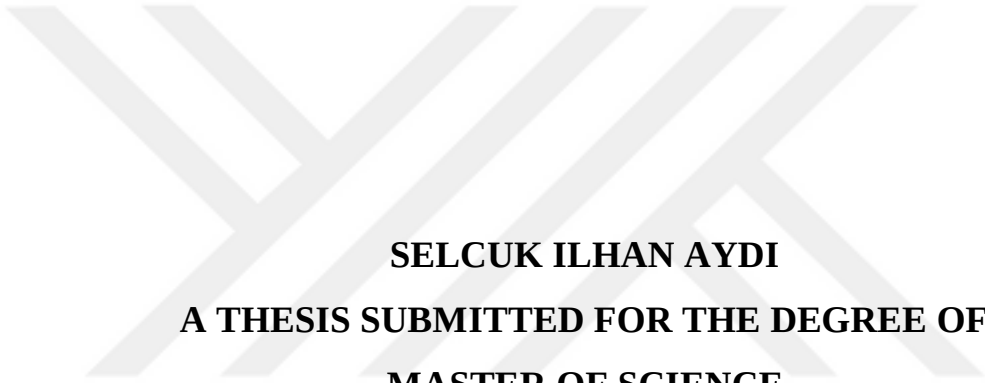


T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

FULL REFERENCE IMAGE QUALITY ASSESSMENT



SELCUK ILHAN AYDI
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING

GEBZE
2022

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

**FULL REFERENCE IMAGE QUALITY
ASSESSMENT**

SELCUK ILHAN AYDI
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF COMPUTER ENGINEERING

THESIS SUPERVISOR
PROF. DR. ERCHAN APTOULA

GEBZE
2022

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TAM REFERANS GÖRÜNTÜ KALİTE
DEĞERLENDİRMESİ

SELÇUK İLHAN AYDI
YÜKSEK LİSANS TEZİ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

DANIŞMANI
PROF. DR. ERCHAN APTOULA

GEBZE
2022



YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 23/06/2022 tarih ve 2022/31 sayılı kararıyla oluşturulan jüri tarafından 30/06/2022 tarihinde tez savunma sınavı yapılan Selçuk İlhan AYDI'nın tez çalışması Bilgisayar Mühendisliği Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE
(TEZ DANIŞMANI) : Prof. Dr. Erchan Aptoula

ÜYE : Prof. Dr. Yusuf Sinan Akgül

ÜYE : Assoc. Prof. Alp Ertürk

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
...../...../..... tarih ve/..... sayılı kararı.

İMZA/MÜHÜR

SUMMARY

In this thesis, a novel full reference image quality assessment method based on sparse neighbor significance is proposed. The method consists of two separate pipelines running in parallel. In the first pipeline, a set of non-overlapping blocks are obtained from the input image, and a central block and four neighboring blocks are chosen to be used in the sparse coding to reconstruct the central block. The neighboring blocks are used as dictionaries and the central block is devised as the signal of interest. The obtained sparse vectors are then utilized to create the sparse significance maps, which are used in the similarity measurement between the reference and the distorted image. In the second pipeline, the gradient information is employed as an auxiliary feature to enhance the overall performance of the proposed method. In this thesis, to be able to determine the performance of the proposed method, several experiments are conducted: the proposed method is compared against the state-of-the-art shallow and deep methods on four image quality assessment datasets. According to the results from the experiments, the proposed method possesses a strong correlation with the Human Visual System and shows a competitive performance with respect to its counterparts.

Key Words: Image Quality Assessment, Sparse Coding, Human Visual System.

ÖZET

Bu tezde, seyrek komşu önemine dayalı yeni bir tam referans görüntü kalitesi değerlendirme yöntemi önerilmiştir. Yöntem esas olarak paralel çalışan iki ayrı işlem hattından oluşur. İlk işlem hattında, giriş görüntüsünden bir dizi örtüşmeyen blok elde edilir ve merkezi bloğu yeniden oluşturmak için seyrek kodlamada kullanılmak üzere bir merkezi blok ve dört komşu blok seçilir. Komşu bloklar sözlük olarak kullanılır ve merkezi blok ilgili sinyal olarak tasarlanır. Elde edilen seyrek vektörler daha sonra referans ve çarpık görüntü arasındaki benzerlik ölçümünde kullanılan seyrek önem haritalarını oluşturmak için kullanılır. İkinci işlem hattında, gradyan bilgisi, önerilen yöntemin genel performansını artırmak için yardımcı bir özellik olarak kullanılır. Bu tezde, önerilen yöntemin performansını belirleyebilmek için bir takım deneyler yapılmıştır. Önerilen yöntem, dört görüntü kalitesi değerlendirme veri kümesi üzerinde en son teknoloji sıg ve derin yöntemlerle karşılaştırılmıştır. Deneylerden elde edilen sonuçlara göre, önerilen yöntem İnsan Görme Sistemi ile güçlü bir korelasyona sahiptir ve benzerleri ile rekabetçi bir performans göstermektedir.

Anahtar Kelimeler: Görüntü Kalitesi Değerlendirmesi, Seyrek Kodlama, İnsan Görme Sistemi.

ACKNOWLEDGEMENTS

I would like to express my deep and sincere gratitude to my supervisor, Prof. Dr. Erchan APTOULA, who not only shared his profound scientific knowledge with me but also taught me great lessons about life. His support, suggestions, and encouragement gave me the drive and will to complete this work.

I am grateful to my parents for their love and support.



TABLE of CONTENTS

	<u>Page</u>
SUMMARY	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
TABLE of CONTENTS	viii
LIST OF ABBREVIATIONS AND ACRONYMS	xii
LIST OF FIGURES	xiii
LIST OF TABLES	xiv
1. INTRODUCTION	1
1.1. Motivation and Objective	1
1.2. Contributions	2
1.3. Thesis Structure	3
2. RELATED WORKS	4
2.1. The Image Quality Assessment Problem	4
2.2. Image Quality Attributes Affecting Quality	5
2.2.1. Human Visual System	5
2.2.2. General Structure of The Human Eye	5
2.2.3. Retina	7
2.2.4. Visual Cortex	9
2.2.5. The Physiological Pathways Hypothesis	10
2.2.6. Visual Perception	11
2.2.7. Adaptation	11
2.2.8. Ambiguous Figures	12
2.2.9. Visual Completion	13
2.2.10. Impossible Objects	13
2.3. Literature Review on IQA Methods	14
2.3.1. Error Sensitivity Based Methods	14
2.3.1.1. Mean Squared Error (MSE)	14
2.3.1.2. Perceptual Image Distortion	15
2.3.1.3. Noise Quality Index	16

2.3.1.4. Visual Signal-to-Noise Ratio	16
2.3.1.5. Most Apparent Distortion	16
2.3.1.6. The Visible Differences Predictor	17
2.3.1.7. Watson’s DCT Model	17
2.3.1.8. Practical Image Quality Index	17
2.3.2. Structural Similarity Based Methods	17
2.3.2.1. Structural Similarity Index	17
2.3.2.2. Feature Similarity Index	18
2.3.2.3. Structural Contrast Quality Index	18
2.3.3. Information-Theoretic	18
2.3.3.1. Visual Information Fidelity	18
2.3.4. Sparse Coding Based Methods	18
2.3.4.1. Sparse Feature Fidelity	19
2.3.4.2. Adaptive Sub-Dictionary Selection Strategy	19
2.3.4.3. Kernel Sparse Coding Based Metric	19
2.3.4.4. Sparseness Significance Ranking Measure	
3. PROPOSED APPROACH	20
3.1. Background	20
3.1.1. Sparse Coding Overview	20
3.1.2. Sparse Coding in IQA	21
3.1.3. Neighboring Blocks	22
3.2. A Novel Full Reference Image Quality Assessment Method Based on Sparse Neighbor Significance	23
3.2.1. Preprocessing	23
3.2.2. Sparse Significance Map	24
3.2.3. Gradient Magnitude	26
3.2.4. Pooling	27
3.2.5. Design Discussion	28
4. EXPERIMENTS	30
4.1. Dataset Details	31
4.1.1. TID2013	31
4.1.2. LIVE	32
4.1.3. TID2008	33

4.1.4. CSIQ	33
4.2. Experimental Setup 1	34
4.2.1. Settings	34
4.2.2. Results	34
4.2.3. Discussion	35
4.3. Experimental Setup 2	35
4.3.1. Settings	36
4.3.2. Results	37
4.3.3. Discussion	38
4.4. Experimental Setup 3	38
4.4.1. Settings	38
4.4.2. Results	38
4.4.3. Discussion	38
4.5. Experimental Setup 4	39
4.5.1. Settings	39
4.5.2. Results	39
4.5.3. Discussion	40
4.6. Experimental Setup 5	41
4.6.1. Settings	41
4.6.2. Results	41
4.6.3. Discussion	41
4.7. Experimental Setup 6	42
4.7.1. Settings	42
4.7.2. Results	43
4.7.3. Discussion	44
4.8. Experimental Setup 7	44
4.8.1. Settings	44
4.8.2. Results	45
4.8.3. Discussion	45
4.9. Experimental Setup 8	46
4.9.1. Settings	46
4.9.2. Results	46
4.9.3. Discussion	46

5. CONCLUSION	48
REFERENCES	49
BIOGRAPHY	54
APPENDICES	55



LIST OF ABBREVIATIONS AND ACRONYMS

<u>Abbreviations</u>	<u>Explanations</u>
<u>and Acronyms</u>	
CNN	: Convolutional Neural Network
CSF	: Contrast Sensitivity Function
DCT	: Discrete Cosine Transform
DM	: Distortion Measure
HVS	: Human Visual System
ICA	: Independent Component Analysis
IQA	: Image Quality Assessment
IQM	: Image Quality Method
LGN	: Lateral Geniculate Nucleus
MOS	: Mean Opinion Score
NQM	: Noise Quality Measure
SSIQM	: Sparse Significance Image Quality Method
SVM	: Support Vector Machine

LIST OF FIGURES

<u>Figure No:</u>	<u>Page</u>
2.1: Cross-sectional diagram of the human eye.	6
2.2: The human visual system.	7
2.3: The human retina.	9
2.4: Receptive field structure of ganglion cells.	9
2.5: The connection map of the layers in the visual cortex.	10
2.6: Schematic diagram of the visual pathways hypothesis.	12
2.7: Ambiguous figure.	14
2.8: An example of an impossible object.	14
3.1: The block diagram of SSIQM.	24
3.2: Robert's edge detector performance on JPEG2000.	28
4.1: Some images from TID2013.	32

LIST OF TABLES

<u>Table No:</u>	<u>Page</u>
4.1: An overview of the experiments conducted in this chapter.	30
4.2: List of distortion types in the TID2013 dataset.	31
4.3: List of distortion types in the TID2008 dataset.	33
4.4: Comparison of SSIQM vs. seven state-of-the-art IQMs on four datasets with five performance metrics.	35
4.5: Comparison of SSIQM vs. ten state-of-the-art FR/NR CNN Based IQMs on three datasets with two performance metrics.	37
4.6: Comparison of SRCC of CNN-Based Models and SSIQM in cross dataset test.	39
4.7: Performance of SSIQM on individual distortions of TID2013.	40
4.8: Comparison of the SSIQM's completion time vs. seven state-of-the-art IQMs in microseconds.	42
4.9: SSIQM's performance with different features.	43
4.10: SSIQM's performance with varying block sizes.	45
4.11: SSIQM's performance with varying numbers of pipelines in the TID2013 dataset.	47

1. INTRODUCTION

Digital images have been a meaningful tool for representing information and communication in many applications. As the usage of images grows the need for a method that controls and ensures the quality of digital images becomes more important. Digital images suffer from distortions during acquisition, compression, transmission, processing, and reproduction. At this point, quantifying the quality of images and determining the degradation becomes a challenging problem, which is called Image Quality Assessment (IQA).

Images are very important for the Human Visual System (HVS), in the context of processing the stimuli, interpreting, and extracting the necessary information. An intuitive approach to IQA is to score images subjectively based on the HVS. However, subjective IQA is expensive and not practical for real-time applications. Therefore, most real-time applications implement objective IQA as an alternative option. In this approach, an image is processed by an objective IQA method and judged concerning a quality score to be used as a metric.

IQA methods may be classified into three categories: Full Reference IQA (FR-IQA), Reduced Reference IQA (RR-IQA), and No Reference IQA (NR-IQA). In the first model, a distortion-free image is always available while partial information is available in the second model. All methods that do not require any prior knowledge of the distorted image may fall into the third model. In this thesis, only a full-reference image quality assessment is studied.

1.1. Motivation and Objective

Since human beings are the final receiver points in many image-based applications, there is great potential in a wide range of applications. A successful development of an objective quality measure can be utilized in many imaging systems [1].

Firstly, a full reference objective quality measure can be employed to benchmark image processing systems. For instance, a fixed scene motoring application/system may use a full-reference image quality method to adjust itself to the conditions according to the instant quality score of the captured frames. Here, it is assumed that

the reference image is the fixed view captured under good conditions, say a scene in the street and the distorted images are the captured frames at any time.

Secondly, using the network bandwidth efficiently has always been important for transmission, not just recently, specifically in video transmission systems. For example, a central server that is responsible for sending and receiving video on the network can use a quality metric to adjust itself automatically.

