT decreases from 2.78 to 2.68 which states that the noisy image has better spatial continuity than the noise-free image. In addition, even though bilinear interpolation have suppressed the additive noise in the image because of its blurring behaviour, the spatial continuity score of the interpolated image is the highest with a score of 8.65. Results clearly conflict with subjective interpretations for noisy images. This shows the spatial continuity measure proposed in [1] fails for superresolution image quality assessment for noisy LR images. As the image observation model introduces various degradations to the LR image to be processed, image quality assessment algorithm in [1] is likely to perform poor in real applications.

The BRISQUE algorithm proposed in [23] evaluates the quality of images in a distortion-generic manner. Rather than other distortion-specific measures that quantify degradations such as ringing, blur or blocking, BRISQUE uses natural scene statistics of locally normalized luminance coefficients to quantify any kind of artifacts that create possible losses of image naturalness.

The algorithm is shown to perform well on different degradation types such as JP2K and JPEG compression, additive White Noise, Gaussian Blur and Rayleigh fast-fading channel simulation in [23]. BRISQUE is tested with bilinear interpolation and EDAT algorithm outputs for degraded and noise-free LR images.

The algorithm is tested with noise-free images and with images degraded with additive White Noise. The results are given in Figure 5.4 and 5.5.

The mean subtracted and contrast normalized features of the images are computed and the quality scores are quantified for different superresolution images, namely bilinear interpolation and EDAT. In noise-free image, EDAT appeared to perform slightly better than bilinear interpolation based on the computed

(a) $D_{br} = 38.7707$ (b) $D_{br} = 38.3199$

Figure 5.4: BRISQUE scores computed for noise-free LR image. (a) Bilinear interpolation vs. (b) EDAT.

(a) $D_{br} = 32.9063$ (b) $D_{br} = 41.2582$

Figure 5.5: BRISQUE scores computed for LR image degraded with Gaussian noise. (a) Bilinear interpolation vs. (b) EDAT.

BRISQUE score. This result correlates well with the subjective interpretations. EDAT was also predicted to perform better than bilinear interpolation since bilinear interpolation introduces blurring to the SR image whereas EDAT has better high frequency reconstruction properties.

Testing the algorithm with noisy images (additive White Noise), bilinear interpolation is predicted to perform better since the blurring behaviour acts as a

noise suppressor. BRISQUE scores complied with the predicted results since bilinear interpolation clearly outperformed EDAT with a score of 32.9 for bilinear interpolation and 41.25 for EDAT.

5.2 Robust Image Quality Evaluator for Image Superresolution Using Natural Scene Statistics

We have proposed a robust image quality assessment metric for image superresolution by combining the frequency falloff statistics model in [1] and mean subtracted and contrast normalized feature statistics in [23] which are both natural scene statistics (NSS) approaches to the no reference image quality assessment problem. By fusing these two approaches, we have reached the SR IQA framework shown in Figure 5.6.

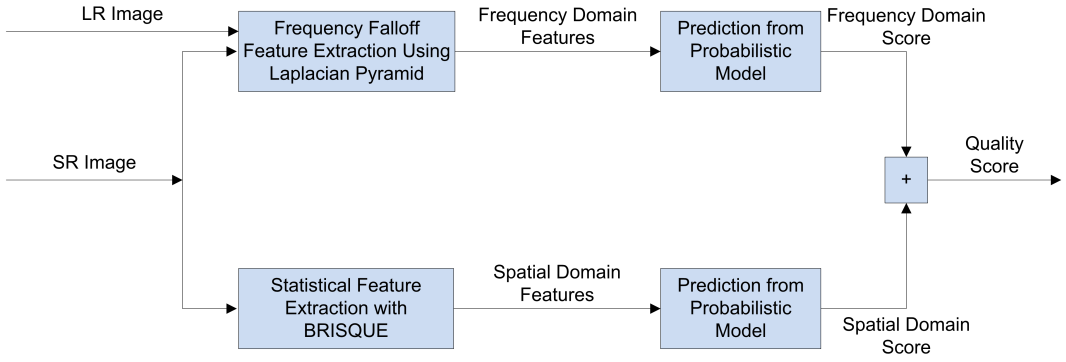


Figure 5.6: The proposed framework for proposed SR IQA method.

The proposed framework computes a quality score which is the sum of the normalized scores computed by spatial and frequency domain features. Scores computed by spatial domain statistics are computed from MSCN coefficient statistics which is described in detail in Chapter 3. In the frequency domain, frequency falloff characteristics which give information about the high frequency reconstruction properties of the SR algorithms is made use of. We constructed a statistical model for frequency energy falloff characteristics by decomposing both the LR and SR images into dyadic scales using the Laplacian Pyramid Transform.

Laplacian pyramid transform is a better choice for computing the total energy

of the images for the constructed statistical model; because the transform is independent of the orientation of high frequency contents in the image whereas the steering pyramid transform is sensitive to the orientation of the high frequency details in the image depending on the subbands selected for the total energy computation. In Laplacian Pyramid Transform, the image f_i is blurred by using a Gaussian kernel forming a low pass image, L_i. Then the low pass image is subtracted from the input image and a high pass image, h_i, is acquired. h_i can also be represented as the convolution of Laplacian of Gaussian function with the original image, which gives the transform its name. The same process is applied in each scale of the pyramid.

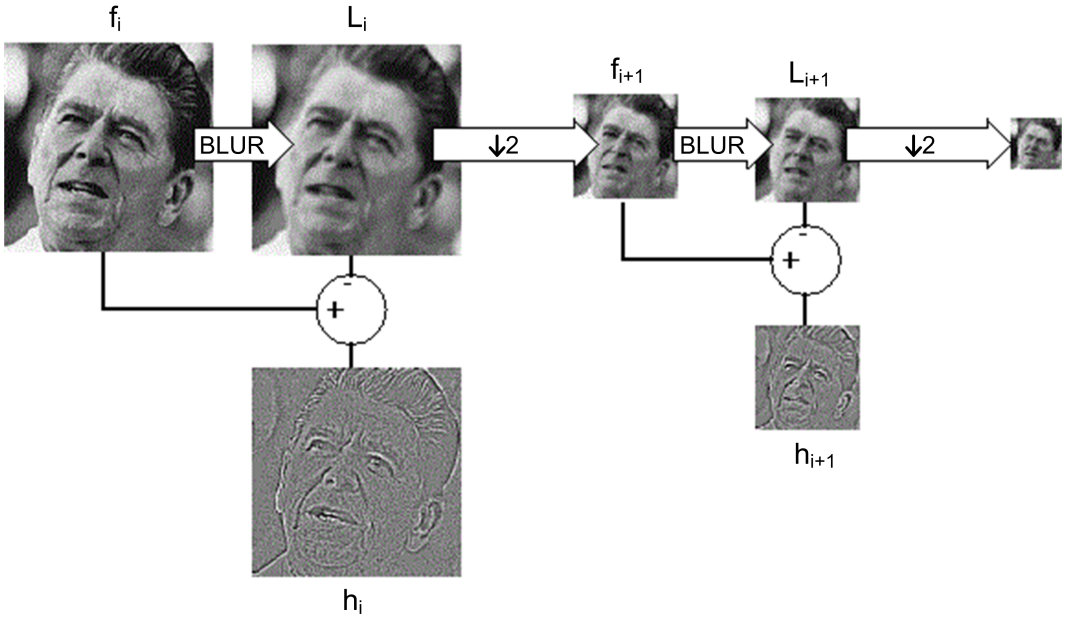


Figure 5.7: The Laplacian Pyramid Transform.

