Perceived Image Quality Assessment for Stereoscopic Vision

by

Roushain Akhter

A thesis submitted to
The Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements
of the degree of

Master of Science

Department of Computer Science
The University of Manitoba
Winnipeg, Manitoba, Canada
May 2011

Thesis advisor

Dr. Jacky Baltes

Author

Roushain Akhter

Perceived Image Quality Assessment for Stereoscopic Vision

Abstract

This thesis describes an automatic evaluation approach for estimating the quality of stereo displays and vision systems using image features. The method is inspired by the human visual system. Display of stereo images is widely used to enhance the viewing experience of three-dimensional (3D) visual displays and communication systems. Applications are numerous and range from entertainment to more specialized applications such as: 3D visualization and broadcasting, robot tele-operation, object recognition, body exploration, 3D teleconferencing, and therapeutic purposes. Consequently, perceived image quality is important for assessing the performance of 3D imaging applications.

There is no doubt that subjective testing (i.e., asking human viewers to rank the quality of stereo images) is the most accurate method for quality evaluation. It reflects true human perception. However, these assessments are time consuming and expensive. Furthermore, they cannot be done in real time. Therefore, the goal of this research is to develop an objective quality evaluation methods (computational models that can automatically predict perceived image quality) correlating well with subjective predictions that are required in the field of quality assessment.

I believe that the perceived distortion and disparity of any stereoscopic display

are strongly dependent on local features, such as edge (non-uniform) and non-edge (uniform) areas. Therefore, in this research, I propose a No-Reference (NR) objective quality assessment for coded stereoscopic images based on segmented local features of artifacts and disparity. Local feature information such as edge and non-edge area based relative disparity estimation, as well as the blockiness, blur, and the zero-crossing within the block of images, are evaluated in this method. A block-based edge dissimilarity approach is used for disparity estimation. I use the Toyama stereo images database to evaluate the performance and to compare it with other approaches both qualitatively and quantitatively.

Contents

List of Figures

List of Tables

Acknowledgments

First and foremost, I would like to thank my GOD. Without Him, I highly doubt that this document would exist. I am grateful for the skills He has blessed me with and the purpose He brings to my life. I am deeply grateful to my advisor Dr. Jacky Baltes for his commitments to my education. I would like to express my thanks for his patience, inspiration, guidance and encouragement along the way.

I appreciate Dr. John Anderson for his valuable advice and suggestions. I would like to thank my thesis examination committee members, especially Dr. Gabriel Thomas for his feedback and supports to complete this thesis successfully.

I wish to thank the members of my family for their unconditional support. I would also like to thank the Media Information and Communication Technology (MICT) laboratory at the University of Toyama, Japan, for providing their subjective image database and subjective assessment scores.

This thesis is dedicated to my parents

Chapter 1

Introduction

The rapid development of three-dimensional (3D) vision technology has led to an increased interest in 3D image capture, process [21], display [7], and transfer over the Internet [42]. Most of these 3D vision technologies have involved generating stereo vision systems (i.e., a pair of cameras is used for capturing images) because of higher accuracy and capacity of handling real-time data [45]. Moreover, stereoscopic displays are desired for complex and realistic medical and health images. [41]. However, stereo vision systems contain a large amount of data which requires a huge storage capacity or high transmission bandwidth. As an example, a stereoscopic uncompressed HDTV channel may cost up to one Gbit/s transmission bandwidth [33]. Due to poor transmission channels, transmission errors or data dropping may occur, which lead to an imperfect image quality and distortion of the received video data. Therefore, digital image compression (JPEG, JPEG2000, MPEG2, H.264 or similar lossy image compression) is required for saving bandwidth and for minimum storage space [6]. Lossy image compression techniques may solve this problem, but introduce image

aritifacts like blockiness, blurring and ringing, which degrade the visual quality of an image. Therefore, perceived image quality assessment is required to ensure the applied compression techniques and the levels of compression.

1.1 Problem Motivation

The importance of the perceived quality measurement is required for many image processing applications such as compression, acquisition, restoration, enhancement, and reproduction [6; 38]. These applications involve trade-offs between the processing systems resources and the visual quality reproductions. In order to make these trade-offs efficient, a perfect image quality assessment is required. The most common methods for assessing image quality is subjective assessment [16] (i.e., ask human viewers to rank the quality of an image) because it directly reflects human perception. However, the subjective experiments are expensive and time consuming [39], and cannot be performed in real time systems because they require expensive instruments and human viewers. As a result, an objective (i.e., mathematical model) image quality evaluation method [35] that correlates well with human perception is desired. Moreover, there is demand for a precise mathematical model that can assess the quality of images without reference images (i.e., uncompressed original images) which is required for the end user. In addition, most of the widely used mathematical methods do not correlate well with human visual characteristics [12]. Under the above circumstances, a no-reference (NR) objective quality assessment metric that correlates well with human perception is desired. The NR method can be used for adjusting image quality, optimization, or for benchmark image-processing systems