Unfortunately, scientists lack knowledge about the HVS and its internal patterns [2]. Therefore, the successful development of an objective quality measure is challenging but necessary. An objective method must be naturally correlated with the HVS otherwise it is not useful in any application. Although there is an ongoing development to increase the knowledge about the HVS, current objective methods cannot predict the quality of a distorted image perfectly [1]. This is because there are many parameters contributing to the quality of the image. Moreover, these parameters compose the general behavior of the HVS on the process of the visual stimuli. As a result, developing a general objective quality method has many gaps to be filled in and, in this thesis, the underlying goal is the develop an effective and efficient objective quality assessment method for digital images based on sparse coding and neighbor blocks.

1.2. Contributions

In this thesis, a novel IQA method is proposed. The sparse vectors are extracted in a new manner. Sparse vectors are adopted from a set of dynamic dictionaries rather than using fixed over-complete orthogonal dictionaries. The dynamic dictionaries are derived from the normalized direct pixel intensities of the neighboring blocks.

Secondly, this thesis studies the contribution of the neighboring blocks to the final perceptual quality of the center block and assumes that this can be revealed using sparse coding. Instead of extracting the features from the reference and distorted images, and comparing them, a novel similarity metric has been developed to better correlate to the HVS. This metric considers the sparse vectors and obtains a similarity score between the reference and distorted patches. This information is used to calculate the final quality score of the distorted image.

1.3. Thesis Structure

This thesis is structured as follows; in the first chapter, a brief overview is given of the IQA problem and the main motivation points are presented. In addition to that, the contributions of the proposed method are given. In the second chapter, more details about the IQA problem, some background information, and a literature overview are given to better understand the details of the proposed method. Moreover, the general structure and the behavior of the HVS are given in detail in the subsections of the second chapter. In the third chapter, the details of the implementation of the proposed method are presented in detail. In the fourth chapter, the experiments conducted in the study are presented. Finally, in the fifth chapter, this study is concluded and further investigations and possible research directions are given.

2. RELATED WORKS

In the last decade, many image quality assessment approaches have been proposed in the state-of-the-art. Each of the methods in the literature utilizes different aspects of the HVS to solve the IQA problem in an efficient and valuable way. In this chapter, a brief overview of the HVS and its internal mechanism is presented. In addition to that, a set of studies related to IQA, advantages, and disadvantages are discussed.

2.1. The Image Quality Assessment Problem

Image quality assessment can be referred to as a process of extracting and evaluating the visually significant attributes of an image [1]. IQA plays an important role in many applications including, image compression, transmitting, processing, and entertainment (e.g., online video games and video streaming platforms require reducing the use of bandwidth). The quality of an image can be determined using two different methods: subjectively and objectively. In the former method, the quality assessment is based on human subjects rating the quality of an image in the sense of perceptual attributes and details. The latter obtains the quality score using computational methods which try to mimic the HVS. Since subjective methods are expensive, objective methods are mainly used in IQA which makes it easier to develop fully-automated applications. [1] divided objective methods into three main categories by the availability of the original image. The list of categories are following:

- Full Reference Methods: The original image can be accessed to be used in the image quality assessment process. The quality of the distorted image and original image are compared using the extracted attributes.
- Reduced Reference Methods: Partial information about the original image is provided. These metrics assess the quality of the distorted image using local partial information extracted from the pair of images.
- No Reference Methods: These metrics try to assess the quality of the distorted image in the absence of the original image.

2.2. Image Quality Attributes Affecting Quality

Many attributes convey [1] important information about the quality of an image. Each of them has a varying weight in the process of quality assessment.

- *Sharpness* points out the amount of information and detail that can be obtained from an image or scene. It can be affected by the camera lens, focal length, distance from the image center, and lighting factors.
- *Noise* can be summarized as random degradations at the pixel level. It can affect the pixel level information including, color and brightness.
- *Contrast* refers to the amount of color/grayscale variation among the image components/features in both analog and digital images. High contrast images show a greater degree of color or grayscale variation.

The above list mentions only a little portion of the attributes contributing to the quality of an image.

2.2.1. Human Visual System

Understanding the HVS and its characteristics is the key to developing more robust IQA metrics correlated with the HVS. Hence, firstly, the details and the internal process in the part of our brain responsible for vision are discussed. Moreover, in this section, a general description of basic anatomy and physiology is given, and a general overview is presented to show what kind of information is encoded and processed in our brain and how the stimuli are assessed.

2.2.2. General Structure of The Human Eye

The main objective of the human eye is to transmit light to the further parts of the brain, which are responsible to convert light to neural activities. This transmission process is based on the physical structure of the human eye, which is approximately spherical with a bulge at the front end. Mainly, the front end of the eyeball consists of fusing functional segments, see Figure 2.1, gathered from [3] including the followings:

- *Cornea*: Located at the very front of the eyeball, responsible to transmit and focus the eye on the incoming light reflected from the objects in the world.
- *Pupil*: Located in the middle of the cornea and the iris, also called the dark center, and changes its size according to the amount of incoming light.
- *Iris*: Located after the pupil, responsible to regulate the amount of light striking the eye by managing the pupil position and its size as well.
- *Lens*: Located after the iris, has a transparent structure, responsible to focus the light rays onto the retina.

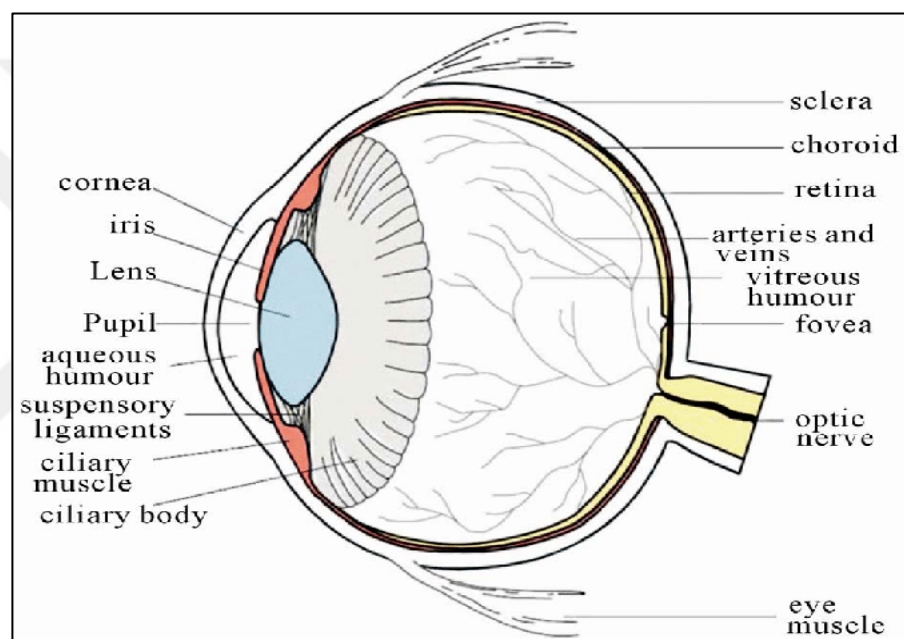


Figure 2.1: Cross-sectional diagram of the human eye.

Extraocular muscles, including rectus, medial, lateral, recti, superior, ciliary, and oblique govern the movement ability of the eyeball. This ability of movement provides the observer to control or adjust the visual perspective and follow the moving objects. The journey of the light rays begins in the very front, at the specific part of the cornea, called aqueous humor, filled in with a water-like fluid similar to plasma, see Figure 2.2, gathered from [4]. In the next stage, the light rays are transmitted to the pupil and iris. Then, the lens receives the lights to focus them on the correct path onto the vitreous humor. In the final stage, the retina receives the photons and converts them into neural activity for further processing by the specific part of the brain, the visual

cortex at the back of the head, generally responsible to process and reduce the redundancy in the visual information.

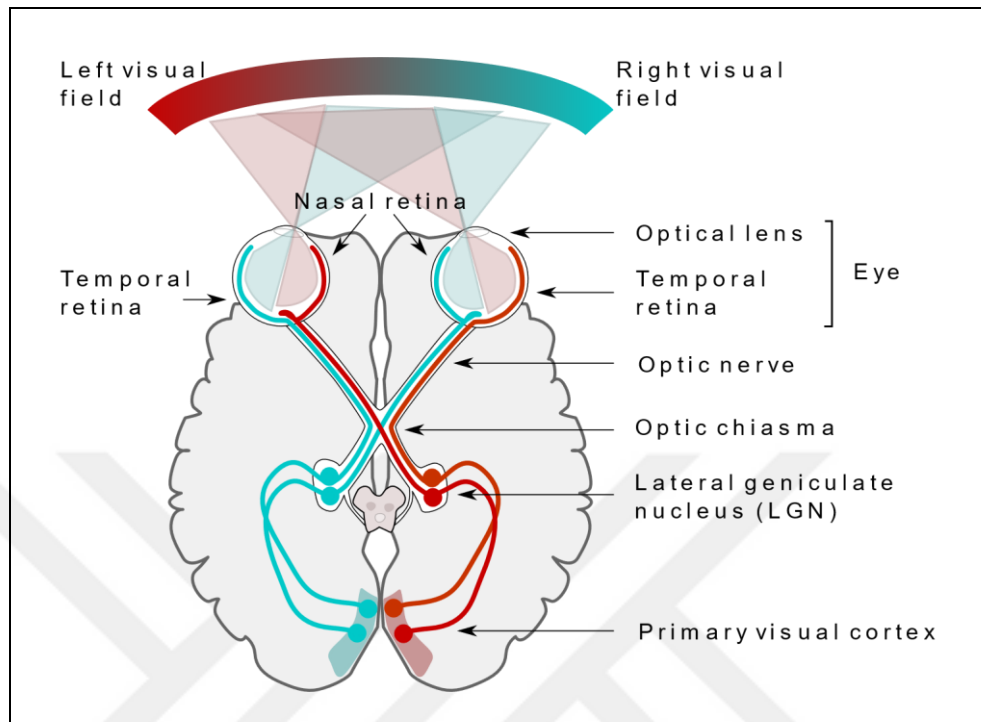


Figure 2.2: The human visual system.

2.2.3. Retina

The output of the optical process has to be carried out by the retinal processing to translate the incoming light into impulse signals so that the further parts of the brain can process and extract the information. The retina is the layer of photoreceptors that are activated by light at different wavelengths, and sample the retinal image projected onto it for further processing. The retina has two types of photoreceptors to obtain light namely rods and cones. The main difference between these cells is that rods activate at low light conditions while cones activate at higher light conditions for vision processing.

There are three distinct classes of cones according to their wavelength: S-cones, M-cones, and L-cones, which tend to be responsive to short-wavelength, medium-wavelength, and longer-wavelength respectively. Cones are mostly found in the fovea, a part of the retina, located at the center. Cones are generally responsible for color information in the incoming light rays at different wavelengths, causing activation of

the corresponding class of cones. Rods contain rhodopsin pigment to be sensitive to light at night vision conditions and have approximately the same wavelength hence does not encode any color information.

The signals that are processed and encoded to neural signals distinctly are postponed to another stage, where the interconnecting neurons (horizontal, bipolar, amacrine, and ganglion cells) are activated and process the incoming signal, see Figure 2.3, gathered from [5]. The ganglion cells are densely located in a specific area, called the ganglion cell layer of the retina of the eye. The ganglion cells receive the encoded visual information from the photoreceptors by intermediate neurons including bipolar cells and amacrine cells. There are two main types of ganglion cells, which vary in their size and responses, see Figure 2.4: off-center cells and on-center cells. The off-center cells are activated by the light rays projected on the nearby region with a reduced response while the on-center cells are activated by the light rays touching the ganglion cells with an enhanced response. In general, ganglion cells have long axons extending to the brain. These axons form specific structures, called optic nerve, optic chiasm, and optic tract. The primary job of these axons is to transmit the encoded and processed information from the optic nerves to the other parts of the brain. The transmission of the encoded information to the cortex is completely crossed – the signals obtained from the left visual field goes to the right-hand side of the brain while the signals obtained from the right visual field go to the left-hand side of the brain. The primary visual cortex (V1) is mainly responsible to reduce the redundancy of information in the incoming signal and process the stimuli to extract relatively higher-level information. The V1 cortex receives this signal through the lateral geniculate nucleus (LGN) and transmits the signal to the upper-level visual cortexes for high-level information extraction.

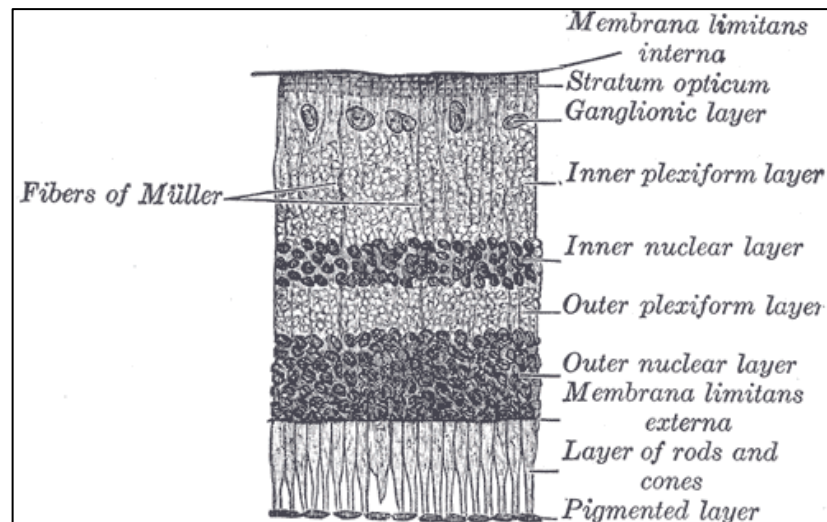


Figure 2.3: The human retina.