In addition, Laplacian pyramid transform is better in terms of computational efficiency compared to the steering pyramid transform which is overcomplete by $4k/3$ where k is the number of orientation bands. A comparison of the properties of steering pyramid transform and Laplacian pyramid transform is given in Table 5.1.

The natural scene statistics model for frequency falloff characteristics is constructed by 2800 natural images which are available in the MIT CVCL Image Database. Firstly, Laplacian Transform is applied to each LR and SR images

Table5.1: Comparison of Laplacian Pyramid and Steering Pyramid representations.

	Steering Pyramid	**Laplacian Pyramid**
self-inverting	yes	no
overcompleteness	$4k/3$	$4/3$
rotated orientation subbands	yes	no

and the energy in each scale is computed by sum of squared transform coefficients. Energy computation is made for 6 scales for HR image and 5 scales for LR image. The frequency falloff curves are plotted from coarse to fine scales. The plot of frequency falloff curves for 24 Kodak Images is given in Figure 5.9.

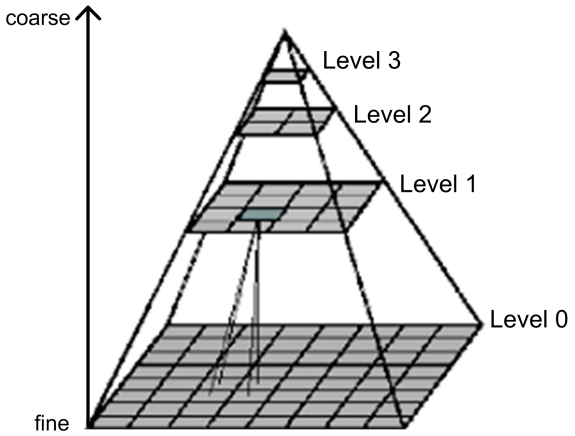


Figure 5.8: Constructed Pyramid for Multiple Levels

Observing the curves, the frequency falloff curves appeared to be straight lines in the log-log scale. This result complies with the $1/f^p$ relationship which states that the amplitude spectrum of the natural images falls with the spatial frequency, f. In addition, the frequency falloff curves computed from LR and HR images are found to be parallel lines in between the same scales. Therefore the frequency falloff lines for HR images are found to be highly predictable from frequency falloff lines of the corresponding LR images. These results show that a statistical model which comprises the idea in [1] can also be constructed by using the Laplacian Pyramid Transform.

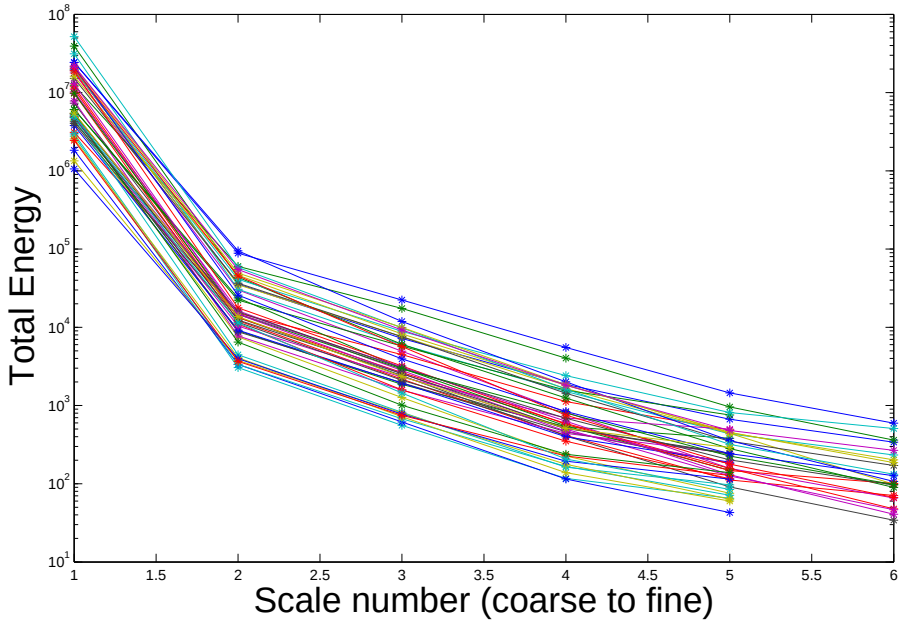


Figure 5.9: Frequency falloff curves computed using Laplacian Pyramid Transform for 24 Kodak Images.

The slopes of frequency falloff of HR image is predicted as

$$\begin{aligned}
\hat{s}_1^H &= s_1^L \\
\hat{s}_2^H &= s_2^L \\
\hat{s}_3^H &= a_0 + a_1 s_3^L \\
\hat{s}_4^H &= b_0 + b_1 s_4^L \\
\hat{s}_5^H &= c_0 + c_1 \hat{s}_3^H + c_2 \hat{s}_4^H
\end{aligned} \tag{5.1}$$

where s_i^L is the falloff curve slope of LR image in i-th scale, $\hat{s}_i^H$ is the predicted falloff curve slope of HR image in i-th scale and $a_0, a_1, b_0, b_1, c_0, c_1$ and c_2 are obtained by simple least square regression using the statistical model built by real LR and HR images. The predicted coefficients from the regression are given in Table 5.2.

Then, normalized frequency energy falloff error between the predicted slope and true slope of HR image in the finest scale is computed by Equation 5.2.

$$e_f = \frac{\hat{s}_5^H - s_5^H}{\hat{s}_5^H} \tag{5.2}$$

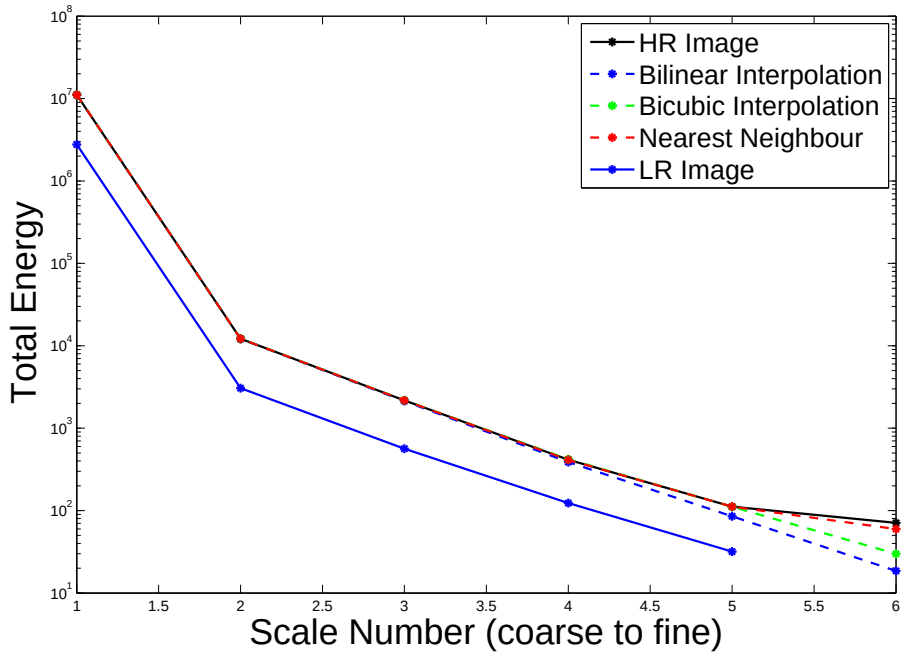


Figure 5.10: Frequency energy falloffs of original HR, SR and LR images.