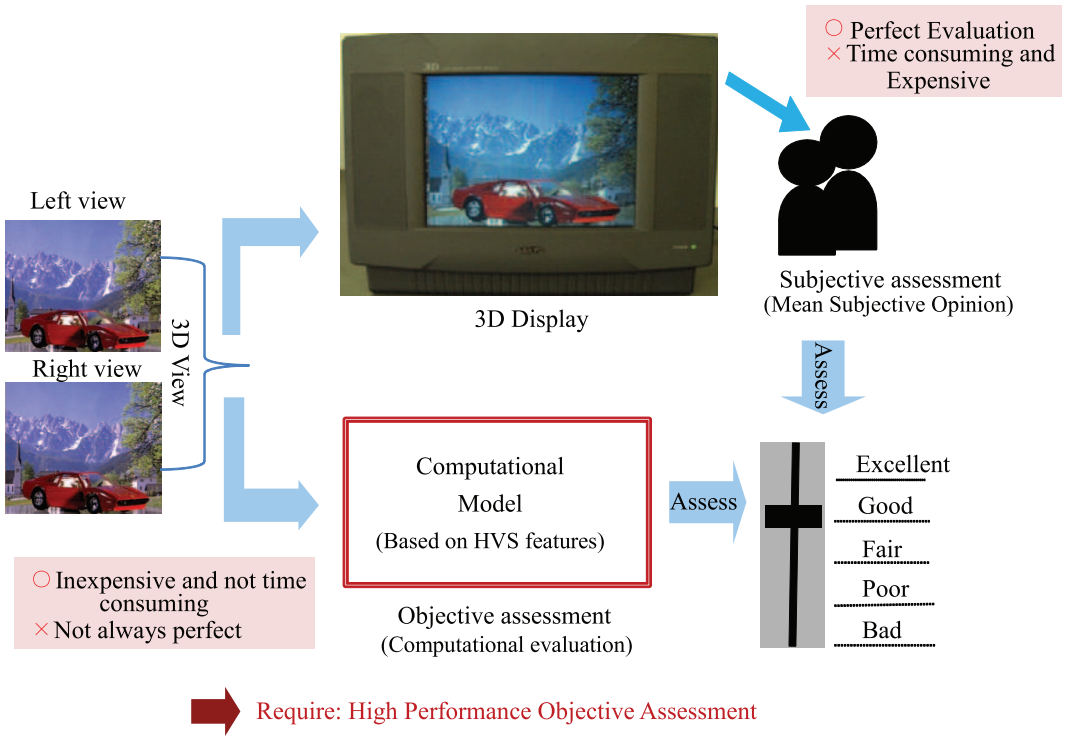


Figure 1.1: Image Quality Assessment: Subjective and Objective

and algorithms. Figure 1.1 shows an image quality assessment scenario, where two primary methods of image quality assessment are shown: subjective and objective methods. In the figure above we can see that the subjective evaluations are dependent on human voting, whereas the objective evaluations are developed by extracting image features.

The objective of my thesis is to develop a NR objective quality assessment method for coded stereoscopic images. The function of Human Visual System (HVS) is to extract structural or edge information from the viewing field of an object [22]. Therefore, humans are very sensitive to edge information, and the distortions they perceive are dependent on local features such as uniform (i.e, area with less edges) and non-

uniform (i.e, area with more edges) areas. Images are the combinations of different edges and the pixels of an image have strong dependencies between each other, especially when they are close in space. Thus, 3D depth perception is strongly dependent on objects/structures or textures edges of stereo images. Therefore, the goal is to predict the perceptual quality of stereoscopic images using local features: uniform and non-uniform areas and the relative disparity estimation.

1.2 Outline of Thesis

The thesis is organized as follows. Chapter 1 gives a general introduction and the motivation of my work. The required background and the past work on quality assessment are reviewed in Chapter 2. Chapter 3 describes my approach and the techniques to construct an automatic NR objective method for stereoscopic image quality assessment. The results and the discussion of my algorithm are given in Chapter 4. Finally, Chapter 5 concludes the thesis. Future work is also mentioned in the same chapter.

Chapter 2

Background and Related Work

2.1 Image Quality Assessment

Image quality assessment plays an important role in various image processing applications. Image quality is defined by an individual who views the image and the qualities of an image itself [3]. Human observer opinions are affected by complex physical and psychological parameters, while the quality of an image depends on various image features. Images are subject to distortions during acquisition, communication, processing and reproduction [38]. Therefore, to maintain, control and to enhance the quality of an image, quality evaluation is important, which is not easy to accomplish [37]. The image quality assessment is an approach to measure the quality of an image i.e., combining the effect of image distortion, quantification, and accumulation into a single score. It is either subjective or objective. Figure 2.1 shows the different assessment methods of image quality.

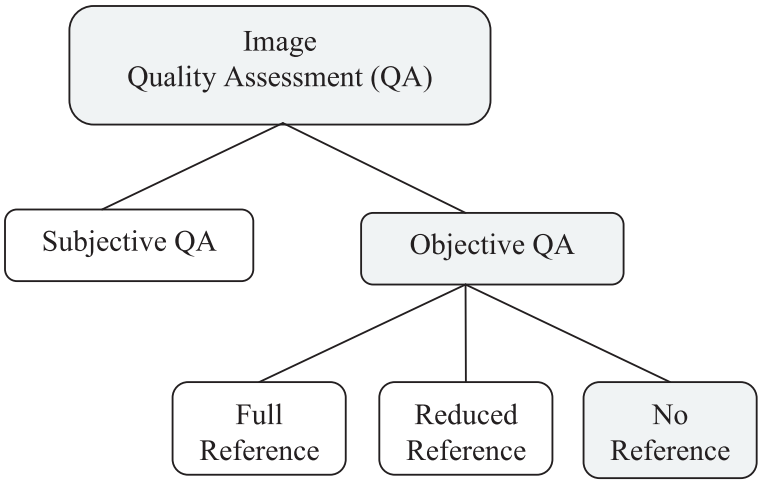


Figure 2.1: Image quality assessment methods

2.1.1 Subjective Assessment

Subjective quality assessment [35] is a method of ranking the quality of an image by human viewers. This is a straightforward approach for assessing the picture quality of digitally coded pictures. Over the years, a number of subjective assessments have been conducted, but few of them have emerged because of limited resources or failure to meet the standards defined by video quality experts group (VQEG) for FR-TV Phase II test [35]. Carefully designed and conducted subjective assessments offer direct and reliable prediction performance of a system being tested [15], but they inevitably have their own shortcomings. Subjective assessments are often very expensive and time consuming [22].