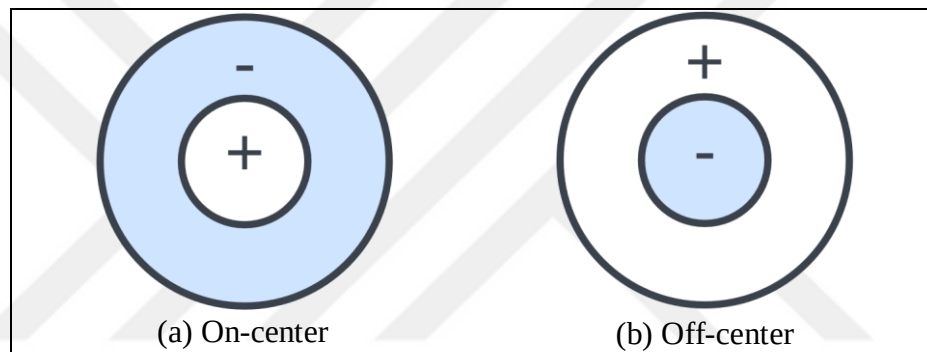


Figure 2.4: Receptive field structure of ganglion cells.

2.2.4. Visual Cortex

The visual information packed and encoded in the retina is now ready to be merged and transmitted to the visual cortex. The information from the left and right sides is merged in LGN. The visual cortex receives this encoded visual signal from the V1, located at the center around the calcarine sulcus. The visual cortex has five distinct classes of areas (V1-V5) concerning their functions and structures. The primary function of the visual cortex is extracting high-level information including spatial location, frequency, orientation, segmenting, and integration of the visual stimuli. The visual cortex has mainly two types of cells: simple and complex cells. The simple cells are specialized for some types of components in the visual scene such as orientation and lines. The simple cells are densely found in the V1 cortex. The complex cells roughly respond to the same stimuli and visual information as simple cells but they do

not represent a single receptive field rather they respond to the summation of other responses of the simple cells lying in the other parts of the brain. There are five well-known visual cortex layers, including V2, V3, V4, MT/V5, and V6 divided according to their functions and structures. Each layer is more complex than the prior layer and extracts more complex information and visual components from the visual stimuli. The connections between the visual cortexes are illustrated in Figure 2.5 gathered from [1]. For example, the V3 layer is responsive to the direction and orientation information in the scene while the V4 extracts high-level color information, the V5 is responsible for motion and moving objects, and the V6 is for depth information.

The visual system of human beings is complex and scientists have started understanding the underlying functions and structures, which lead the authors to develop more robust IQA methods correlated to the HVS. In general, these IQA metrics have the goal of mimicking the visual processing performed in the HVS and producing more compact frameworks.

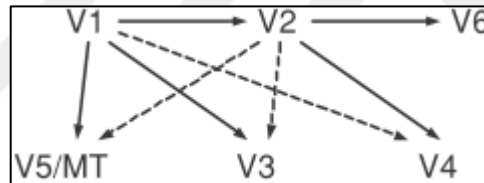


Figure 2.5: The connection map of the layers in the visual cortex.

2.2.5. The Physiological Pathways Hypothesis

There has been a huge amount of ongoing research to find out a relation between the anatomical structure and the physiological function of the HVS. The general assumption is that the different visual fields of the brain process different visual properties. For example, the color, depth, motion, and shape properties of a visual stimulus are processed in distinct regions in the brain. [2] claims that these four types of visual properties are processed in separated subregions of visual cortex areas such as V1 and V2 as demonstrated in Figure 2.6.

The information processing in V1 and V2 are forwarded to the higher layers in different pathways. For example, movement and depth information are transmitted to area V5. On the other hand, the V6 cortex area is mostly responsible to process the color information. The ventral ‘what’ and ‘where’ systems receive this highly complex

information of form and color, and depth and motion, respectively. The ‘what’ system uses this information for object recognition and cognition while the ‘where’ system uses its information for object localization and prediction.

2.2.6. Visual Perception

Optical information carries important aspects and knowledge about environmental objects and events. This type of information must be extracted from the light projected on the retina. This process of extracting information addresses the knowledge about the objects and events. Hence, it can be said that visual perception includes the reinterpretation of the signals made by the sensory responses. These responses are the product of the visual stimuli processed and transformed by the retina and interconnecting neurons. Moreover, the extraction of meaningful information from the visual scene is not only based on cognitive visual processing but also includes the experiences of the observer.

2.2.7. Adaptation

Visual perception has a dynamic structure, and changes over time, depending on the current scene. The human eye adapts itself under low or high light conditions to extract visual information including color, objects, orientation, size, motion, etc. Different visual experience affects the subjective judgment of distorted images in terms of visual quality by the HVS. Hence, an objective metric must consider the light adaptation property of the HVS and include the light information in the scene which is being processed.

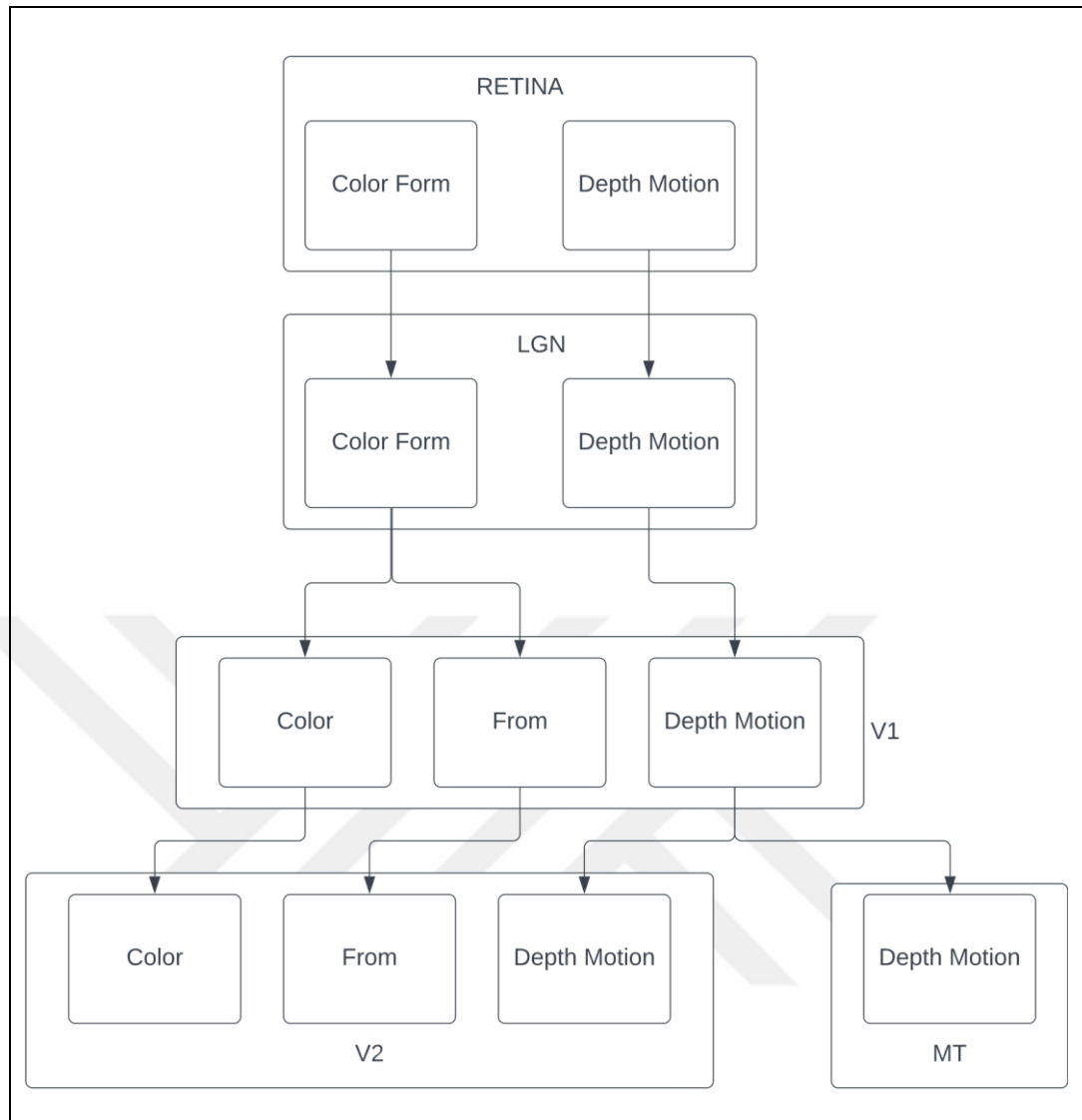


Figure 2.6: Schematic diagram of the visual pathways hypothesis.

2.2.8. Ambiguous Figures

Visual information is transformed into three-dimensional space, hence it is an interpretive process. As a result, the two-dimensional visual information or the components that compose the visual stimuli are the interpretations of our brain rather than certain registered objects. Based on the historical experience of each brain, an image might be interpreted distinctly. A good example of this phenomenon is demonstrated in Figure 2.7. What do you see, a rabbit or a duck or something else? The HVS can only perceive one object at a time. This is mainly the fact that ambiguity is produced because of the illusory contour, perceived by the HVS without any gradient information. As a result, the perception of the mechanism of the HVS

constructs the missing parts by adding high-level visual components in the further parts of the brain after the early perception phase.

2.2.9. Visual Completion

The HVS has a constructive model [2] rather than a sensory stimulation on objects perceived from the outworld and interpreted in the internal mechanism of the visual region of the brain. Thus, the perception of an object or a scene might be illusory and ambiguous. By following this, perceptual models must produce reasonable interpretations of the detected visual scene.

It is a well-known fact that visual perception involves the construction of the hidden surfaces of the visual models. This means our brain completes the missing parts of the objects based on the experience it has. These missing parts are generally not visible from the viewpoint of the observer. This process is performed without any additional effort and it happens whenever we observe or perceive a visual scene.

2.2.10. Impossible Objects

An impossible object can be defined as a two-dimensional drawing that triggers the process of completion of the missing parts in the brain. These objects give a fake view sense of a three-dimensional object, for example, see Figure 2.8. This behavior of the HVS supports the fact that our brain actively constructs environmental objects rather than simply registering the actual view. It can be concluded that perceiving the impossible objects initially starts with the perception of some local features and only later fit the outputs of these results into a global framework.

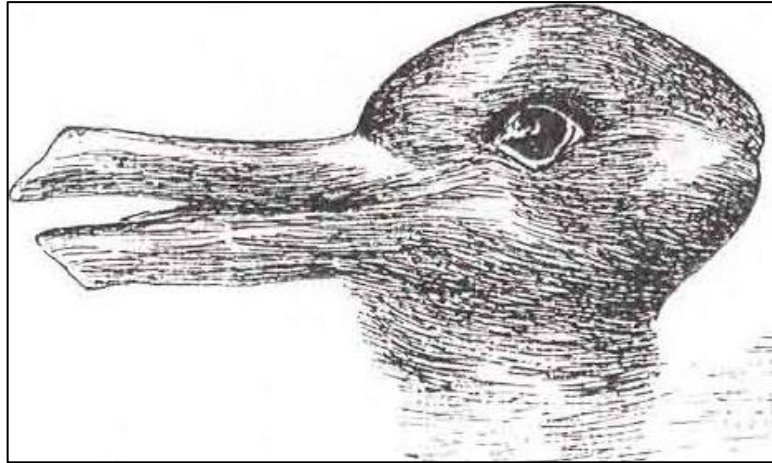


Figure 2.7: Ambiguous figure. This figure can be perceived as a duck or a rabbit at a time but not both at the same time.

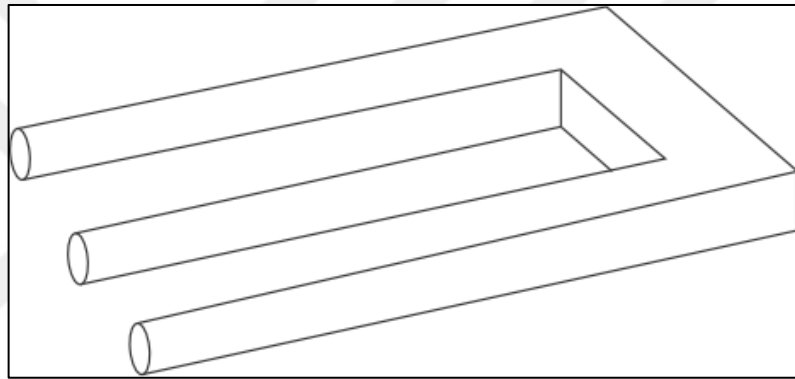


Figure 2.8: An example of an impossible object.