Table5.2: Least Square Regression Results for Frequency Falloff Curve Slopes

Parameter	**Regression Result**
a_0	0.0155
a_1	1.0036
b_0	-0.0589
b_1	0.9897
c_0	-1.1841
c_1	-0.9671
c_2	1.3798

Normalized error values, which are computed from 2800 images in the MIT CVCL Image Dataset is shown on an histogram in Figure 5.11. In ideal case when the SR image is close to the HR image, e_f approaches to zero.

From the natural scene statistics computed from HR images in the dataset, it is clear that the histogram has a Gaussian (even peakier than Gaussian) characteristic and GGD function in Equation 5.3 can be fitted.

$$p_{e_f}(e_f) = \frac{1}{Z_f} \exp\left[-\left(\frac{|e_f - \mu_f|}{\alpha_f}\right)^{\beta_f}\right] \tag{5.3}$$

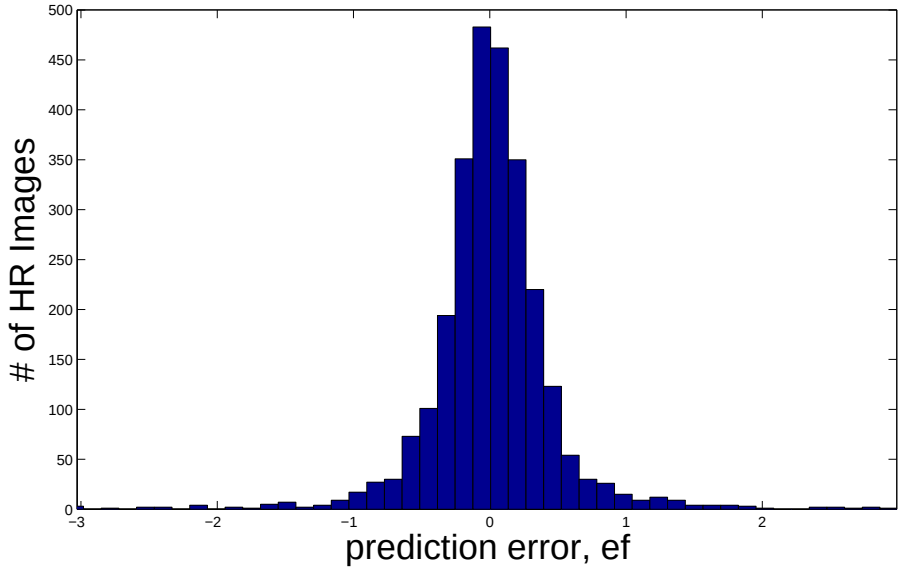


Figure 5.11: Histogram of normalized frequency energy falloff error, e_f, for HR Images.

where $Z_f = \dfrac{\beta_f}{2\alpha_f \Gamma\left(1/\beta_f\right)}$ is a normalization factor. Our results from Figure 5.11 complies with the results of Yeganeh et. al. in [1], therefore the same maximum likelihood based fitting technique is used for estimating the frequency energy falloff error distribution function computed by the Laplacian Pyramid Transform coefficients. The fitted parameters for our GGD model is $\mu_f = 0.029, \alpha_f = 0.0608$ and $\beta_f = 0.6124$.

From the natural image probability model p_{ef}, SR image which is more natural and similar to the HR image tend to achieve maximum values with high probabilities. The probability-based measure is then converted to a quality assessment measure by Equations 5.4 and 5.5 where $K = 1/Z_f$ is the normalization factor such that the naturalness measure p_f is upper-bounded by 1.

$$p_f = \frac{1}{K} p_{ef} \tag{5.4}$$

$$D_f = -log p_f \tag{5.5}$$

The spatial domain feature extraction and prediction from the probabilistic model is done with BRISQUE algorithm which is described in detail in Chapter

3.2. The BRISQUE score, D_{br} is then added with the frequency energy falloff naturalness measure D_f and the final image quality assessment model is achieved as in Equation 5.6.

$$D_{fs} = D_f + D_{br} \tag{5.6}$$

Both the BRISQUE score (D_{br}) and frequency falloff score (D_f) are normalized in the proposed metric. For better correlation with subjective evaluations, the spatial domain and frequency domain components of the achieved model is weighted as in Equation 5.7. The final weighting factor, w, is found to be 0.7. By choosing $w = 0.7$, frequency energy falloff component of the metric which characterizes the high frequency reconstruction properties in the SR image is given more weight, which in result was found to correlate better with human perception.

$$D_{fs} = wD_f + (1 - w)D_{br} \tag{5.7}$$

CHAPTER 6

EXPERIMENTAL RESULTS

In this Chapter, the algorithm proposed in Chapter 5 is tested with SR images computed from noise-free LR images and SR images from LR images degraded with additive white noise and Gaussian Blur. The results of the proposed SR IQA algorithm will be compared with the SR IQA algorithm proposed in [1]. Five different SR algorithm performances are going to be measured in the experiments. These algorithms are namely; nearest neighbour interpolation, bicubic interpolation, bilinear interpolation, EDAT and Kim's example based SR algorithm which is discussed in Chapter 4.

6.1 Experimental Results

The algorithm is tested with 11 images with different contents. Images used in the experiments does not contain images from MIT CVCL Image Database; because it was used to estimate the NSS model of the frequency energy falloff characteristics which is used within the proposed image quality assessment metric.

Before starting the experiments, the tested SR images are evaluated subjectively. All the images appeared to have better edge reconstruction and better high frequency component estimation for Kim's SR method. EDAT appeared to have sharper edge reconstruction properties compared to bilinear and bicubic interpolation methods. Bilinear interpolation performs worst in high frequency detail estimation. Details of the images showing the high frequency reconstruction

success of the SR algorithms are given in Figure 6.1.