In the Media Information and Communication Technology (MICT) laboratory [25], the University of Toyama, Japan, conducted a subjective experiment on 24 bit per pixel RGB color stereoscopic images. The image database consists of 10 stereo-

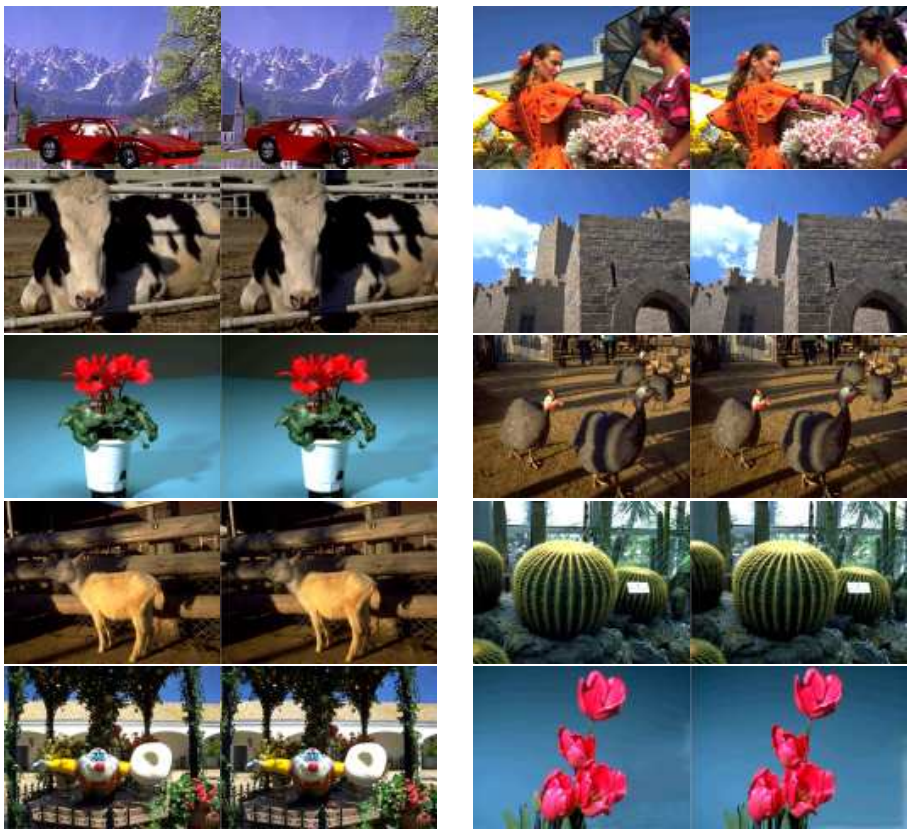


Figure 2.2: Reference stereoscopic images [25]

scopic image pairs. The stereo pairs are Car, Cattle, Flower, Goat, Mongyo, Women, Ningyou, Peacock, Saboten, and Flower2 (see Figure 2.2). Seven different levels of compression were applied for each of the images. There are 70 symmetric and 420 asymmetric coded image pairs of size 640 × 480. In the symmetric stereoscopic coding, an equal level of compression is applied to the left and right images whereas in the asymmetric coding, the compression level of the two images is different. The seven compression levels or the quality scales (QS: 10, 15, 27, 37, 55, 79, and the reference/original) were selected for the JPEG encoder [18]. QS (10–79) correspond to very low to high bit rates. Each level of compression was applied to both in

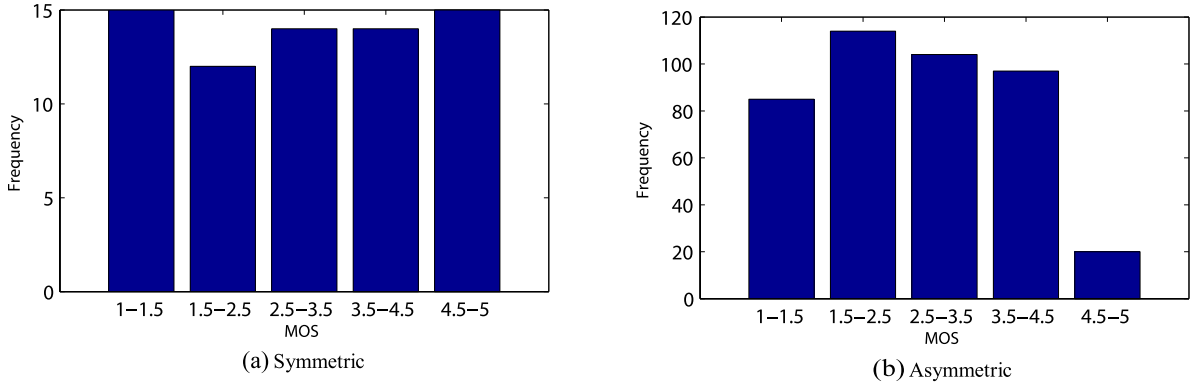


Figure 2.3: Histogram of MOS scores: (a) Symmetric, (b) Asymmetric.

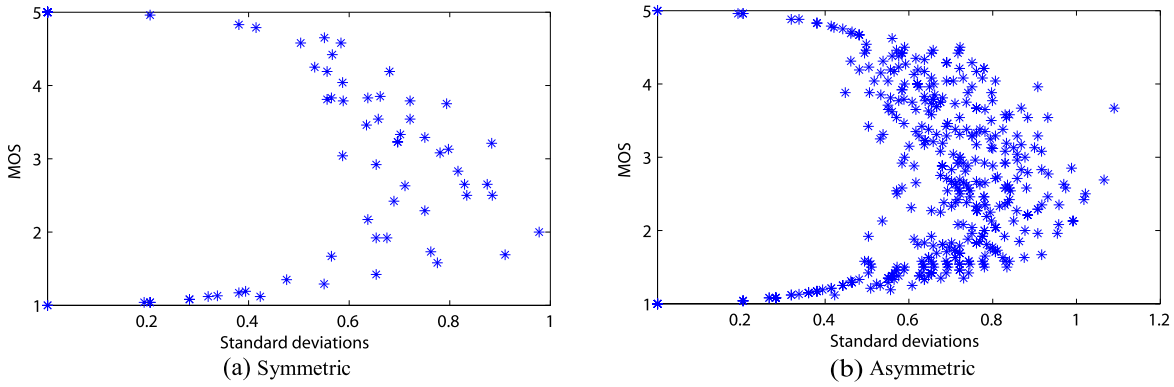


Figure 2.4: Standard deviations of MOS scores: (a) Symmetric, (b) Asymmetric.