2.3. Literature Review on IQA Methods

There have been extensively published methods, aiming to solve the IQA problem from different perspectives and using the various phenomena of the HVS. In this section, a brief review of the state-of-the-art statistical methods, dividing them into four main classes based on their categories, namely error sensitivity, structural similarity, sparse coding based, and information-theoretic is given.

2.3.1. Error Sensitivity Based Methods

2.3.1.1. Mean Squared Error (MSE)

The simplest metric might be considered as Mean Squared Error in image quality assessment. Its mathematical formulation and desirable features make it more popular.

Assume that the reference image of size $M \times N$ with the perfect quality is donated by $I^r \in \mathbb{R}^{M \times N}$, and the corresponding distorted image of size $M \times N$ is represented by $I^d \in \mathbb{R}^{M \times N}$, and $x \in \{0, 1, \dots, M-1\}$ and $y \in \{0, 1, \dots, N-1\}$ are the spatial coordinates. In this case, MSE can be calculated as

$$MSE(I^r, I^d) = \frac{1}{M \times N} \sum_{y=0}^{N-1} \sum_{x=0}^{M-1} (I^r(x, y) - I^d(x, y))^2 \quad (2.1)$$

Even though MSE has a relatively good correlation with perceived image quality when it is used in the CIELAB color space it does not consider the known features of HVS [6]. As a result, it has been more important to study more compact and meaningful methods/approaches to measuring perceived image quality in the way of addressing the features of HVS including texture masking, contrast sensitivity function, and structural distortion process and its behavior in the visual cortex.

2.3.1.2. Perceptual Image Distortion

The model [7] adopts the steerable pyramid transform [8] to decompose the image into its spatial frequency and orientation components so that it can determine the visible distortion persists in the image by considering the HVS's sensitivity to the different level of frequencies.

2.3.1.3. Noise Quality Index

Based on the concept of the psycho-visual effects of filtering and noise injection are separate matters, two image degradation measures are developed: distortion measure (DM) and noise quality measure (NQM) [9].

The degradation measure is computed in three stages. First, the frequency distortion is determined. Second, the deviation of the frequency domain distortion is obtained. In the final step, the deviations in the frequency distortion are weighted using a low-pass contrast sensitivity function.

The noise quality measure is computed using a contrast pyramid that calculates the contrast at each pixel and spatial frequency. Finally, the NQM is formed by calculating the Signal-to-Noise Ratio (SNR) between the degraded and restored image that is already obtained in the process of calculating the DM.

2.3.1.4. Visual Signal-to-Noise Ratio

Two different thresholds are used, including near-threshold and suprathreshold, as reported in [10]–[13], in the design of Visual Signal-to-Noise Ratio (VSNR). VSNR mainly proposes a wavelet-based IQA metric to quantify the visual fidelity of the preprocessed distorted image.

VSNR operates a two-stage approach using low-level and mid-level properties of the HVS including contrast sensitivity and visual masking. In the first stage, the visibility of the distortions is determined concerning a visible threshold value. If the distortions are near-threshold then the distorted image is considered as having perfect fidelity with the reference image. Otherwise, the second stage is executed, where the low-level and mid-level properties of the HVS are together integrated as a measure of structural degradation.

2.3.1.5. Most Apparent Distortion

Most Apparent Distortion (MAD) [14] is based on the assumption that the HVS may have a varying sensitivity function which represents the different behavior to the different degrees of distortion in image blocks. The main idea is to develop a quality metric that supports the idea of HVS is more sensitive to the distorted components in relatively more quality images, whereas the HVS seeks the main scene content in low-quality images. Hence, MAD proposes a metric, consisting of two main models simulating the underlying assumption of the HVS.

2.3.1.6. The Visible Differences Predictor

The Visible Differences Predictor (VDP) [15] is designed to produce a map that represents the probability of difference between two images. This map (probability-of-detection map) is then used to obtain an overall quality score for the given distorted image.

The items that are in the probability map, are obtained using the following equation:

$$P(c) = 1 - e^{-(c/a)^\beta} \quad (2.2)$$

where $P(c)$ is the probability of detecting a signal of contrast $c \in \mathbb{R}$. The threshold is given by the constant $a \in \mathbb{R}^*$ and $\beta \in \mathbb{R}$ describes the slope of the psychosomatic function.

2.3.1.7. Watson's DCT Model

The main component is the Discrete Cosine Transform (DCT) coefficients of the local blocks, obtained from the reference and distorted image patches. Watson's DCT model [16] is based on the visibility threshold obtained from the DCT sub-blocks. The DC coefficients of DCT sub-blocks are then normalized by the average intensity/luminance. Moreover, all the coefficients are utilized to address the contrast/texture masking property of the HVS.

2.3.1.8. Practical Image Quality Index

A full reference image quality metric, DCTex [17] is developed based on two important properties (contrast sensitivity function and texture masking) of the HVS. The texture masking effect is modeled based on a wavelet decomposition method in which the image is divided into different sub-bands. The transforms in the frequency domain are modeled using the contrast sensitivity function. DCTex has been used successfully in JPEG compression standards.

2.3.2. Structural Similarity Based Methods

2.3.2.1. Structural Similarity Index

The method Structural Similarity Index (SSIM) [18] is composed of three stages to measure the quality of the distorted image. The first stage compares the luminance of the reference and distorted image. The luminance is estimated by the mean intensity of the local patches. In the second stage, the contrast information of the local patches estimated by standard deviation is compared. In the last stage, the structural information is obtained from signal mean and standard deviation and compared. Finally, the three components are then combined yielding a single similarity measure.

2.3.2.2. Feature Similarity Index

The main assumption is that the HVS may be sensitive to the congruent phase in the components of Fourier at varying frequencies relative to the contrast sensitivity function. [19]–[22]. Based on this conclusion Feature Similarity Index (FSIM) [23]

uses phase congruency (PC) as the primary feature. In addition to that, considering that the local contrast information has an impact on the perceptual quality FSIM employs gradient magnitude as the second feature, encoding contrast information.

2.3.2.3. Structural Contrast Quality Index

Structural Contrast Quality Index (SC-IQ) [24] is a recent framework, that solves the IQA problem by considering various structural degradations. In this design, Structural Contrast Quality Index (SCI) [24] is used as the backbone, which can detect perceived distortion for varying distortion types. SCI defines the structural distortion as the ratio of structures and contrast intensity, where the structures are defined as the kurtosis of the magnitudes of DCT AC coefficients, and contrast intensity is defined as the ratio of mean and square of N , where N is the height (= width) of $N \times N$ DCT block. Hence, SC-IQ composes the SCI properties to propose a new metric. SC-IQ also utilizes the color components to detect the impact of the chrominance values on the perceived quality [23], [25].

2.3.3. Information-Theoretic

2.3.3.1. Visual Information Fidelity

Visual Information Fidelity (VIF) [26] is based on the concept of the information theory, treating the IQA problem as an information fidelity issue. The main idea is to develop a model consisting of a communication channel between the distortion and perception process, aiming to measure the quantity of the common information among the reference and distorted image.

VIF extracts the information from the reference and distorted image patches, measuring the amount of information shared. This model is developed on top of a former method Information Fidelity Criterion (IFC) [27].

2.3.4. Sparse Coding Based Methods

2.3.4.1. Sparse Feature Fidelity

Sparse Feature Fidelity (SFF) [28] learns the weighting matrix by machine learning and, the resulting matrix is used to find the sparsest feature map by using Independent Component Analysis (ICA). The calculated sparse features are utilized as

a feature set to be used in quality assessment. Sparse Feature Fidelity (SFF) also combines luminance correlation with sparse features to enhance the correlation with the HVS.

2.3.4.2. Adaptive Sub-Dictionary Selection Strategy

In this method [29], a global dictionary is learned using various natural image patches. Then, the global dictionary is used in the phase of feature extraction from the reference and distorted image blocks. Moreover, the global dictionary is utilized to obtain the sub-dictionaries for each distorted image, to be used to represent them, respectively. Since it is important to reflect the HVS's aspects in the design of any IQA method, several auxiliary features are extracted to be employed, including, contrast, color, and luminance. As the proposed method heavily depends on the global dictionary learned from various image patches, the selection of the dataset has a big impact on the overall performance.

2.3.4.3. Kernel Sparse Coding Based Metric

Kernel Sparse Coding Based Metric (KSCM) [30] aims to solve the IQA problem from a different perspective. It proposes a nonlinear method based on the assumption that the nonlinear information might be useful to better represent the structures of the image. In addition to that, sparse vectors and reconstruction errors are utilized to obtain the predicted objective quality score as the output of the whole method. This method requires various parameters and often parameter tunes to gather a better performance on some specific domains.

2.3.4.4. Sparseness Significance Ranking Measure

Sparseness Significance Ranking Measure (SSRM) [31] is based on weighting the Fourier components in a meaningful and consistent way with the HVS. The weights are obtained from the sparse vectors, called significance values of the corresponding Fourier components. AC and DC coefficients are also considered in the sense of sparse analysis and pooling is performed to produce the normalized objective predicted quality score for the given distorted image.

3. PROPOSED APPROACH

In this chapter, first, a brief introduction to the main concepts regarding the sparse coding and neighboring blocks is given, then a discussion of the main motivation points is presented. At the end of this chapter, the proposed method is presented.

3.1. Background

The two important components of this thesis are sparse coding used as the main framework and neighboring blocks. In this section, a brief review of sparse coding is presented. Also, the usage of sparse coding is discussed in the problem of image quality assessment. Moreover, the neighbor blocks are considered concerning their responsibility and contribution to the overall quality score.

3.1.1. Sparse Coding Overview

There have been many methods and approaches to represent a signal in a compact form. This way provides many benefits, providing an easy way to analyze the signal by its meaningful contributing parts. In the last few decades, sparse coding has become a trending approach to representing a signal in a linear relation [32]. Sparse coding expresses a given signal lying in a high dimensional space, $x \in \mathbb{R}^u$ in a linear relation between sparse vector $a \in \mathbb{R}^v$ and an orthonormal basis dictionary $D \in \mathbb{R}^{u \times v}$ with $v \in \mathbb{Z}^{*+}$ atoms and $u \in \mathbb{Z}^{*+}$ molecules

$$x = Da \tag{3.1}$$

In sparse coding, the objective is to represent the signal of interest x with its simplest possible explanation. More clearly, the signal of interest x is represented by as few as possible atoms obtained from D . Hence, the sparse coding is to be defined as determining the sparsest a that satisfies Equation (3.1). The sparsest representation of a can be found by solving the following minimization problem

$$\min_a \|a\|_0 \text{ subject to } x = Da \quad (3.2)$$

where $\|\cdot\|_0$ counts the nonzero elements or sparsity in the sparse vector a . Unfortunately, solving Equation (3.2) yields many possible solutions, and finding the sparsest solution among those possible solutions is known to be an NP-hard problem. Alternatively, ℓ_1 can be used in place of ℓ_0 to obtain convex relaxation where $\|a\|_1$ is defined as the absolute sum of the values of the items in a . As far as the noise is concerned this alternative solution results in a ℓ_1 -norm regularized minimization problem as

$$\min_a \|x - Da\|_2^2 + \lambda \|a\|_1 \quad (3.3)$$

where $\lambda \in \mathbb{R}^{*+}$ is a regularization parameter to balance the reconstruction error and sparsity. The details of sparse coding can be found in [32]. Each coefficient in the sparse vector corresponds to the weight or significance of each item in the dictionary. Each item in the dictionary is called an atom that will be used to reconstruct the actual signal and their corresponding weight is the sparse vector.

The most famous orthonormal bases D include Fourier Basis and wavelet basis [32]. The choice of D is crucial as it has an impact on a sparse vector a . In most natural signals, the coefficients are zero or close to zero. By following this observation, overcomplete frames are used as orthogonal bases. These overcomplete frames are called dictionaries. Since they have more dimensions than the signal of interest, representing a signal more compactly becomes easy and possible.

3.1.2. Sparse Coding in IQA

The HVS has a complex perceptual process and most of the work is performed in the visual cortex [2]. The internal process of a visual signal in the V1 cortex has the following stages; firstly, a continuous visual stimuli/signal is projected on the retina. Then this stimulus is transmitted to the V1 cortex through LGN to further abstraction and reduction of redundancy. Numerous experimental [33]–[35] shreds of evidence strongly suggest that the underlying mechanism of the neural activities in the brain implies that not only the visual cortex but also other parts of the brain follow a similar sensory functionality against visual stimuli. By following this activity in the brain, a

basic assumption can be concluded that for each visual stimulus a different corresponding neural activity with a different level of importance happens to reduce the redundancy. This internal model of the brain can be modeled [31] using sparse coding such that an over-complete dictionary consisting of atoms with corresponding weights (significance) provides a similar functionality happening in the brain.

From the perspective of sparse coding, a given image can be reconstructed using a set of basis components/functions, each of them indicating a corresponding activation at a different level [36]. As a result, the sparse features can be considered as the actual responses of the neurons against a visual stimulus. The behavior of the neurons in the visual cortex is closely related to the visual perception of the HVS and its rating mechanism implicitly.