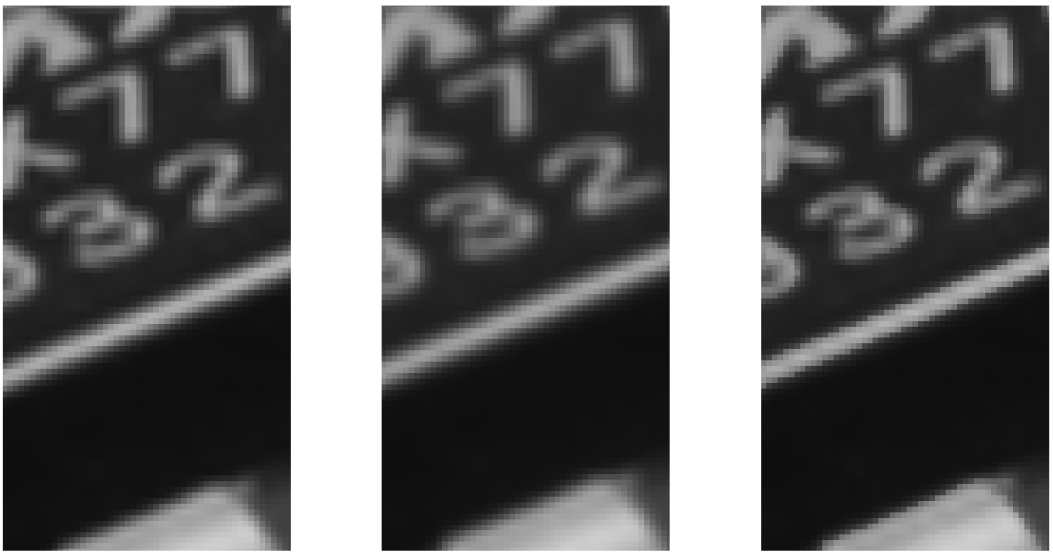


Figure 6.1: Edge Details of SR Images (left to right: bilinear, bicubic, EDAT)

In the first experiment, the quality scores are measured for SR images computed from noise-free LR images. Computed scores from the SR IQA algorithm in [1] and the proposed algorithm is given in Table 6.1.

SR images computed from noise-free LR pairs are given in Figure 6.2 for subjective evaluation. For all of the test images except Chip image, according to the scores from SR IQA algorithm in [1], EDAT appeared to have the best reconstruction. In EDAT, the structure component of the image is interpolated in an edge adaptive manner and it recovers the sharp edges in the image well. However, example based SR method by Kim appeared to have best high frequency detail reconstruction according to the subjective evaluation. Therefore, subjective evaluation does not comply with the SR IQA results by [1]. However, subjective evaluation clearly comply with the scores computed from the proposed SR IQA metric where example based SR method by Kim clearly outperforms other SR algorithms. It is seen that the proposed SR IQA metric is successful in characterizing the high frequency reconstruction properties of SR algorithms

Subjective evaluation also shows that bicubic interpolation outperforms bilinear interpolation. There is excessive loss of high frequency details and added blur

Table6.1: Comparison of SR Image Quality Assessment Algorithms (Noise-Free LR images).

SR IQA in [1]					
	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	60.4562	12.8840	10.5184	**6.7547**	16.1115
Building	69.3070	15.5868	13.1546	**5.5918**	15.3108
Cartoon	53.0112	9.3280	8.1596	**3.7965**	9.3540
Child	58.6642	8.9588	7.5221	**6.4576**	11.2646
Chip	37.8639	7.0748	6.5620	5.7466	**4.2999**
Flowers	70.9156	11.3057	9.4977	**5.4594**	11.9863
Girl	46.8716	8.4534	7.0977	**5.7877**	8.3280
Obama	50.6122	8.1627	6.6908	**5.7409**	8.6066
Roman	62.1855	12.0108	9.9200	**5.8171**	16.4117
Text	68.4226	10.5081	7.5490	**3.9147**	29.9323
Train	81.4756	17.4008	14.1500	**5.9375**	25.2450
Proposed SR IQA					
	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	21.8536	22.6788	25.0026	24.9481	**17.8570**
Building	17.3102	17.6890	19.1113	20.1846	**15.8313**
Cartoon	20.9799	20.1835	22.0950	20.1875	**17.8600**
Child	13.8322	17.2968	18.9466	17.5206	**12.9688**
Chip	23.7895	25.8179	27.9332	23.7799	**23.2761**
Flowers	17.5378	16.5787	19.1224	17.5454	**10.6110**
Girl	16.9676	17.2533	19.8195	17.6366	**11.6887**
Obama	19.9406	19.8244	21.9978	19.5786	**16.9733**
Roman	16.8417	17.2547	19.9248	19.0946	**13.0411**
Text	30.5320	22.4072	24.5243	20.3790	**19.0959**
Train	23.4504	25.5440	26.5980	27.3110	**23.2471**

on bilinear interpolated images. However, SR IQA algorithm in [1] evaluates the performance of bilinear interpolation better than the bicubic interpolation method for all of the test images. This evaluation shows that SR IQA algorithm in [1] fails to characterize blurring in SR images. This failure is evident because the spatial continuity measure in [1] characterizes blurry images as in perfect spatial continuity. However, according to the scores from the proposed SR IQA metric, bicubic interpolation has better performance than bilinear interpolation on all images tested. This shows that the proposed SR IQA algorithm evaluates blurring well which may be involved after SR process. The proposed SR IQA algorithm performs better as the computed features are distortion-unaware and characterize any kind of degradations.

Scores computed by proposed SR IQA method for noise-free images in Table 6.1 shows that nearest neighbour interpolation method performs almost as good as SR method by Kim for some of the test images. However, subjective evaluations

show that nearest neighbour method has evident artefacts in the spatial domain, mainly at the edge locations. To further analyze this observation in detail, D_f and D_{br} scores computed by the proposed SR IQA method is decomposed and compared in Table 6.2. Results show that nearest neighbour method has better frequency energy falloff scores, D_f, than SR by Kim method for some of the test images. This result is not surprising as the frequency energy falloff curves of SR images created by nearest neighbour method appears to approach the frequency falloff curves of natural HR image, as shown in Figure 5.10. In nearest neighbour method, high frequency components are reconstructed but these high frequency components are erroneous. However, the spatial domain score D_{br} characterizes the degradations involved from nearest neighbour method and after weighting in between D_f and D_{br}, the final quality score shows that nearest neighbour method does not outperform SR by Kim method. The results from the proposed metric correlate well with the subjective evaluations, which is given in Chapter 6.2.

Table6.2: Frequency and Spatial Domain Score Comparison for Nearest Neighbour and SR by Kim methods.

	Proposed SR IQA			
	Nearest		**Kim**	
	D_f	D_{br}	D_f	D_{br}
Boat	7.0907	56.3005	7.6831	41.5961
Building	13.5683	26.04132	12.8930	22.6872
Cartoon	7.9345	51.4192	7.2821	42.5418
Child	7.8518	27.78626	7.9127	24.7662
Chip	7.5594	61.6597	6.5115	62.3935
Flowers	5.5188	45.5822	6.3731	20.4995
Girl	7.6184	38.7826	7.4944	21.4754
Obama	5.5208	53.5867	5.1485	44.5647
Roman	6.4835	41.0106	6.2333	28.9259
Text	5.6327	88.6304	5.9090	49.8654
Train	16.2311	40.2955	16.0610	40.0148

Then, SR IQA algorithms are tested for SR images created from their LR pairs degraded with additive white (Gaussian) noise. The tests are performed for 2 different levels of additive white noise, $\sigma^2 = 0.0005$ and $\sigma^2 = 0.005$. Computed scores from the SR IQA algorithms are given in Table 6.3 and Table 6.4.

Table6.3: Comparison of SR Image Quality Assessment Algorithms (degraded LR image with Additive White Gaussian Noise, $\sigma^2 = 0.0005$).