left and right stereo image pairs. According to the recommendation of the International Telecommunication Union (ITU) for still-image coding [16], the quality scale was divided into five categories. Each category was assigned with a single numerical value of adjectives "Bad =1", "Poor=2", "Fair=3", "Good=4", and "Excellent=5". The human viewers were asked to provide their assessments on a discrete quality scale. Mean opinion scores (MOSs) were then computed for each stereo image after the screening of post-experiment results according to ITU-R 500-10 recommenda-

tion [16]. A detailed description about the experimental method and procedure is given in Appendix A. The MOS histogram and the standard deviations for all MOSs of symmetric and asymmetric images are shown in Figure 2.3 and Figure 2.4, respectively. Both histograms indicate that the image database covers all ranges of quality scale (i.e., the MOS scores from 1 to 5). Figure 2.4 shows the variation of MOS scores (subjective quality) for various compression levels. For my research, I am using the MICT stereo image database to evaluate the performance of my objective (i.e., mathematical) quality assessment technique. Therefore, I will be referring these stereo image pairs throughout my thesis and appendixes.

2.1.2 Objective Image Quality Assessment

An objective assessment is a computational method for predicting the perceived quality of images without actual human viewers. According to Wang et al. [38] "*an objective image quality model predicts the image quality sensation of an average human observer*". In other words, the strong correlation between the subjective observations and the objective quality metrics is essential when developing an objective metric. Therefore, the goal of an objective image quality assessment is to achieve a good correlation with the subject's opinion for compressed images. There are three types of objective image quality measures: Full-Reference (FR), Reduced-Reference (RR), and No-Reference (NR). The schematic diagrams of three different objective methods are shown in Figure 2.5.

Full-Reference Method: In a full-reference method [38], reference images are required to assess the quality of distorted images. Most of the quality assessment meth-

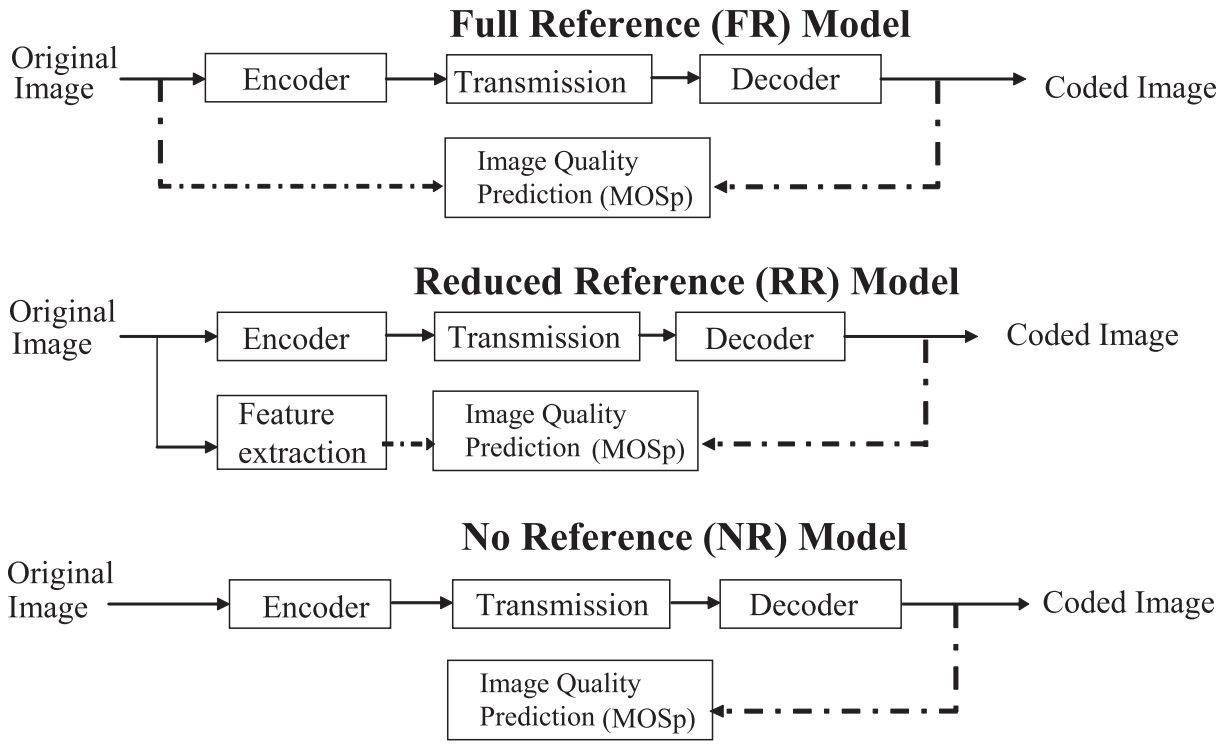


Figure 2.5: Objective image quality assessment models

ods belong to this category. However, in most of the real-time applications reference images are unavailable so a FR metric is not useful for those applications.

Reduced-Reference Method: In a reduced-reference method [38], partial information of the reference images is required to predict the visual quality of distorted images. This method is used as a substitute of FR method.

No-Reference Method: The method that does not require any reference to assess the quality of a distorted image is known as a NR method [38]. In many real-time applications, the reference image is not available, and a *no-reference* or "blind" quality assessment approach is desirable. Moreover, a human observer can assess the quality

of a distorted image without the use of any reference image. However, developing an objective NR quality measurement metric is difficult due to the limited knowledge of HVS. Therefore, a NR quality assessment is reliable only when the prior knowledge about the image distortion is available [30]. My thesis focus on this method.

2.2 Image Artifacts

Image artifacts or distortions can be defined as any features that appear in compressed images, but do not exist in the reference image. Due to many reasons, excessive image compression may cause image artifacts such as blocking, blurring, and ringing [17]. Some of these artifacts are easily detectable, but few are always out of balance. However, in my research, I will consider the following two major artifacts of coded images.