3.1.3. Neighboring Blocks

Biological pieces of evidence [37], [38] refer that the neighboring block that may have an impact on the visual quality of the center block that is being processed. Visual agnosia [39], referring to the impairment of visual stimuli in the process of recognition of objects, is an example of such an idea to consider the neighboring blocks in the process of quality assessment. Typically, two types of agnosia are referred to in the literature impacting the IQA problem, including apperceptive agnosia and associative agnosia. Patients suffering from apperceptive agnosia can recognize the local features in the visual scene whereas they fail to perceive the surrounding objects or components that are contributing to the actual structure of the object that is being observed. Therefore, this thesis studies the neighboring blocks as additional information to the idea of sparse coding in the domain of the IQA problem.

Inspired by the above facts, in this thesis, a novel IQM is proposed called Sparse Significance Image Quality Measure (SSIQM) based on the sparse vectors extracted from the reference and distorted image patches of the center and its neighboring blocks.

3.2. A Novel Full Reference Image Quality Assessment Method Based on Sparse Neighbor Significance

In this section, the details of the proposed method for the image quality assessment problem, SSIQM, are described.

SSIQM consists of two pipelines executing in parallel to each other. The first pipeline is responsible for extracting a sparse significance map while the second pipeline extracts gradient magnitude information for each reference I^r and distorted I^d image patch. SSIQM can be summed up as

$$SSIQM = F(I^r, I^d, \beta, \gamma, \theta) \quad (3.4)$$

where $\beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}$ are the weights of the sparse significance map and gradient magnitude similarity scores, respectively, θ is defined as a list of the internal parameters including window size and sparse coding parameters. I^r and I^d are already defined in Equation (2.1). An overview of SSIQM design is illustrated in Figure 3.1.

3.2.1. Preprocessing

In the preprocessing stage, the input images are converted from RGB color space to grayscale using $Y = 0.2125 R + 0.7154 G + 0.0721 B$, where $R, G, B \in \mathbb{R} \cap [0, 1]$. This is because that SSIQM does not take account of chrominance components and does not need to decorrelate the luminance and color components at all. In addition to that, a unity-based min-max normalization is employed to rescale intensity values to scale the range in $[0, 1]$. Finally, the normalized images are then downscaled to a fixed size of 128×128 pixels by performing interpolation for the sake of computational complexity. In this thesis, high-resolution images are not treated specifically rather the downscaling is performed regardless of the dimension of the input image.

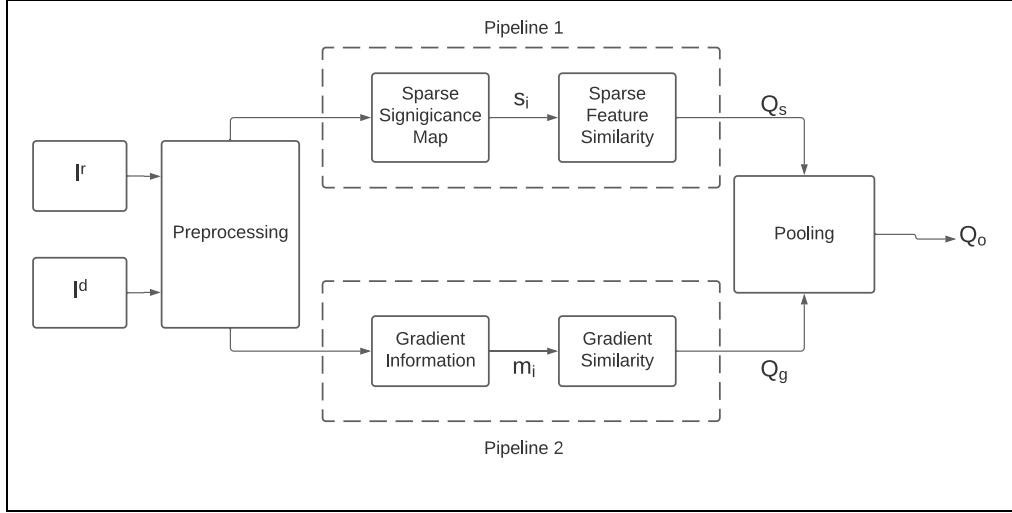


Figure 3.1: The block diagram of SSIQM.

3.2.2. Sparse Significance Map

SSIQM assumes that sparse vectors are considered as a simulation of the internal brain responses. By following this, SSIQM takes the sparse vectors as the primary feature. Specifically, in the first pipeline, the input images with size $M \times N$, are divided into non-overlapping K blocks of size 15×15 pixels, then each block is divided into sub-blocks of size 3×3 pixels, resulting in 25 sub-blocks in total per outer block. Each outer block should be represented by only one center block. Hence, a center block $c_i \in \mathbb{R}^{3 \times 3}, i \in \{1, \dots, K\}$ is selected among 25 sub-blocks. The center block is assumed to be the signal of interest in Equation (3.3). The neighboring blocks $n_{i,j} \in \mathbb{R}^{3 \times 3}, j \in \{1, 2, 3, 4\}$ are considered as a set of dynamic dictionaries that are used in the construction of the signal of interest, the center block. In this study, only four adjacencies are considered to keep the overall design simple and efficient. Hence each neighboring block $n_{i,j}$ is used as D in Equation (3.3). Since the center block is reconstructed by the neighboring dictionaries, a significance map, $S \in \mathbb{R}^{3 \times 3}$ is created for each pair of $(c_i, n_{i,j})$, the center block and its j -th neighboring block. The sparse significance maps, $S_{i_r} \in \mathbb{R}^{3 \times 3}, S_{i_d} \in \mathbb{R}^{3 \times 3}$ of the reference and distorted center blocks c_{i_r} and c_{i_d} , respectively, can be obtained as

$$S_{i_r} = (n_{i,j})_r^T c_{i_r} \quad (3.5)$$

$$S_{i_d} = (n_{i,j})_d^T c_{i_d} \quad (3.6)$$

where $(n_{i,j})_r$ and $(n_{i,j})_d$ are the reference and distorted $j - th$ neighboring blocks of the corresponding c_{i_r} and c_{i_d} . For the sake of simplicity, the significance maps for each pair of the center block and its neighboring blocks are altered to concatenate them horizontally, resulting in $\hat{S}_{i_r} \in \mathbb{R}^{6 \times 3}$ and $\hat{S}_{i_d} \in \mathbb{R}^{6 \times 3}$, the reference and distorted sparse significance maps for $i - th$ center block, respectively. The resulting four significance maps are then forwarded to the next step where a local similarity function $SIM(\rho, \varrho) \in \mathbb{R}$, and $\rho, \varrho \in \mathbb{R}^{6 \times 3}$ measures the similarity between the reference and distorted significance maps. Here, the similarity measure $SIM(\rho, \varrho)$ should meet the following conditions:

- Symmetry: $SIM(\rho, \varrho) = SIM(\varrho, \rho)$.
- Boundedness: $SIM(\rho, \varrho) \leq 1$.
- Unique maximum: $SIM(\rho, \varrho) = 1$ if and only if $\rho = \varrho$.

As stated in [18], the structural similarity measure meets the above requirements and is used in the SSIM index successfully. Hence, in this thesis, the structural similarity measure is adapted to measure the similarity between the sparse significance maps as the following

$$SIM(\hat{S}_{i_r}, \hat{S}_{i_d}) = \frac{2\hat{S}_{i_r}\hat{S}_{i_d} + C}{\hat{S}_{i_r}^2 + \hat{S}_{i_d}^2 + C} \quad (3.7)$$

where the constant $C \in \mathbb{R}^{*+}$ is added to avoid instability when the square sum of the significance maps is very close to zero. The global and normalized similarity measure $Q_s \in \mathbb{R}$ between the reference and distorted significance maps can be obtained using Equation (3.7) as

$$Q_s = \frac{1}{K} \sum_{k=1}^K SIM(\hat{S}_{i_r}, \hat{S}_{i_d}) \quad (3.8)$$

As the final output of the first pipeline, Q_s is then forwarded to the pooling stage where the final quality score is calculated. The details of the pooling process are explained in Section 3.2.4. Please notice that in Equation (3.7), only the local similarity measure of the sparse significance maps of the reference and distorted patches consisting of 15×15 pixels of blocks is calculated. Hence in Equation (3.8), the global similarity score for the given reference and the distorted image is calculated.

3.2.3. Gradient Magnitude

It is a well-known fact that the HVS is highly sensitive to image local structures [40], [41]. Capturing these local structures and determining the degradations in these fields is an important and effective pattern in the IQA problem. Most of the image distortions including, noise corruption, blur effect, and compression artifacts, result in visible structural changes in the gradient domain.

This importance leads this study to utilize the gradient information and integrate it with the SSIQM design. In the second pipeline, the gradient magnitude information is extracted from the reference and distorted images. The magnitude P can be obtained as the following

$$P = \sqrt{g_x^2 + g_y^2} \quad (3.9)$$

where $g_x \in \mathbb{R}$ and $g_y \in \mathbb{R}$ are the gradients in both axes. Since Robert's cross-edge detector is relatively more sensitive to noise it is deployed as the gradient extraction tool because it is desirable to keep the noise information that might be important in the similarity phase where the local quality is calculated. The produced gradient magnitude map $m \in \mathbb{R}^{M \times N}$ by Equation (3.9) is then divided into 15×15 pixels K non-overlapping blocks. As the final step in this pipeline, the mean similarity score $Q_m \in \mathbb{R}$ is calculated between the reference and distorted magnitude maps. Since the structural similarity measure is used successfully in Gradient Magnitude Similarity Deviation (GMSD) [42], in this thesis, it is adapted to measure the gradient map similarity mean as shown below

$$Q_m = \frac{1}{K} \sum_{k=1}^K \frac{2m_{i_r}m_{i_d} + C}{m_{i_r}^2 + m_{i_d}^2 + C} \quad (3.10)$$

where $m_{i_r} \in \mathbb{R}^{15 \times 15}$ and $m_{i_d} \in \mathbb{R}^{15 \times 15}$ are the i – th local blocks of the magnitude maps of the reference and distorted images, respectively, and $C \in \mathbb{R}^*$ is used to avoid instability.

In Figure 3.2, the first row represents the reference and distorted images sequentially. The distortion used in this example is JPEG2000 compression. The second row shows Robert's edge detection results with a magnitude map of the reference image and distorted respectively.

3.2.4. Pooling

The objective quality score is produced in the final stage, called pooling. The obtained Q_s and Q_m are both pooled to provide the final quality score $Q_o \in \mathbb{R}$ against the distorted image as

$$Q_o = \beta Q_s + \gamma Q_m + b \quad (3.11)$$

Each similarity scores Q_s and Q_m have their weights $\beta \in \mathbb{R}$ and $\gamma \in \mathbb{R}$. To avoid a zero-quality score $b \in \mathbb{R}$ is used. Instead of determining the parameters by ad-hoc methods, Ordinary Least Squares Linear Regression is utilized to obtain more effective and efficient results. Let $\{X_1, X_2, \dots, X_l\}$ and $\{y_1, y_2, \dots, y_l\}$ denote the independent and dependent variable sets, respectively. Each $X_i = (Q_s)_i, (Q_m)_i$ represents the 2-D vector consisting of the sparse significance and gradient similarity scores. The objective score (target value) is associated with each i – th image by the notation y_i . As a result, given the training data $(X_1, y_1), \dots, (X_l, y_l)$, the weight vector $w = (\beta, \gamma)$ and b are found by solving the following optimization problem:

$$\min_{w, b} \sum_i^l (y_i - (w^T X_i + b))^2 \quad (3.12)$$



Figure 3.2: Robert's edge detector performance on JPEG2000.

In Equation (3.12), $(w^T X_i + b)$ approximates the training data by minimizing the sum of squared errors. TID2008 [43] is chosen as the training dataset. In this training process, only 15% of this dataset is used to fit the model and the rest is used in the tests.

3.2.5. Design Discussion

In the design of the proposed method, only two pipelines are deployed to predict the objective quality score of the distorted image given as input. As demonstrated in the experiment conducted in Section 4.9, adding more pipelines increases the completion time and decreases the correlation performance of the SSIQM. Moreover, more pipelines mean more output that has to be pooled at the final stage to obtain the predicted score. In addition to that, the features that are chosen for each pipeline have to be consistent with the HVS and of course each other as well.

The gradient information is a consistent feature with the HVS that can be used in the design of an FR-IQA method. Gradient-based methods have generally better performance compared to other methods that rely on different features of the HVS. A good example, [42] has a superior correlation to the HVS against many distortion types. As stated in Section 3.2.3, the HVS is sensitive to the local structures and the deviations in these structures. As a result, the second pipeline is employed for gradient information extraction to increase the proposed method's performance.



4. EXPERIMENTS

This section presents the performance analysis and results gathered from different experiments conducted on well-known datasets, including TID2008 [43], TID2013 [44], CSIQ [14] and LIVE [45]. The details of all these datasets were given in the previous sections.

The experiments conducted in this section consist of two categories: comparing SSIQM with statistical and Convolutional Neural Network (CNN) based methods. In addition, the performance of SSIQM on different distortion types is studied. To be able to measure the pure contribution of the SSIQM an ablation study is accomplished as follows: the SSIQM is executed on the first pipeline while the auxiliary features are executed on the second pipeline. For example, in the first test case, the SSIQM is executed with one pipeline to show the underlying contribution of the proposed method. In the subsequent test cases, gradient, color, contrast, and Gabor features are used as the auxiliary features and executed on the second pipeline.