SR IQA in [1]

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	80.3760	18.3776	14.2634	**6.4244**	32.3479
Building	89.2175	20.6667	16.5154	**5.2402**	32.2779
Cartoon	73.4363	15.5291	12.2154	**3.3611**	29.2838
Child	77.6167	14.9729	11.6164	**6.1269**	27.0623
Chip	65.8116	14.6043	11.4822	**4.7066**	26.9986
Flowers	86.1586	16.2217	12.6931	**4.9212**	27.7601
Girl	65.6570	14.2321	10.9826	**4.9150**	27.5597
Obama	75.2340	14.7734	11.2445	**5.2818**	28.8353
Roman	82.0229	17.1898	13.4318	**5.3411**	31.4847
Text	97.1539	17.8885	13.0504	**3.8078**	46.2584
Train	99.0357	22.0110	17.2666	**5.8138**	38.5359

Proposed SR IQA

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	26.9124	19.9360	21.1985	21.9228	**19.1378**
Building	37.3496	29.6145	**30.0069**	32.6892	30.7537
Cartoon	28.8982	**20.5053**	21.5482	22.7403	21.5365
Child	24.4231	19.4661	19.7624	20.3931	**19.3231**
Chip	27.8136	**19.2287**	19.6108	21.3017	24.5666
Flowers	22.4774	16.6629	18.5954	18.7788	**15.7755**
Girl	29.0578	**22.0513**	22.9084	22.8418	22.5673
Obama	25.9003	**17.6906**	17.9040	18.4770	21.6383
Roman	25.0196	**18.5627**	19.8326	20.3778	19.5088
Text	26.1761	20.1335	**19.3930**	22.9294	24.1648
Train	39.3712	34.9735	36.0649	35.1984	**33.4717**

SR images computed from noisy LR pairs are given in Figure 6.3 for subjective evaluation. SR image qualities are not easy to distinguish in between SR methods computed from noisy LR pairs. When SR algorithms are applied to LR images with added noise, noise masks the high frequency reconstructed parts in the SR image and reconstruction properties of SR algorithms are not evident. However different SR methods amplify noise in different degrees. While bilinear interpolation blurs sharp edges within the image, it also filters out the added noise. On the other hand, bicubic interpolation, EDAT and example based SR by Kim are found to amplify added noise in different degrees in subjective evaluations When SR images are computed from LR images with added noise, SR IQA algorithm in [1] fails drastically. Results from SR IQA method in [1] show that performance of EDAT clearly beats other SR algorithms; however subjective evaluation shows that the performance of SR images are very close and hard to distinguish. The results clearly fail that even the scores of SR images

Table6.4: Comparison of SR Image Quality Assessment Algorithms (degraded LR image with Additive White Gaussian Noise, $\sigma^2 = 0.005$).

SR IQA in [1]

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	143.4593	34.2873	25.3170	**5.3099**	81.5292
Building	151.0179	36.1716	27.1318	**4.4066**	81.7179
Cartoon	140.0075	32.7449	24.2523	**2.7534**	82.3142
Child	139.7092	31.7943	23.4545	**4.5627**	82.3701
Chip	138.0636	31.9951	23.5294	**3.2330**	81.7848
Flowers	147.7582	32.9568	24.1357	**3.6917**	82.5354
Girl	128.7154	30.8234	22.7782	**4.1663**	77.7122
Obama	139.9133	30.6441	22.3130	**3.7180**	80.0266
Roman	145.5876	33.5448	24.7360	**4.2603**	84.1922
Text	160.3236	33.9919	24.4010	**3.5310**	86.8044
Train	152.9488	36.2685	27.0171	**4.9522**	81.3277

Proposed SR IQA

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	38.3629	22.6529	**22.2959**	22.1269	30.2431
Building	37.2306	22.7804	**22.3338**	22.4976	30.5004
Cartoon	36.8763	21.1502	**20.3495**	21.1238	29.6421
Child	35.2674	23.0808	**21.5293**	22.7449	30.6955
Chip	38.6244	24.4201	**23.7469**	24.4777	32.3164
Flowers	38.5890	27.4534	**27.3187**	27.8643	31.8079
Girl	37.0760	21.4547	**20.5405**	21.5783	28.8873
Obama	68.5168	69.1726	72.4384	67.2633	**47.8590**
Roman	37.8971	26.7880	**26.4310**	27.0325	32.0890
Text	35.2503	23.5911	**22.0912**	25.6701	30.9457
Train	36.5902	22.8058	**22.4244**	22.6671	29.8390

computed from noisy LR pairs beats the scores of SR images computed from noise-free LR pairs in Table 6.1, which is not possible in practice. Proposed SR IQA algorithm computes the scores of all SR algorithms relatively close to each other such that the results correlate well with subjective evaluation. As the level of added noise increases up to $\sigma^2 = 0.005$, bilinear interpolation starts to beat all other SR methods. The reason is, although blurring the high frequency details in the image, bilinear interpolation suppresses additive noise whereas all other SR methods under test amplifies added noise while interpolating the LR image. When the amount of additive noise increases, the suppressive effect of bilinear interpolation increase, which increased the performance score of this method among others.

Lastly, SR IQA algorithms are tested for SR images created from their LR pairs degraded with Gaussian blur, with a window size of 3x3 and $\sigma^2 = 1$. Computed scores from the SR IQA algorithms are given in Table 6.5.

Table6.5: Comparison of SR Image Quality Assessment Algorithms (degraded LR image with Gaussian Blur).

SR IQA in [1]

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	38.1345	4.5507	**3.7931**	6.6165	4.0783
Building	42.5897	7.8667	7.3183	**6.4851**	6.5457
Cartoon	36.1613	4.4802	4.1104	4.7060	**3.5365**
Child	40.6226	2.8749	**2.3225**	5.9721	2.6829
Chip	30.7277	4.0910	**3.5838**	5.1303	3.9733
Flowers	49.9634	5.2862	4.7419	5.4918	**4.2707**
Girl	32.5312	3.0461	2.5311	5.3608	**2.5134**
Obama	37.5812	2.9662	**2.4028**	5.6357	2.7350
Roman	41.3476	5.1717	4.5542	6.1950	**4.5087**
Text	56.6468	7.4130	6.9176	**5.5158**	7.1757
Train	44.3067	5.9838	5.2923	6.3586	**4.7334**

Proposed SR IQA

	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	23.9967	23.9706	25.5336	22.7168	**22.4267**
Building	14.0052	16.2427	18.1891	16.1197	**13.7944**
Cartoon	20.6341	22.9490	24.8005	21.4629	**20.0530**
Child	**19.5120**	22.2849	23.6487	21.3466	20.2453
Chip	**26.3155**	30.2764	31.1038	26.6983	29.0993
Flowers	25.8086	22.1569	23.6566	**20.0762**	20.9374
Girl	22.3707	23.9516	25.6027	23.3191	**22.3509**
Obama	**21.8161**	25.8264	27.0788	24.7189	24.4091
Roman	22.4182	23.0800	24.5765	21.9345	**21.0493**
Text	27.7228	25.6356	26.7654	23.2936	**22.7318**
Train	18.7328	20.5763	22.0845	19.7198	**17.5228**

SR images computed from Gaussian blurred LR pairs are given in Figure 6.4 for subjective evaluation. Bilinear interpolation method appears to blur the SR image even more than its blurred LR pair. Nearest Neighbour, Bicubic, EDAT and example based SR by Kim appeared to enhance the image in similar performances as the high frequency details in the LR image was already removed. Scores from both SR IQA methods claimed that all the SR algorithms performed in close performances which also complies with the subjective evaluation. The proposed SR IQA method also shows that bilinear interpolation performs worst which complies with predicted results, as bilinear interpolation degrades (blurs) the SR image even more than its blurred LR pair. However, there appears an evident failure in quality assessment by [1] as the scores of SR images computed by bilinear interpolation method beats other SR methods, even though the SR image appears to have excessive blur. This failure is evident because the spatial continuity measure in [1] characterizes blurry images as in perfect spatial

continuity.