2.2.1 Blocking Artifacts

The most common artifact in digital compressed images is blockiness [17; 40], which appears as a small square block all over the image. Blockiness arises when a Discrete Cosine Transform (DCT) [29] based compression algorithm uses a high compression ratio. In DCT data are presented as a sum of cosine functions of various magnitudes and frequencies. The DCT quantitized an image by combining a large amount of image data into a small number of coefficients. In a DCT-based coder, an image is divided into several blocks of 8×8 pixels. Each of the blocks is then encoded, from the top left corner to keep only the important information of an image. As it processes forward from the upper left hand corner, the blocks are encoded with

(a) (b)

Figure 2.6: Image artifact: (a) Reference image [25] (b) JPEG Compressed Image with blockiness

fewer and fewer bits and finally, the whole image is DCT quantized [29]. During compression, the DCT coefficients of zeros are discarded without affecting the quality of the image. The compression rate and the quality of an image depend on the level of quantization of the DCT coefficients [29].

From a technical point of view, blockiness occurs due to the discontinuity at block boundaries which is generated during block-based quantization of a DCT-based coder. During compression the DCT transform operates on blocks of $M \times N(row \times column)$ pixels; as a consequence, horizontal and vertical lines at the edges of the DCT boundaries (i.e., blockiness) are exhibited in compressed images. The presence of a periodic 8×8 edge structure in an image is called blocking artifact. A reference image and the image with blocking artifacts are shown in Figure 2.6.

(a) (b)

Figure 2.7: Image artifact: (a) Reference image [25] (b) Blurred Image using Gaussian low-pass filter with 5×5 Pixels

2.2.2 Blur Artifacts

Blurring [17; 23] reduces the sharpness of edges within an image. The image becomes smoother than the reference image and causes "vague dark smudges" around the sharp edges. Blur reduces the contrast of an image. Therefore, the visibility of small objects are reduced and making it difficult to clearly view the details of the image. Blur is normally the worst artifact in most images [40]. Excessive image compression may cause this distortion. In a DCT-coded image, blur appears due to the truncation or quantization of high frequency DCT coefficients which smooth the image signal of a block [29]. Human eyes are good to spot the smoothness and blur artifacts that are occurred frequently in compressed images. Figure 2.7 gives an example of blur artifacts.

2.3 Disparity Estimation

Disparity estimation is an essential step in stereo images [24]. Stereo vision consists of two images (left and right views) captured by two closely-located cameras. The distance between the cameras is approximately the distance between two eyes. These views constitute a stereo pair and can be perceived as a 3D/virtual view by human observers. In a stereo vision, disparity is estimated by extracting features from the left image and locating the corresponding features in the right image to perceive disparity information. The relative pixels shifting from the left to the right of images are called *disparity*, which is inversely proportional to the depth of the corresponding feature points in the scene.

Disparity map construction algorithms can be divided into two groups: local and global algorithms. The local algorithms deal with small neighbourhoods of points of search areas whereas the global algorithms deal with image lines or entire images [6]. Local algorithms are efficient but they are sensitive to locally ambiguous regions (for example, to regions with a homogeneous structure). Global algorithms are less sensitive to such regions; however, they are more computationally intensive. The block-matching algorithm belongs to the group of local algorithms and is the simplest in realization and performance. Currently, in the field of disparity estimation, the block-based matching [17] (i.e., the global algorithm) has attracted much attention because of its better performance and effective implementation. The idea of block matching is to segment the target image into fixed-size blocks and find the best matches into the reference image. It determines disparity by comparing a small region (block) around a point in the first image with regions on the second image.

The sum of squares of differences of intensities and the sum of absolute differences of intensities are widely used used in measuring block similarities [36].

Eslami et al. [9] developed a different block-based method for the image distortion search called "sparse" representations of image data (i.e., where most information is packed into a small number of samples) that employ different weights to the smooth and high contrast areas of an image. In fixed-sized block matching (FSBM) [17], the original image is divided into fixed-sized blocks and similar blocks within the reference image disrespecting its location are found. These approaches fail in low texture areas. It has been proven that larger block-size increases not only the robustness but also the magnitude of errors. A hybrid approach of block-based matching is called region-based matching [4]. It allows more accurate estimation of pixel displacement. Region-based schemes are attractive because of their comparability with the HVS characteristics and follow the standards of image compression techniques.

By considering the limitations of the aforementioned approaches, I propose a simple block-based edge dissimilarity measure which considers the edge differences between the left image to the corresponding right image to measure the relative disparity. The details of my disparity assessment algorithm are discussed in chapter 3.

2.4 Objective Quality Assessment Methods Review

A large number of objective metrics have been developed that are capable of mimicking subjective assessment to measure the visible differences between a pair of images [33; 39; 31; 5]. Considering this wide range of applications, I have separated the objective research into two main categories: first, the methods that consider only

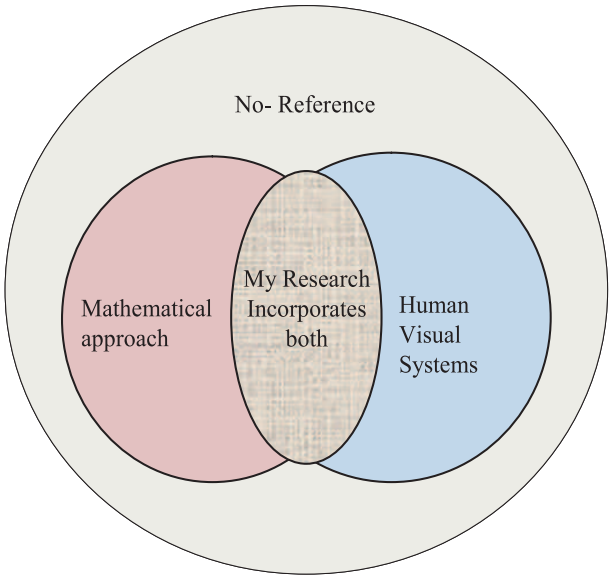


Figure 2.8: NR research scope

statistical or mathematical measurement (i.e., the image features extraction), and, second, methods that consider the HVS characteristics. In my approach, I consider mathematical measures with incorporation of HVS, that is, image feature extraction using HVS characteristics. A diagram of my proposed approach is shown in Figure 2.8.