It is assumed that an ideal IQA method should ensure a correlation with the HVS without any internal layer between objective and subjective scores. That is why any logistic regression method between objective and subjective scores has not been performed in this study. Since this study focuses on structural degradation, contrast and mean-shift distortion types are excluded from all experiments. An overview of the experiments conducted in this chapter is listed in Table 4.1.

Table 4.1: An overview of the experiments conducted in this chapter.

Experimental Setup	Experiment Detail
1	SSIQM is compared with the state-of-the-art shallow methods on TID2013, TID2008, LIVE, and CSIQ datasets.
2	SSIQM is compared with the state-of-the-art CNN-based models on TID2013, LIVE, and CSIQ datasets.
3	A cross-dataset test is conducted with the state-of-the-art CNN-based models and SSIQM on TID2013, LIVE, and CSIQ.
4	SSIQM's performance on individual distortion types (gathered from TID2013) is benchmarked.
5	SSIQM's time complexity.
6	SSIQM with one pipeline.
7	SSIQM with Gabor.
8	SSIQM with Color.
9	SSIQM with Contrast.
10	SSIQM's performance with more than two pipelines.

4.1. Dataset Details

The proposed method in this thesis is validated on four image quality assessment datasets that have reference images, distorted images, and the corresponding subjective scores. Each dataset has its own distortion set and distortion level. In this section, the datasets used in the experiments of benchmarking and performance are presented in detail.

4.1.1. TID2013

The dataset [44] has 25 reference images and 3000 distorted images with 5 levels of distortions. A Mean Opinion Score (MOS) is assigned to each distorted image. This dataset has a broad distortion set such as additive Gaussian noise, JPEG2000 compression, contrast change, comfort noise, etc. Table 4.2 demonstrates all the distortion types involved in the TID2013 dataset. Moreover, in Figure 4.1, several images obtained from the TID2013 dataset are shown with their corresponding scores.

Table 4.2: List of distortion types in the TID2013 dataset.

Number	Type of Distortion
1	Additive Gaussian noise
2	Additive noise in color components is more intensive than additive noise in the luminance component
3	Spatially correlated noise
4	Masked noise
5	High-frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non-eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change
18	Change of color saturation
19	Multiplicative Gaussian noise
20	Comfort noise
21	Lossy compression of noisy images
22	Image color quantization with dither
23	Chromatic aberrations
24	Sparse sampling and reconstruction



Figure 4.1: Some images from TID2013. Reference images, JPEG2000 transmission error distorted and GB distorted images are on the first, second, and third columns, respectively.

4.1.2. LIVE

The LIVE dataset [45] contains 29 reference images with high resolution 24 bits/pixels in RGB color space and the corresponding 779 distorted images at different distortion levels. The set of distortions used in the LIVE dataset includes JPEG compression, JPEG2000 compression, Gaussian blur, White noise, and bit errors in JPEG2000 bitstream.

4.1.3. TID2008

The dataset [43] contains 25 reference images and 1700 distorted images. For each distorted image MOS is assigned. All the reference images are subjected to 17 different distortion types with 4 levels. Table 4.3 lists all the distortion types that exist in TID2008.

4.1.4. CSIQ

The dataset [14] consists of 30 reference images that are subjected to 4 different types of distortions at 4 to 5 levels of distortion. The distortion types include JPEG compression, JPEG-2000 compression, global contrast decrements, additive pink Gaussian noise, and Gaussian blurring. This resulted in 866 distorted versions of the original images.

Table 4.3: List of distortion types in the TID2008 dataset.

Number	Type of Distortion
1	Additive Gaussian noise
2	Additive noise in color components is more intensive than additive noise in the luminance component
3	Spatially correlated noise
4	Masked noise
5	High-frequency noise
6	Impulse noise
7	Quantization noise
8	Gaussian blur
9	Image denoising
10	JPEG compression
11	JPEG2000 compression
12	JPEG transmission errors
13	JPEG2000 transmission errors
14	Non-eccentricity pattern noise
15	Local block-wise distortions of different intensity
16	Mean shift (intensity shift)
17	Contrast change

4.2. Experimental Setup 1

4.2.1. Settings

In the first experiment, Peak Signal-to-Noise Ratio (PSNR), SSIM [18], Multi-Scale SSIM (MS-SSIM) [46], Visual Information Fidelity (VIF) [26], Feature Similarity Index Measure with Color (FSIMc) [23], GMSD [42] and Unique Quality Index (UQI) [47] are compared to the proposed method SSIQM in terms of correlation with the HVS. All these methods are implemented and performed with their default values. To be able to evaluate the performance of each IQM four metrics are employed, including Pearson's linear correlation coefficient (PLCC), Spearman rank-order correlation coefficient (SRCC), Kendall rank-order correlation coefficient (KRCC), and root mean squared error (RMSE) as suggested in [48].

4.2.2. Results

The overall performance of SSIQM is shown in Table 4.4 with seven alternative methods and four metrics on four datasets. The most successful results are highlighted in bold.

4.2.3. Discussion

From Table 4.4, on TID2013, it can be observed that SSIQM has a superior performance over the other full reference IQA methods specifically in terms of prediction accuracy while FSIMc shows a better monotonic performance than SSIQM in this dataset. Moreover, RMSE is higher than SSIQM and PSNR. This is mainly due to the fact that SSIQM does not have an optimized pooling strategy. On the TID2008 dataset, although GMSD and FSIMc have a higher correlation, SSIQM presents a competitive performance in terms of prediction accuracy. On the other hand, SSIQM has a worse PLCC score. The proposed method has competitive performance against SSIM, MS-SSIM, FSIMc, and GMSD on the LIVE dataset. On CSIQ, SSIQM is not at the first rank, however, it exhibits a competitive prediction accuracy and a fair level of PLCC.

Table 4.4: Comparison of SSIQM vs. seven state-of-the-art IQMs on four datasets with five performance metrics.

Method	Criteria	TID2013	TID2008	LIVE	CSIQ
PSNR	SRCC	0.754	0.672	0.819	0.887
	PLCC	0.732	0.641	0.800	0.897
	KRCC	0.546	0.453	0.617	0.685
	RMSE	22.90	21.65	28.19	29.94
SSIM [18]	SRCC	0.728	0.694	0.874	0.890
	PLCC	0.735	0.679	0.713	0.807
	KRCC	0.535	0.500	0.687	0.695
	RMSE	3.700	3.687	0.291	0.322
MS-SSIM [8]	SRCC	0.799	0.769	0.877	0.918
	PLCC	0.770	0.734	0.665	0.826
	KRCC	0.597	0.567	0.687	0.742
	RMSE	3.659	3.634	0.334	0.336
VIFp [26]	SRCC	0.538	0.594	0.869	0.886
	PLCC	0.538	0.523	0.858	0.890
	KRCC	0.432	0.439	0.677	0.790
	RMSE	4.032	4.040	0.187	0.163
FSIMc [23]	SRCC	0.869	0.902	0.919	0.899
	PLCC	0.849	0.845	0.776	0.827
	KRCC	0.686	0.718	0.748	0.709
	RMSE	3.644	3.609	0.356	0.356
GMSD [41]	SRCC	0.810	0.896	0.909	0.947
	PLCC	0.859	0.868	0.868	0.938
	KRCC	0.644	0.714	0.730	0.796
	RMSE	4.502	4.430	0.510	0.095
UQI [47]	SRCC	0.707	0.567	0.786	0.589
	PLCC	0.530	0.418	0.450	0.419
	KRCC	0.505	0.393	0.579	0.420
	RMSE	3.616	3.576	0.441	0.354
SSIQM proposed	SRCC	0.883	0.890	0.914	0.937
	PLCC	0.790	0.794	0.679	0.855
	KRCC	0.702	0.711	0.728	0.772
	RMSE	7.770	7.709	11.05	11.22

4.3. Experimental Setup 2

When deep models are trained on one dataset they can produce good results if the test dataset is the same as the training set that has the same distortion types. Moreover, cross dataset tests show the real generalization capacity of deep models where they are trained on dataset X and evaluated on dataset Y. However, the general challenging problem is the lack of dataset to be used in the training phase which leads to over-fitting in most cases if the number of parameters of the deep model is relatively

big enough. To tackle this problem, most of the papers use augmentation techniques to increase the number of items in a dataset which will help avoid the over-fitting problem. This may provide some performance improvement but local qualities and their importance are not distributed uniformly in most of the distortion types.

4.3.1. Settings

In the second experiment setup, SSIQM is compared against CNN-based methods including, Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning (H-IQA) [49], Blind Image Quality Assessment With Active Inference (AIGOA) [50], Blind predicting similar quality map for image quality assessment (BPSQM) [51], Blind image quality assessment using a deep bilinear convolutional neural network (DB-CNN) [52], Fully deep blind image quality predictor (BIECON) [53], Deep CNN-based blind image quality predictor (DIQA) [54], End-to-end blind image quality prediction with cascaded deep neural network (CaHDC) [55], Deep blind image quality assessment based on multiple instance regression (DSIR-IQA) and DMIR-IQA [56].

In this experiment, only DB-CNN and BIECON are trained individually while other CNN models are directly used with pre-trained settings explained in their corresponding papers published. Because of the lack of training data for the IQA problem, the datasets are split into two distinct: one for training and one for testing.

In the settings of DB-CNN [52], all the distorted images in TID2013, CSIQ, and LIVE, are divided into two splits, 70% of which are used for the training phase and the rest 30% for testing. In this experiment, any nonlinear logistic mapping function is not used to map the predicted scores to the actual scores. All the parameters of this model are initialized with the variables explained in [52]. Training is optimized using Adam [57] with a mini-batch of 64. The learning rate is chosen as 0.10548. In the training phase, all the input images are rescaled to $256 \times 256 \times 3$ and randomly cropped to $224 \times 224 \times 3$ for data augmentation. Since resizing the image in the training has an impact on the accuracy, precision, and recall [58] in this experiment it is used only in the training phase.

In the settings of BIECON [53], TID2008 [43] and CSIQ [14] datasets are used, 80% for training and 20% for testing. Again, any nonlinear logistic mapping function

is not used to map the predicted scores to the actual scores. Adam [57] is used for training optimization. In this experiment, as stated in [53], variable mini-batch-based optimization is adopted. Overfitting is always a problem because of the lack of datasets in the IQA problem that's why image patches are flipped horizontally in the training set. Since flipping vertically may alter the semantic information in the image, it is not employed.

4.3.2. Results

The overall performance of each deep model is illustrated in Table 4.5, showing only SRCC and PLCC scores and comparing them with SSIQM. All the models are trained and validated in the same dataset in each row.

Table 4.5: Comparison of SSIQM vs. ten state-of-the-art FR/NR CNN Based IQMs on three datasets with two performance metrics.

Method	Criteria	TID2013	LIVE	CSIQ
H-IQA [49]	SRCC	0.790	0.982	0.885
	PLCC	0.880	0.982	0.910
AIGQA [50]	SRCC	0.871	0.960	0.885
	PLCC	0.893	0.957	0.952
BPSQM [51]	SRCC	0.862	0.973	0.927
	PLCC	0.885	0.963	0.952
DB-CNN [52]	SRCC	0.816	0.968	0.874
	PLCC	0.865	0.971	0.915
BIECON [53]	SRCC	0.717	0.958	0.946
	PLCC	0.762	0.960	0.959
DIQA [54]	SRCC	0.825	0.975	0.884
	PLCC	0.850	0.977	0.915
CaHDC [55]	SRCC	0.862	0.965	0.903
	PLCC	0.878	0.964	0.914
DSIR-IQA [56]	SRCC	0.782	0.967	0.820
	PLCC	0.816	0.969	0.878
DMIR-IQA [56]	SRCC	0.796	0.967	0.823
	PLCC	0.821	0.971	0.881
SSIQM proposed	SRCC	0.883	0.914	0.937
	PLCC	0.790	0.679	0.855

4.3.3. Discussion

It is possible to notice that SSIQM has a superior prediction accuracy on TID2013 with the best SRCC performance while its PLCC value shows a relatively low correlation. AIGQA seems to have the best performance on PLCC and its SRCC performance is the closest to SSIQM's SRCC performance. On the LIVE dataset, H-IQA shows a very strong correlation on both SRCC and PLCC benchmarks whereas SSIQM has competitive performance on the SRCC benchmark. However, SSIQM's PLCC performance is the worst on the LIVE dataset. BIECON is in the first rank on the CSIQ dataset, having the best SRCC and PLCC values. It can be noticed that SSIQM has a very similar performance result with BIECON on SRCC while SSIQM's performance is relatively worse than BIECON on PLCC. It can be said that still SSIQM has a promising performance on the three datasets used in this experiment.

4.4. Experimental Setup 3

In this experiment, the cross dataset test is conducted to measure the power of SSIQM on LIVE, CSIQ, and TID2013 databases.

4.4.1. Settings

For this experiment, the common distortion types among the three databases are selected including, GB, White Noise (WN), JPEG, and JPEG200, and the methods used include, AIGQA, DIQA, DSIR-IQA, and DMIR-IQA.

4.4.2. Results

Table 4.6 demonstrates the SRCC results of the cross dataset test.