The empirical results show that the proposed SR IQA method clearly performs better while evaluating the degradations added either by the superresolution or by image observation model while acquiring the LR image. SR IQA method in [1] performs as good as the proposed SR IQA method while measuring the high frequency reconstruction properties of SR algorithms; however the main difference in between the SR IQA performances becomes evident in spatial domain statistics measures. D_s component of SR IQA in [1] plainly fails to characterize distortions amplified by SR algorithms as it is a straightforward approach to characterize spatial continuity. However in D_{br} component of proposed SR IQA, 36 features in multiple scales are computed which detects and quantifies any kind of degradation (distortion-unaware) in multiple scales, which may be evident after SR process.

6.2 Validation by Subjective Tests

We have conducted a subjective experiment to better validate the proposed objective SR IQA algorithm. In the subjective test, 10 subjects were asked to score 5 SR images that is created from different 11 LR images. In total, 55 SR images were evaluated by subjects. The SR images are created by different SR methods which are namely nearest neighbour, bicubic, bilinear interpolation, EDAT and SR by Kim methods.

Subjects are shown the SR images created from each LR image simultaneously, and asked to score each image within a scale of 1-5 after pairwise comparison in between SR images. Before computing the average mean opinion scores (MOS) for the test images, outliers are separated from the collected data. The results are compared with the scores computed by proposed SR IQA method in Table 6.6.

Subjective test results showed that SR by Kim method clearly outperforms the other 4 SR methods. This result correlate well with the proposed SR IQA scores. Subjective results also showed that in some images, it is hard for subjects

Table6.6: Comparison of Proposed SR IQA Method with Mean Opinion Scores (Noise-Free LR images).

Average Mean Opinion Scores (MOS)					
	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	3.0	2.9	2.4	2.4	**5.0**
Building	2.4	3.6	2.1	3.3	**4.9**
Cartoon	2.4	3.1	**4.0**	3.7	3.4
Child	2.1	3.3	2.1	3.6	**5.0**
Chip	2.9	3.4	2.4	3.9	**4.9**
Flowers	3.0	3.1	2.6	3.3	**4.9**
Girl	2.3	3.4	2.7	4.0	**5.0**
Obama	2.1	3.1	3.3	3.8	**3.9**
Roman	2.1	3.6	2.1	3.7	**5.0**
Text	2.6	3.6	2.7	4.0	**4.1**
Train	2.4	3.3	2.3	3.7	**4.9**
Proposed SR IQA					
	Nearest	**Bicub.**	**Bilin.**	**EDAT**	**Kim**
Boat	21.8536	22.6788	25.0026	24.9481	**17.8570**
Building	17.3102	17.6890	19.1113	20.1846	**15.8313**
Cartoon	20.9799	20.1835	22.0950	20.1875	**17.8600**
Child	13.8322	17.2968	18.9466	17.5206	**12.9688**
Chip	23.7895	25.8179	27.9332	23.7799	**23.2761**
Flowers	17.5378	16.5787	19.1224	17.5454	**10.6110**
Girl	16.9676	17.2533	19.8195	17.6366	**11.6887**
Obama	19.9406	19.8244	21.9978	19.5786	**16.9733**
Roman	16.8417	17.2547	19.9248	19.0946	**13.0411**
Text	30.5320	22.4072	24.5243	20.3790	**19.0959**
Train	23.4504	25.5440	26.5980	27.3110	**23.2471**

to distinguish the performance of bicubic and EDAT algorithms. This result also validated the proposed algorithm; because bicubic interpolation and EDAT have close objective scores for all of the test images. For bilinear and nearest neighbour interpolation methods, subjective tests gave the lowest scores. In the proposed SR IQA, bilinear interpolation also gave the lowest scores. However proposed SR IQA metric showed that nearest neighbour method beats bilinear and bicubic interpolation methods for some test images. The reason is because of the method's erroneous high frequency reconstruction properties, which could not be distinguished by frequency energy falloff statistics. Even though there is pixelation and aliasing artefacts in NN method, total energy in the frequency domain for finest scale is higher than other methods, therefore giving better score D_f. Subjective results clearly validated that the proposed SR IQA method is an objective SR IQA method which correlates well with human perception whereas on the other hand being robust against any kind of degradation in the LR image.

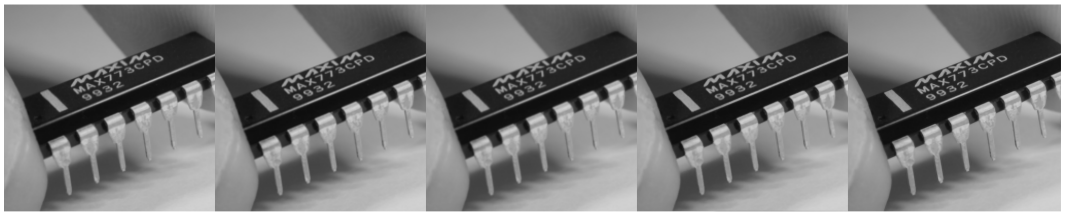
(a) Boat

(b) Building

Figure 6.2: SR Results for Noise-Free LR Pairs (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(c) Cartoon

(d) Child



(e) Chip

Figure 6.2: SR Results for Noise-Free LR Pairs (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(f) Flowers

(g) Girl

(h) Obama

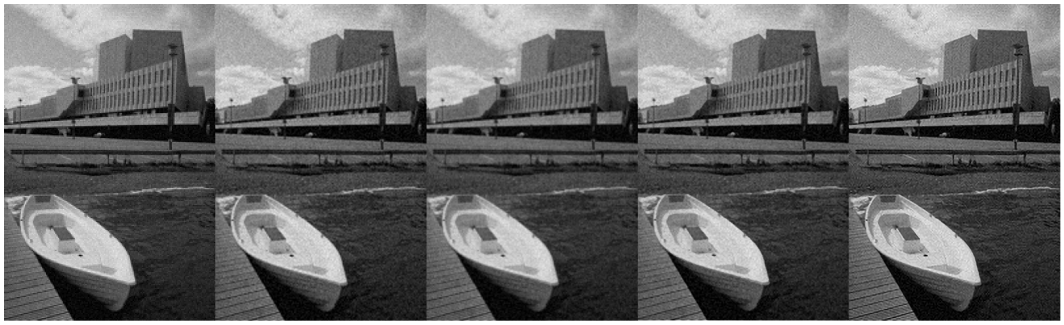
Figure 6.2: SR Results for Noise-Free LR Pairs (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(i) Roman