Mathematical Model Metrics

Most of the objective metrics consider the statistical or mathematical measurement for finding the image artifacts. The mean squared error (MSE) [14] and the peak signal-to-noise ratio (PSNR) [12] are the most widely used pixel-based image quality metrics. These techniques are simple and fast, but widely criticized for not correlating well with human visual perception and require reference images [38]. PSNR

is a simple pixel-based comparison method whereas MSE is designed on statistical features for finding differences between reference and original images. They do not consider the relationship between pixels. Although MSE or PSNR are considered as a quality metrics but these are not consistent with the HVS as they measure every pixel within with equal priority. In addition, no information of structure, contrast, visibility, etc. are considered in these methods. These metrics consider the power of the error signal, but not how it affects the image. In reality pixels at different position create various effects on the HVS. Since image quality is strongly based on subjective observations, these metrics rarely work accurately on quality judgement.

MSE is the differences between corresponding pixels of the reference and the distorted images and it can be defined as:

$$MSE = \frac{1}{M \times N} \sum_{m=1,n=1}^{M,N} |I(m,n) - I_d(m,n)|^2 \tag{2.1}$$

where M×N is image size. I(m,n) and $I_d(m,n)$ represent pixels of reference and distorted images, respectively.

PSNR maps the MSE in a logarithmic way which is defined as:

$$PSNR = 10log_{10}\frac{MAX_I}{MSE} \tag{2.2}$$

where MAX_I is the maximum value that a pixel can have. As an example, the MAX value for a 8-bit grayscale image is 255.

PSNR is a popular and widely used metric to evaluate and quantify performance of image processing algorithms. But it exhibits weak performance in perceived image quality assessment due to pixel-wise error computation. Human vision is sensitive to contrast sensitivity of an image. Therefore, these mathematical models do not always correlate with human perception and fail to predict the perceived quality of an image.

I ran several experiments to analyze the prediction performance of MSE and PSNR over MOS. In order to show the performance of MSE over MOS for different level of compression of three stereoscopic image pairs e.g., Cattle, Saboten and Women (stereo images from the MICT database [25]) are given in the Figures 2.9, 2.10, and 2.11. These figures show increasing and decreasing trends for MOS and MSE scores respectively for low level of compressions (i.e., increasing bit rate) for the three different images. However, the relationship between MOS and MSE are not always inversely proportional. For example, in Figure 2.9(a) the MSE of the stereoscopic coded pairs (37-27) (indicate the compression level for left and right images, respectively) and (37-37) are different but their MOS scores are almost equal. Stereo pairs (37-79) and (37-Ref) give similar results. Figure 2.9 (c) and (d) show a similar trend among the pairs (79-37), (79-55), and (79-79) and in Figure 2.9 (d) between the stereo pairs (Ref-55) and (Ref-79). From the figures above we can see the large variation of MSE for different compression levels, whereas the MOS show a small variations for the same compression ratio. Subsequently, Figure 2.10, and 2.11 give similar trends but not exactly in the same coded levels of combinations which have already been discussed. Figure 2.9 (b) shows that variations of MSE and MOS between the stereo pairs (55-15), and (55-27) are high and low, respectively. This is because the MOS scores are strongly dependent on the texture contents of an image. So, it has been proven that MSE, a widely used mathematical approach, does not truly reflect human perception or the subjective score. Similarly, PSNR does not reflect true subjective because it depends on MSE. A similar results for PSNR are shown in appendix A.

An accurate objective image quality model predicts the image quality sensation of