4.4.3. Discussion

It can be observed that SSIQM performs best when deep models are trained on TID2013 and tested with other datasets. DIQA has a better generalization capability compared to other deep models. Training on LIVE and validating on other datasets

gives better scores for DIQA while other models follow a stable graph. On the other hand, SSIQM's performance is the best in the CSIQ dataset compared to others. When the training dataset is CSIQ, deep models have higher prediction accuracy than SSIQM.

Table 4.6: Comparison of SRCC of CNN-Based Models and SSIQM in cross dataset test.

Trained	Tested	AIGQA [50]	DIQA [54]	DSIR-IQA [56]	DMIR-IQA [56]	SSIQM proposed
TID2013	LIVE	0.886	0.904	0.879	0.880	0.914
	CSIQ	0.823	0.877	0.856	0.867	0.937
LIVE	TID2013	0.698	0.922	0.878	0.891	0.883
	CSIQ	0.847	0.915	0.903	0.922	0.937
CSIQ	LIVE	-	0.926	0.936	0.947	0.914
	TID2013	-	0.923	0.906	0.907	0.883

4.5. Experimental Setup 4

Each distortion type has its effect on different aspects of an image that potentially may alter the perception of the scene by the HVS. To determine the performance of SSIQM on a variety of distortions an experiment is conducted.

4.5.1. Settings

For this experiment, the TID2013 dataset is used since it has the largest distortion set. SSIQM is validated on the 24 distortion types individually. SRCC, PLCC, KRCC, and RMSE are used for benchmarking. In this experiment, any specific task (color channel separation and processing) for color-based distortions is not performed. The SSIQM is performed with the default values.

4.5.2. Results

The results of this experiment are shown in Table 4.7. The failed distortion types are in bold.

4.5.3. Discussion

It can be seen that SSIQM has a strong correlation with the HVS in a wide range of distortion sets in TID2013. It performs best in structural and non-color distortions while its performance is not sufficient in mean shift, contrast change, change of color saturation, and local block-wise distortion and the corresponding rows are in bold. Since SSIQM only relies on two extracted features it does not address all kinds of degradation, however, the performance on failed distortions might be enhanced by adding new features which are important for the perception of the HVS. The underlying features might be sensitive to color components, contrast information, and light adaptation.

Table 4.7: Performance of SSIQM on individual distortions of TID2013.

Type	SRCC	PLCC	KRCC	RMSE
Additive Gaussian noise	0.938	0.850	0.759	7.693
Noise in color components	0.876	0.866	0.683	7.260
Spatially correl. Noise	0.920	0.857	0.737	7.697
Masked noise	0.827	0.747	0.632	7.310
High frequency noise	0.928	0.878	0.748	7.626
Impulse noise	0.885	0.795	0.700	8.361
Quantization noise	0.906	0.773	0.721	7.536
Gaussian blur	0.974	0.876	0.862	7.878
Image denoising	0.963	0.890	0.840	7.431
JPEG compression	0.948	0.918	0.793	7.762
JPEG2000 compression	0.960	0.887	0.832	8.023
JPEG transm. Errors	0.887	0.822	0.656	7.67
JPEG2000 transm. Errors	0.889	0.779	0.701	8.005
Non ecc. patt. Noise	0.785	0.701	0.535	7.130
Local block-wise dist	0.613	0.561	0.433	8.687
Mean shift	0.592	0.539	0.409	3.690
Contrast change	0.429	0.575	0.297	5.997
Change of color	0.478	0.305	0.329	8.334
Multipl. Gauss. Noise	0.911	0.842	0.727	7.787
Comfort noise	0.924	0.861	0.755	7.471
Lossy compr. of noisy images	0.941	0.868	0.783	7.742
Image color quant. w. dither	0.921	0.809	0.748	7.672
Chromatic aberrations	0.884	0.826	0.720	7.365
Sparse sampl. and reconstr	0.966	0.874	0.836	7.890

4.6. Experimental Setup 5

The completion time is an important indicator for real-time systems. In this experiment, the SSIQM's time complexity is compared with the state-of-the-art shallow methods.

4.6.1. Settings

PSNR, SSIM [18], MS-SSIM [46], VIF [26], FSIMc [23], GMSD [42] and UQI [47] are compared to the SSIQM in terms of completion time in microseconds. Each method is performed on every four datasets and the average completion time is obtained. Since the datasets do not have the same number of references and distorted images, the metrics are iterated over the four datasets 100 times.

4.6.2. Results

The results for each metric, including the SSIQM, are shown in Table 4.8 as sorted in ascending order according to the completion time.

4.6.3. Discussion

According to Table 4.8, it can be said that the SSIQM has a competitive time complexity overall while PSNR, GMSD, and SSIM have superior performance. This is mainly because of that PSNR consists of a primitive calculation and does not consider the aspects of the HVS. In the case of GSMD and SSIM, the completion time drastically increases as they deal with the gradient, color, and frequency domain components extracted from the images.

Table 4.8: Comparison of the SSIQM's completion time vs. seven state-of-the-art IQMs in microseconds.

Method	Completion Time (μs)
PSNR	18541
GMSD [41]	29858
SSIM [18]	49879
SSIQM proposed	70437
MS-SSIM [46]	209878
VIFp [26]	214200
UQI [47]	230003
FSIMc [23]	298967

4.7. Experimental Setup 6

In this experiment, the pure performance and the contribution of the sparse significance maps are investigated, running only one pipeline that produces the predicted objective quality score against the distorted image. Moreover, the contribution of the Gabor filter, color, and contrast features are benchmarked. Each feature is executed in the second pipeline.

4.7.1. Settings

In the case of one pipeline test case, the sparse significance map is executed and the produced predicted quality score is directly benchmarked without any pooling phase.

In the case of the Gabor filter test case, a second pipeline is added to extract the Gabor features. The kernel that is used in this experiment is constructed with 0.05 frequency, sigma, and theta as 1.

In the case of the color test case, the input image is converted to the CIELAB color space and the luminance component is extracted in the second pipeline. The produced similarity score is pooled at the final stage.

In the case of the contrast test case, the contrast information is extracted and processed in the second pipeline. Pooling is applied to the similarity scores produced by the pipelines at the final stage.

SRCC, PLCC, KRCC, and RMSE performance of the SSIQM with one pipeline is investigated. This experiment is conducted on four datasets including TID2013, TID2008, LIVE2016, and CSIQ.

4.7.2. Results

The results of this experiment are demonstrated in Table 4.9. Each row indicates one test case on different datasets with different criteria. The results are not normalized rather, they are gathered directly from the experiments. The best scores performed obtained from the test cases are in bold.

Table 4.9: SSIQM's performance with different features.

Test Case	Criteria	TID2013	TID2008	LIVE	CSIQ
SSIQM with one pipeline	SRCC	0.871	0.879	0.898	0.901
	PLCC	0.753	0.726	0.646	0.811
	KRCC	0.676	0.692	0.698	0.737
	RMSE	8.083	8.093	11.57	11.64
SSIQM with the Gabor features	SRCC	0.848	0.863	0.896	0.891
	PLCC	0.762	0.765	0.649	0.674
	KRCC	0.660	0.678	0.712	0.707
	RMSE	7.734	7.668	11.07	11.14
SSIQM with contrast	SRCC	0.870	0.888	0.875	0.906
	PLCC	0.755	0.748	0.586	0.776
	KRCC	0.693	0.710	0.682	0.724
	RMSE	8.005	7.994	11.38	11.52
SSIQM with color	SRCC	0.765	0.727	0.826	0.717
	PLCC	0.683	0.630	0.546	0.592
	KRCC	0.572	0.533	0.622	0.535
	RMSE	7.890	7.835	11.28	11.17
SSIQM proposed	SRCC	0.883	0.890	0.914	0.937
	PLCC	0.790	0.794	0.679	0.855
	KRCC	0.702	0.711	0.728	0.772
	RMSE	7.770	7.709	11.05	11.22

4.7.3. Discussion

According to the results obtained from this experiment, the SSIQM with one pipeline still shows a good correlation with the HVS on the TID2013 and the LIVE datasets while it has a competitive performance on the CSIQ dataset. On the other hand, it is obvious that when the SSIQM deploys only one pipeline with sparse significance the performance relatively decreased when it is compared to the original SSIQM with two pipelines.

The impact of the Gabor features extracted by using the Gabor kernels can be seen in the TID2013, TID2008, and the LIVE datasets on PLCC, KRCC, and RMSE scores. It can be observed that the Gabor features do not contribute to the correlation as much as the SSIQM with one pipeline.

Interestingly, the SSIQM that uses contrast information has superior performance on the TID2008 and the CSIQ datasets. Moreover, it has similar results to the SSIQM which has one pipeline. As a result, the contrast information shows a meaningful contribution to the overall.

In the fifth test case, the SSIQM which uses the color information does not show a good performance on any dataset and it causes performance degradation. It can be concluded that the proposed method, in the last test case, has a better performance compared to other variants of the SSIQM with varying numbers of pipelines.

4.8. Experimental Setup 7

To be able to measure the impact of the outer block size on the performance of the SSIQM, varying block sizes are performed. Indeed, the block size has an impact not only on the correlation but also on the complexity. However, in this experiment, the impact on the completion time is not investigated. The TID2013, TID2008, LIVE, and CSIQ datasets are chosen for this experiment.

4.8.1. Settings

In the original design of the SSIQM, a 15×15 outer block size is used as stated previously. Now, in this experiment, the original SSIQM with two pipelines is benchmarked using 9×9 , 12×12 , 18×18 , and 21×21 . All the other parameters

are kept with their default values. SRCC, PLCC, KRCC, and RMSE are used as the benchmark criteria for the test cases.

4.8.2. Results

The results conducted from this experiment are shown in Table 4.10. Each row points to one test case with different criteria on each dataset.

4.8.3. Discussion

From Table 4.10, it can be observed that the block size has a significant impact on the correlation of the SSIQM. In the first test case, it is obvious that the SSIQM shows the worst performance with a 9×9 outer block size. On the other hand, the proposed method shows the best performance in the second test case in the TID2013 and TID2008 datasets. Also, the SSIQM has fewer RMSE values and the best PLCC scores in the LIVE and CSIQ datasets. In the third test case, the performance decreases however, the SSIQM still obtains the best SRCC values in the LIVE and CSIQ datasets. In the last test case, the scores are pretty close to the second and the third case but the time complexity is much more. From the test cases, it can be concluded that the outer block size affects the correlation score and the complexity of the proposed method.

Table 4.10: SSIQM's performance with varying block sizes.

Test Case	Criteria	TID2013	TID2008	LIVE	CSIQ
9×9	SRCC	0.789	0.782	0.835	0.885
	PLCC	0.714	0.752	0.700	0.740
	KRCC	0.599	0.611	0.654	0.693
	RMSE	8.148	7.895	11.89	11.64
12×12	SRCC	0.881	0.872	0.885	0.882
	PLCC	0.740	0.745	0.724	0.768
	KRCC	0.693	0.710	0.689	0.718
	RMSE	11.64	7.953	10.01	11.23
18×18	SRCC	0.860	0.870	0.887	0.887
	PLCC	0.720	0.712	0.604	0.604
	KRCC	0.676	0.689	0.696	0.696
	RMSE	8.066	8.071	11.57	11.57
21×21	SRCC	0.855	0.862	0.886	0.887
	PLCC	0.696	0.685	0.601	0.743
	KRCC	0.667	0.679	0.695	0.697
	RMSE	8.062	8.065	11.58	11.65

4.9. Experimental Setup 8

In this experiment, the SSIQM's performance is measured with more than two pipelines in terms of time complexity and correlation.

4.9.1. Settings

Color information, Gabor feature, gradient, and contrast are used as the main features and deployed to the pipelines in each test case. All the features are extracted and processed in the same way as stated in the previous experiments. In each test case, more pipelines are added consecutively and the performance of the SSIQM with the newly added pipelines is measured in the TID2013 dataset. The time complexity is measured in microseconds.

4.9.2. Results

The results obtained from this experiment are shown in Table 4.11. Each added pipeline is shown with '+' sign.

4.9.3. Discussion

From Table 4.11, it can be observed that the SSIQM has a good performance with one and two pipelines. However, as more pipelines are added the correlation decreases and time complexity increases, affecting the general performance of the SSIQM with more than two pipelines. Moreover, as the time complexity is not efficient with more than two pipelines the SSIQM becomes less attractive for real-time systems thus the usage area of the proposed method gets smaller. Furthermore, pooling all these features and assigning proper weight to each of them is a cumbersome process at the pooling stage.

Table 4.11: SSIQM's performance with varying numbers of pipelines in the TID2013 dataset.

	Sparse Significance	+Gradient	+Contrast	+Color	+Gabor
Completion Time (μs)	40437	70437	99437	112643	251008
SRCC	0.871	0.883	0.753	0.731	0.744
PLCC	0.753	0.790	0.620	0.590	0.610
KRCC	0.676	0.702	0.611	0.582	0.632
RMSE	8.083	7.770	12.410	13.801	10.161

5. CONCLUSION

In this work, an IQA method based on sparse coding coefficients and neighboring blocks, contributing to the overall quality of the center block, is proposed. The sparse vectors are devised to measure the significance of each contributing block. In the calculation of the sparse vectors, a new approach, called dynamic dictionaries, obtained from the neighboring blocks, is utilized in the process of similarity measurement.