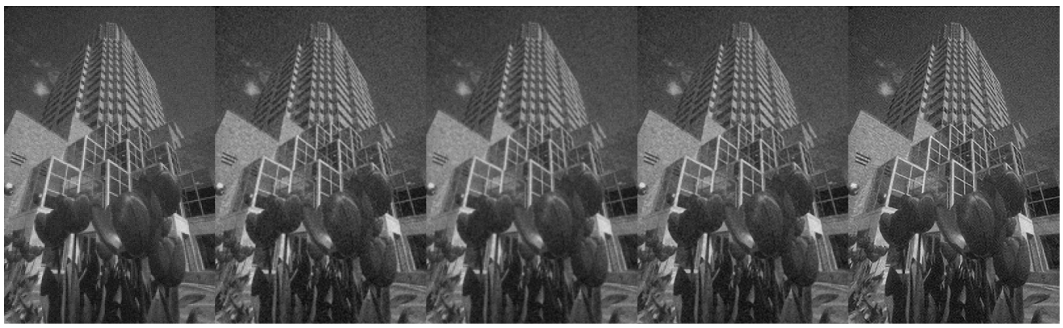
(j) Text

(k) Train

Figure 6.2: SR Results for Noise-Free LR Pairs (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)



(a) Boat

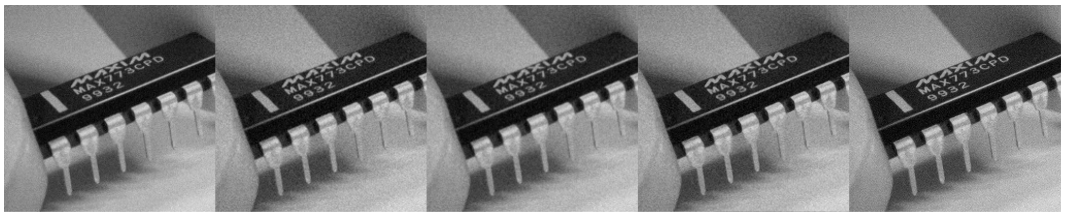


(b) Building

Figure 6.3: SR Results for LR Pairs Degraded with White Noise, $\sigma^2 = 0.0005$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(c) Cartoon

(d) Child

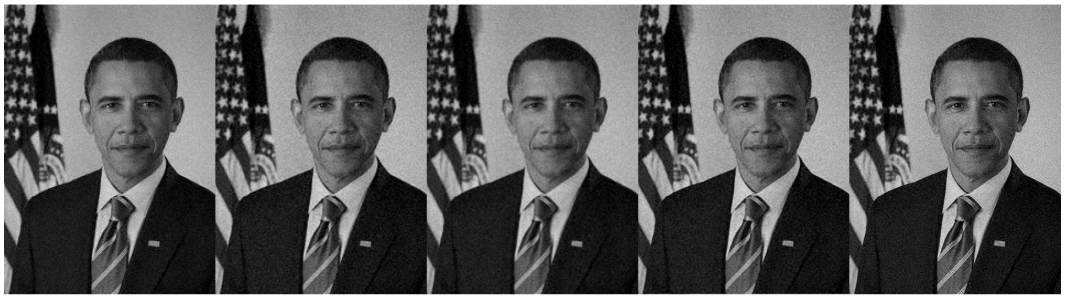


(e) Chip

Figure 6.3: SR Results for LR Pairs Degraded with White Noise, $\sigma^2 = 0.0005$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(f) Flowers

(g) Girl



(h) Obama

Figure 6.3: SR Results for LR Pairs Degraded with White Noise, $\sigma^2 = 0.0005$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(i) Roman

(j) Text

(k) Train

Figure 6.3: SR Results for LR Pairs Degraded with White Noise, $\sigma^2 = 0.0005$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

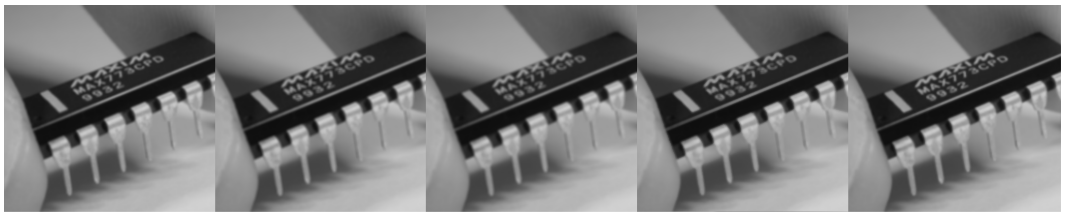
(a) Boat

(b) Building

Figure 6.4: SR Results for LR Pairs Degraded with Gaussian Blur, 3x3 window size, $\sigma^2 = 1$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(c) Cartoon

(d) Child



(e) Chip

Figure 6.4: SR Results for LR Pairs Degraded with Gaussian Blur, 3x3 window size, $\sigma^2 = 1$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(f) Flowers

(g) Girl

(h) Obama

Figure 6.4: SR Results for LR Pairs Degraded with Gaussian Blur, 3x3 window size, $\sigma^2 = 1$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

(i) Roman

(j) Text

(k) Train

Figure 6.4: SR Results for LR Pairs Degraded with Gaussian Blur, 3x3 window size, $\sigma^2 = 1$ (left to right: NN, bicubic, bilinear, EDAT, Example Based SR by Kim)

CHAPTER 7

CONCLUSION AND FUTURE WORK

7.1 Conclusions

In this thesis, we have proposed a no-reference image quality assessment metric for image superresolution which correlates well with human perception as well as being robust against any type of degradation.

The proposed SR IQA algorithm is a natural scene statistics approach to the objective SR IQA problem. In the algorithm, multiple natural scene statistics models are computed for natural HR images. Different features both in frequency domain and spatial domain are computed for SR images under test and compared with the developed NSS model of HR images.

In the frequency domain, frequency energy falloff statistics are computed by using Laplacian Pyramid Transform which computes the total energy of the image in multiple scales very efficiently in time compared to other available pyramid transforms. The frequency falloff curve is predicted for the finest scale and compared with the NSS model which quantifies the high frequency detail reconstruction capability of the SR process. The departures from the NSS model constructed from natural HR images are quantified as the overall image quality score in the frequency domain, D_f.

In the spatial domain, Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE) algorithm is implemented. This method is a NSS based no-reference method. The method is chosen to be used in the proposed algorithm as it

is a distortion-generic (distortion-unaware) NR quality assessment algorithm which correlates well with the subjective tests for any kind of degradations. The method computes image quality from MSCN coefficients and statistics of pairwise products of locally normalized neighbouring pixel luminance values. The features are computed for 2 scales as distortion in an image affects structure across the scales. Each of these features extracted in the algorithm correlate well with the human perception. Again, the departures from the NSS model constructed from natural HR images is quantified as the overall image quality score in the spatial domain, D_{br}. From the subjective evaluation, empirically, weighting factors are found for image quality scores in spatial and frequency domain and a final SR image quality assessment metric, D_{fs}, is found.

The proposed SR IQA algorithm is tested for 5 different SR algorithms, nearest neighbour interpolation, bilinear interpolation, bicubic interpolation, EDAT and example based SR by Kim, which has better high frequency reconstruction properties compared to other methods. In addition, robustness of the algorithm is tested with LR images distorted with different degradations, namely Gaussian Blur and additive White Noise. The results of the proposed algorithm is compared with the results of SR IQA algorithm in [1] which is a first attempt to design an NSS-based objective method to assess the quality of HR images created using SR methods. Subjective test results show that the proposed SR IQA metric correlates well with the human perception. The results also show that the proposed SR IQA metric is robust against any kind of distortion, which is likely to be introduced to the LR image as a result of the image observation model or during SR process. In addition, the proposed algorithm is efficient in time with low computational complexity; therefore applicable to real applications. No reference metrics are important not only for superresolution applications, but also for applications involving image formation.