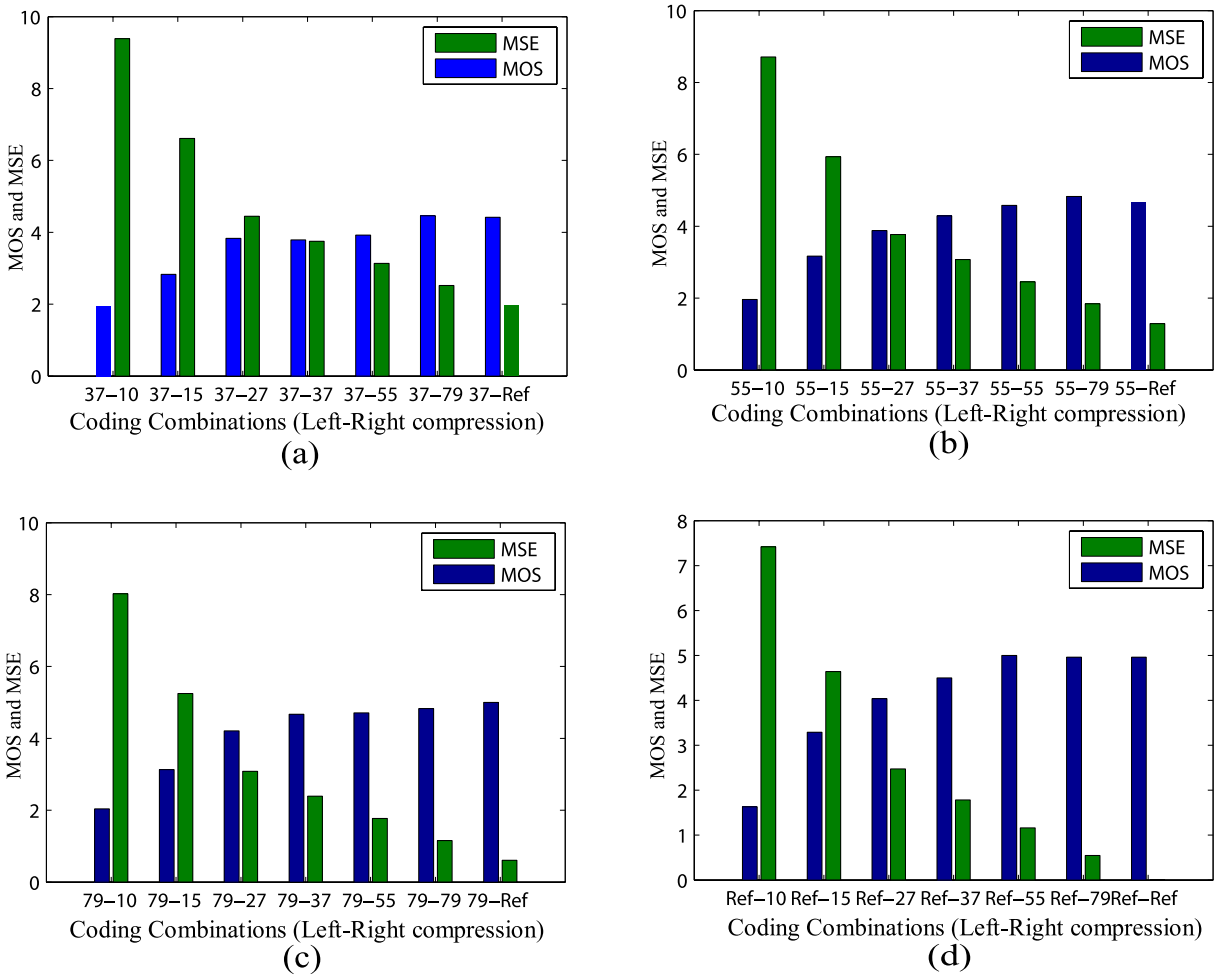


Figure 2.9: MSE and MOS of different coding levels for Cattle stereo pair

an average human observer, so a good objective model must exploit the knowledge of the HVS. In the following subsection, the development of HVS-based quality metrics are reviewed and discussed.

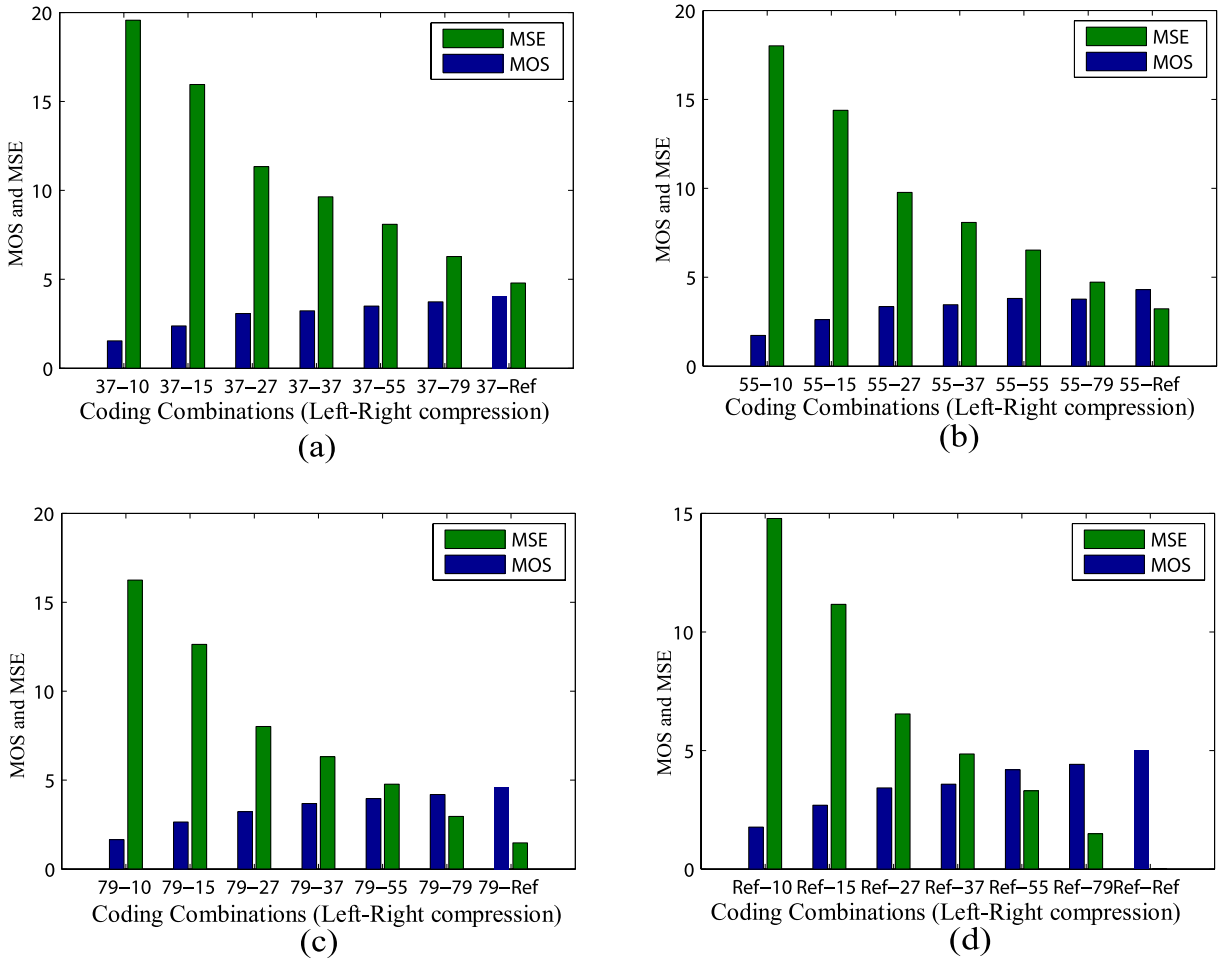


Figure 2.10: MSE and MOS of different coding levels for Saboten stereo pair

HVS-based Feature Metrics

As image quality assessment should depend on the assessments made by humans, a better understanding of features of the HVS should lead to more effective comparisons [28], which in turn will assess more realistic and reliable perception. Recently, a great deal of effort has been made to the development of visual models that take advantage of the known characteristics of the HVS [33; 14; 10]. The aim of the HVS-