In conclusion, it can be established that the proposed method works well with most of the distortion types except mean shift and contrast. To be able to determine the performance of the proposed method, intensive experiments are conducted where SSIQM is first compared with the most successful statistical FR-IQA methods. Furthermore, SSIQM's performance is compared with deep models on different datasets via cross dataset tests. Since each distortion has different effects on perceptual quality SSIQM's performance on each distortion type has been measured. From the experiments, it can be concluded that the proposed method has a good correlation with HVS on many distortion types. Its performance is very competitive with a simple design when its counterparts are considered.

According to the experiments conducted in Sections 4.6, 4.7, 4.8, and 4.9, the proposed method has a competitive time complexity compared to other state-of-the-art shallow methods. Also, it can be concluded that the number of pipelines is an important parameter in the design of the SSIQM, impacting the correlation and completion time. Moreover, another important parameter that has a huge impact on the performance of the SSIQM is the size of the outer blocks. In conclusion, all these parameters have to be chosen carefully and applied to the design of the SSIQM. On the other hand, as the number of parameters increases the complexity of the SSIQM also increases which affect the overall performance of the proposed method.

A further study would be adding more features considering the aspects of HVS to provide a more robust IQA method on all the distortion types and databases. It is also possible to integrate sparse significance maps with modern deep models. Finally, it would make sense to remark that SSIQM shows a well moderate correlation with the HVS and holds the promise to evaluate the images robustly and effectively when it is used in video codecs and applications that consider image quality.

REFERENCES

- [1] Wang Z., Bovik A. C., (2006), “Modern image quality assessment. Synthesis Lectures on Image, Video, and Multimedia Processing”, 2nd Edition, Morgan-Claypool.
- [2] Palmer S. E., (1999), “Vision science: Photons to phenomenology”, 1st Edition, MIT press.
- [3] Mansour R. F., Abdelrahim E. M., Al-Johani A. S., (2013), “Identification of diabetic retinal exudates in digital color images using support vector machine”, Journal of Intelligent Learning Systems and Applications, 5(3), 135-142.
- [4] Web 1, (2015), https://commons.wikimedia.org/wiki/User:Perellonieto#/media/File:Human_visual_pathway.svg, (Accessed: June 10, 2022).
- [5] Gray H., (1878), “Anatomy of the human body”, 20th Edition, Lea-Febiger.
- [6] Rajashekar U., Wang Z., Simoncelli E. P., (2009), “Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions,” 16th IEEE International Conference on Image Processing, 2213-2216, Cairo, Egypt, 7-10 November.
- [7] Teo P. C., Heeger D. J., “Perceptual image distortion”, (1994), 1st International Conference on Image Processing, 982-986, Austin, TX, USA, 13-16 November.
- [8] Simoncelli E. P., Freeman W. T., (1995) “The steerable pyramid: A flexible architecture for multi-scale derivative computation”, International Conference on Image Processing, 444-447, Washington, DC, USA, 23-26 October.
- [9] Damara-Venkata N., Kite T. D., Geisler W. S., Evans B. L., Bovik A. C., (2000), “Image quality assessment based on a degradation model”, IEEE transactions on image processing, 9(4), 636-650.
- [10] Ramos M. G., Hemami S. S., (2001), “Suprathreshold wavelet coefficient quantization in complex stimuli: psychophysical evaluation and analysis”, Journal of the Optical Society of America A, 18(10), 2385-2397.
- [11] Chandler D. M., Hemami S. S., (2003), “Effects of natural images on the detectability of simple and compound wavelet subband quantization

distortions”, Journal of the Optical Society of America A, 20(7), 1164-1180.

- [12] Chandler D. M., Hemami S. S., (2003), “Suprathreshold image compression based on contrast allocation and global precedence”, Human Vision and Electronic Imaging VIII in SPIE Proceedings, 5007, 73-86.
- [13] Chandler D. M., K. H. Lim., Hemami S. S, (2006), “Effects of spatial correlations and global precedence on the visual fidelity of distorted images”, Proceedings of Human Vision and Electronic Imaging XI, 131-145, San Jose, California, United States, 15-19 January.
- [14] Larson E. C., Chandler D. M., (2010), “Most apparent distortion: full-reference image quality assessment and the role of strategy”, Journal of Electronic Imaging, 19(1), 1-21.
- [15] Daly S. J., (1992), “Visible differences predictor: an algorithm for the assessment of image fidelity,” Proceedings of Symposium on Electronic Imaging: Science and Technology; Human Vision, Visual Processing, and Digital Display III, 2-15, San Jose, CA, United States, 9-14 February.
- [16] Watson A. B., (1993), “DCT quantization matrices visually optimized for individual images”, Proceedings of IS&T/SPIE's Symposium on Electronic Imaging: Science and Technology; Human Vision, Visual Processing, and Digital Display IV, 202-216, San Jose, CA, United States, 31 January - 5 February.
- [17] Zhang F., Ma L., Li S., Ngan K. N., (2011), “Practical image quality metric applied to image coding”, IEEE Transactions on Multimedia, 13(4), 615-624.
- [18] Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P., (2004), “Image quality assessment: from error visibility to structural similarity”, IEEE Transactions on Image Processing, 13(4), 600-612.
- [19] Kovess P., (1999), “Image features from phase congruency”, Videre: Journal of computer vision research, 1(3), 1-26.
- [20] Morrone M. C., Owens R. A, (1987), “Feature detection from local energy”, Pattern Recognit Letters, 6(5), 303-313.
- [21] Morrone M. C., Ross J., Burr D. C., Owens R., (1986), “Mach bands are phase dependent” Nature, 324(6094), 250-253.
- [22] Morrone M. C., Burr D. C., (1988), “Feature detection in human vision: A phase-dependent energy model” Proceedings of the Royal Society of London. Series B. Biological Sciences, 235(1280), 221-245.

- [23] Zhang Lin., Zhang Lei., Mou X., Zhang D., (2011), "FSIM: A Feature Similarity Index for Image Quality Assessment", IEEE Transactions on Image Processing, 20(8), 2378-2386.
- [24] Bae S. H., Kim M., (2016), "A Novel Image Quality Assessment With Globally and Locally Consistent Visual Quality Perception", IEEE Transactions on Image Processing, 25(5), 2392-2406.
- [25] Zhang L., Shen Y., Li H., (2014), "VSI: A Visual Saliency-Induced Index for Perceptual Image Quality Assessment", IEEE Transactions on Image Processing, 23(10), 4270-4281.
- [26] Sheikh H. R., Bovik A. C. (2006), "Image information and visual quality," IEEE Transactions on Image Processing, 15(2), 430-444.
- [27] Sheikh H. R., Bovik A. C., Veciana G., (2005), "An information fidelity criterion for image quality assessment using natural scene statistics", IEEE Transactions on image processing, 14(12), 2117-2128.
- [28] Chang H. W., Yang H., Gan Y., Wang M. H., (2013), "Sparse feature fidelity for perceptual image quality assessment", IEEE Transactions on Image Processing, 22(10), 4007-4018.
- [29] Li L., Cai H., Zhang Y., Lin W., Kot A. C., Sun X., (2016), "Sparse Representation-Based Image Quality Index With Adaptive Sub-Dictionaries", IEEE Transactions on Image Processing, 25(8), 3775-3786.
- [30] Zhou Z., Li J., Quan Y., Xu R., (2021), "Image Quality Assessment Using Kernel Sparse Coding", IEEE Transactions on Multimedia, 23, 1592-1604.
- [31] Ahar A., Barri A., Schelkens P., (2018), "From Sparse Coding Significance to Perceptual Quality: A New Approach for Image Quality Assessment", IEEE Transactions on Image Processing, 27(2), 879-893.
- [32] Wang Z., et al., (2015), "Sparse coding and its applications in computer vision", World Scientific.
- [33] Kruger N., et al., (2013), "Deep Hierarchies in the Primate Visual Cortex: What Can We Learn for Computer Vision?", IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(8), 1847-1871.
- [34] Barlow H. B., (1972), "Single Units and Sensation: A Neuron Doctrine for Perceptual Psychology?", Perception, 1(4), 371-394.
- [35] Karklin Y., Lewicki M. S., (2006), "Is Early Vision Optimized for Extracting Higher-Order Dependencies?", Proceedings of the 18th International Conference on Neural Information Processing Systems, 18, 635-642, Vancouver, British Columbia, Canada, 5-8 December.

- [36] Olshausen B. A., Field D. J., (1996), “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”, *Nature*, 381(6583), 607-609.
- [37] Qian J., Yang J., Xu Y., (2013), “Local Structure-Based Image Decomposition for Feature Extraction With Applications to Face Recognition”, *IEEE Transactions on Image Processing*, 22(9), 3591-3603.
- [38] Khellah F. M., (2011), “Texture Classification Using Dominant Neighborhood Structure,”, *IEEE Transactions on Image Processing*, 20(11), 3270-3279.
- [39] Farah M. J., (2004), “Visual agnosia”, 2nd Edition, MIT press.
- [40] Zhu J., Wang N., (2011), “Image quality assessment by visual gradient similarity”, *IEEE Transactions on Image Processing*, 21(3), 919-933.
- [41] Xue W., Zhang L., Mou X., Bovik A. C., (2014), “Gradient Magnitude Similarity Deviation: A Highly Efficient Perceptual Image Quality Index” *IEEE Transactions on Image Processing*, 23(2), 684-695.
- [42] Xue W., Mou X., Zhang L., Feng X., (2013), “Perceptual Fidelity Aware Mean Squared Error”, *IEEE International Conference on Computer Vision*, 705–712, Sydney, NSW, Australia, 01-08 December.
- [43] Ponomarenko N., Lukin V., Zelensky A., Egiazarian K., Carli M., Battisti F., (2009), “TID2008-a database for evaluation of full-reference visual quality assessment metrics”, *Advances of Modern Radio electronics*, 10(4), 30-45.
- [44] Ponomarenko N., et al., (2015), “Image database TID2013: Peculiarities, results and perspectives”, *Signal Processing: Image Communication*, 30, 57–77.
- [45] Sheikh H. R., Sabir M. F., Bovik A. C., (2006), “A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms”, *IEEE Transactions on Image Processing*, 15(11), 3440–3451.
- [46] Wang Z., Simoncelli E. P., Bovik A. C., (2003), “Multiscale structural similarity for image quality assessment”, *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers*, 1398–1402, Pacific Grove, CA, USA, 09-12 November.
- [47] Wang Z., Bovik A. C., (2002), “A universal image quality index”, *IEEE Signal Processing Letters*, 9(3), 81-84.
- [48] Rapporteur ITU-T Q21/9 (Co-Chair VQEG), (2003), Final report from the video quality experts group on the validation of objective models of video quality assessment, Phase II.

- [49] Lin K. Y., Wang G., (2018), “Hallucinated-IQA: No-Reference Image Quality Assessment via Adversarial Learning”, IEEE/CVF Conference on Computer Vision and Pattern Recognition, 732–741, Salt Lake City, UT, USA, 18-23 June.
- [50] Ma J. et al., (2021), “Blind Image Quality Assessment with Active Inference”, IEEE Transactions on Image Processing, 30, 3650-3663.
- [51] Pan D., Shi P., Hou M., Ying Z., Fu S., Zhang Y., (2018), “Blind predicting similar quality map for image quality assessment”, Proceedings of the IEEE conference on computer vision and pattern recognition, 6373-6382, Salt Lake City, UT, USA, 18-23 June.
- [52] Zhang W., Ma K., Yan J., Deng D., Wang Z., (2018), “Blind image quality assessment using a deep bilinear convolutional neural network”, IEEE Transactions on Circuits and Systems for Video Technology, 30(1), 36-47.
- [53] Kim J., Lee S., (2016), “Fully deep blind image quality predictor”, IEEE Journal of selected topics in signal processing, 11(1), 206-220.
- [54] Kim J., Nguyen A. D., Lee S., (2018), “Deep CNN-based blind image quality predictor”, IEEE Transactions Neural Network Learning Systems, 30(1), 11-24.
- [55] Wu J., Ma J., Liang F., Dong W., Shi G., Lin W., (2020), “End-to-end blind image quality prediction with cascaded deep neural network”, IEEE Transactions on Image Processing, 29, 7414-7426.
- [56] Liang D., Gao X., Lu W., Li J., (2021), “Deep blind image quality assessment based on multiple instance regression”, Neurocomputing, 431, 78-89.
- [57] Kingma D. P., Ba J., (2014), “Adam: A method for stochastic optimization”, arXiv:1412.6980.
- [58] Sergio A. S., Elhanashi A., (2022), “Impact of Image Resizing on Deep Learning Detectors for Training Time and Model Performance”, Applications in Electronics Pervading Industry, Environment and Society, 10–17.

BIOGRAPHY

Selcuk Ilhan AYDI got his bachelor's (degree) in computer engineering at Karabuk University in 2016. He started his Master's degree program at Gebze Technical University in 2018 and generally focused on image processing. Currently, he works at Afiniti as a senior software developer.



APPENDICES

Appendix A: List of Publications

Aydi. S. (2022). An Image Quality Assessment Method based on Sparse Neighbor Significance. In Proceedings of the 2nd International Conference on Image Processing and Vision Engineering - IMPROVE, (pp. 34–44). SciTePress. doi:10.5220/0011058700003209