7.2 Future Work

In our work, we have validated our proposed SR IQA algorithm by conducting subjective experiments with 10 subjects, with noise-free images. As a future

work, a large scale subjective experiment can be conducted with a larger image set which contain HR images created using other different SR methods. Subjective evaluations might also be conducted for SR images created from LR images degraded with different types distortions.

The NSS model in the frequency domain in Chapter 5.2 is constructed from 2800 natural HR images in MIT CVCL image database. The contents of the images in MIT CVCL database are very similar. Model might be constructed by using a wider set of HR images with various contents which will be a better representation of natural scene statistics. This approach might improve the performance of the frequency domain computation in the proposed SR IQA method.

The proposed algorithm is tested for robustness by testing it with LR images degraded with 2 different distortion types, additive White Noise and Gaussian Blur. Tests can be conducted with a wider image set that contains images with various other distortions and the proposed SR IQA measure can be tested for robustness against these distortion types.

REFERENCES

[1] Hojatollah Yeganeh, Mohammad Rostami, and Zhou Wang. Objective quality assessment for image super-resolution: A natural scene statistics approach. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 1481–1484. IEEE, 2012.

[2] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004.

[3] ITUR Rec. Bt. 500-11, methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union Std*, 2002.

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

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

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

[7] Chaofeng Li and Alan C Bovik. Content-partitioned structural similarity index for image quality assessment. *Signal Processing: Image Communication*, 25(7):517–526, 2010.

[8] Zhou Wang and Alan C Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2(1):1–156, 2006.

[9] Dacheng Tao, Xuelong Li, Wen Lu, and Xinbo Gao. Reduced-reference iqa in contourlet domain. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(6):1623–1627, 2009.

[10] Xinbo Gao, Wen Lu, Dacheng Tao, and Xuelong Li. Image quality assessment based on multiscale geometric analysis. *Image Processing, IEEE Transactions on*, 18(7):1409–1423, 2009.

[11] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. A no-reference perceptual blur metric. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 3, pages III–57. IEEE, 2002.

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

[13] Zhou Wang, Alan C Bovik, and BL Evan. Blind measurement of blocking artifacts in images. In *Image Processing, 2000. Proceedings. 2000 International Conference on*, volume 3, pages 981–984. IEEE, 2000.

[14] Shizhong Liu and Alan C Bovik. Efficient dct-domain blind measurement and reduction of blocking artifacts. *Circuits and Systems for Video Technology, IEEE Transactions on*, 12(12):1139–1149, 2002.

[15] Zhou Wang, Hamid R Sheikh, and Alan C Bovik. No-reference perceptual quality assessment of jpeg compressed images. In *Image Processing. 2002. Proceedings. 2002 International Conference on*, volume 1, pages I–477. IEEE, 2002.

[16] Lydia Meesters and Jean-Bernard Martens. A single-ended blockiness measure for jpeg-coded images. *Signal Processing*, 82(3):369–387, 2002.

[17] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: application to jpeg2000. *Signal Processing: Image Communication*, 19(2):163–172, 2004.

[18] Anush K Moorthy and Alan C Bovik. A two-step framework for constructing blind image quality indices. *Signal Processing Letters, IEEE*, 17(5):513–516, 2010.

[19] Michele A Saad, Alan C Bovik, and Christophe Charrier. A dct statistics-based blind image quality index. *Signal Processing Letters, IEEE*, 17(6):583–586, 2010.

[20] Anush Krishna Moorthy and Alan Conrad Bovik. Blind image quality assessment: From natural scene statistics to perceptual quality. *Image Processing, IEEE Transactions on*, 20(12):3350–3364, 2011.

[21] Michele A Saad, Alan C Bovik, and Christophe Charrier. Dct statistics model-based blind image quality assessment. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 3093–3096. IEEE, 2011.

[22] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Learning a blind measure of perceptual image quality. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 305–312. IEEE, 2011.

[23] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *Image Processing, IEEE Transactions on*, 21(12):4695–4708, 2012.

[24] Antonio Torralba and Aude Oliva. Statistics of natural image categories. *Network: computation in neural systems*, 14(3):391–412, 2003.

[25] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12):2379–2394, 1987.

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

[27] Eero P Simoncelli and Bruno A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–1216, 2001.

[28] Cem Tarhan. Real time single frame superresolution. Master's thesis, Middle East Technical University (METU), Turkey, 2014.

[29] Kwang In Kim and Younghee Kwon. Example-based learning for single-image super-resolution. In *Pattern Recognition*, pages 456–465. Springer, 2008.

[30] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical imaging and vision*, 20(1-2):89–97, 2004.

[31] K-W Hung and W-C Siu. Fast image interpolation using the bilateral filter. *Image Processing, IET*, 6(7):877–890, 2012.

[32] Stanley Osher and Leonid I Rudin. Feature-oriented image enhancement using shock filters. *SIAM Journal on Numerical Analysis*, 27(4):919–940, 1990.

[33] Eero P Simoncelli and William T Freeman. The steerable pyramid: A flexible architecture for multi-scale derivative computation. In *Image Processing, International Conference on*, volume 3, pages 3444–3444. IEEE Computer Society, 1995.

[34] Anush K Moorthy and Alan C Bovik. Statistics of natural image distortions. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 962–965. IEEE, 2010.

[35] Anuj Srivastava, Ann B Lee, Eero P Simoncelli, and S-C Zhu. On advances in statistical modeling of natural images. *Journal of mathematical imaging and vision*, 18(1):17–33, 2003.

[36] Peng Ye and David Doermann. No-reference image quality assessment using visual codebooks. *Image Processing, IEEE Transactions on*, 21(7):3129–3138, 2012.

[37] Yi Zhang and Damon M Chandler. An algorithm for no-reference image quality assessment based on log-derivative statistics of natural scenes. In *IS&T/SPIE Electronic Imaging*, pages 86530J–86530J. International Society for Optics and Photonics, 2013.

[38] Amy R Reibman, Robert M Bell, and Sharon Gray. Quality assessment for super-resolution image enhancement. In *Image Processing, 2006 IEEE International Conference on*, pages 2017–2020. IEEE, 2006.

[39] Zianou Ahmed Seghir, Fella Hachouf, and Frederic Morain-Nicolier. Blind image quality metric for blurry and noisy image. In *Image Information Processing (ICIIP), 2013 IEEE Second International Conference on*, pages 193–197. IEEE, 2013.

[40] Zhou Wang and Alan C Bovik. Reduced-and no-reference image quality assessment. *Signal Processing Magazine, IEEE*, 28(6):29–40, 2011.

[41] Zhou Wang. Applications of objective image quality assessment methods [applications corner]. *Signal Processing Magazine, IEEE*, 28(6):137–142, 2011.

[42] Dong Zhang, Yong Ding, and Ning Zheng. Nature scene statistics approach based on ica for no-reference image quality assessment. *Procedia Engineering*, 29:3589–3593, 2012.