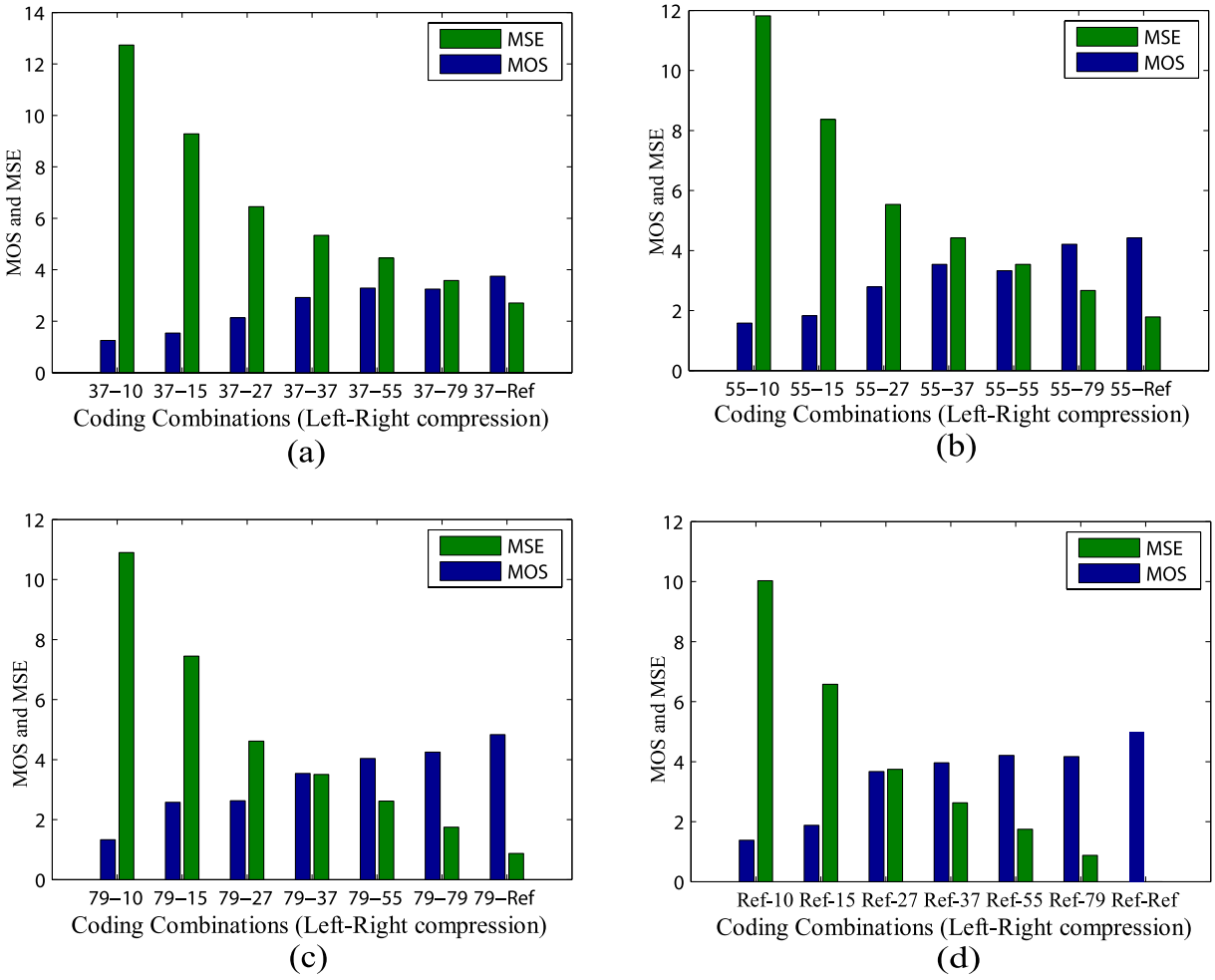


Figure 2.11: MSE and MOS of different coding levels for Women stereo pair

based objective assessments are to evaluate how strong the distorted information is perceived by the metric, according to the characteristics of the HVS. However, most of these proposed approaches require the original image as a reference, therefore, these metrics can not be used in real time applications.

Recently, extensive research into designing objective metric based on artificial neural network has done. Gastaldo et al. [11] proposed a neural network based

objective metric using PSNR. They present their quality assessment on a circular back-propagation neural network model. Though the performance of their metric is good, it was not verified widely by other databases. Horita et al. [14] introduced a FR metric for stereoscopic color images by considering the HVS. They considered all the components such as luminance and color of the CIELab color space (i.e., colors are considered as combinations of red and yellow, red and blue, green and yellow, and green and blue). Image impairments were quantified by the differences of coded and reference images. The MSE algorithms were used in their metrics for finding the pixel distortions. Gorley et al. [12] developed a point matching technique to identify the contrast and the luminance changes between the left and the right images. Most of these metrics were developed based on the structural similarity [22], segmentation and/or quality-awareness [34]. Therefore, they fail when the reference image or at least the partial information of the reference image is unavailable.

Wang et al. [39] designed their metric by finding two major image artifacts, blockiness and blur within an image. The image impairments were identified from the shifting of the left and right image pair. The major draw back of their metric is high computational cost. A similar approach was proposed by Sazzad et al. [31] where the metric was developed to find the texture areas of an image. However, the metric was developed for 2D image quality assessment. In the region-based matching [4], researchers investigated the sensitivity of the human eyes for occluded regions. Marziliano et al. [23] presented a NR blur metric by measuring average edge transition widths, and this blur measure was used to predict the quality of JPEG2000 compressed images.

Many researchers proved the sensitivity of human eye on image features such as occluded, edge, or smooth regions of an image [14; 4]. Some of them focused on the brightness, contrast, or structural similarities [40; 9] of an image. A similar approach was developed by Sazzad et al. [32] where three local features: edge, plain (i.e., uniformly colored areas), and textured areas have been considered to find the image impairments. Their metric has a large number of parameters, which poses a danger of over-fitting the metric. Moreover, the computational cost of their metric is very high.

The discussion above implies that for the past few decades, a lot of research has been done in the field of 2D objective image quality assessment [33; 6; 39], but no comparable effort has been taken towards the quality assessment of stereoscopic images. More specifically, very little research has been done on NR stereo image assessment by considering the HVS. Emerging 3D technologies still require a larger number of quality metrics and methodologies by taking into account the fundamental characteristics of the HVS and typical distortions of stereoscopic image content. Therefore, I strongly believe that my research would be recognized as a great contribution in the field of 3D imaging.

Chapter 3

Implementation

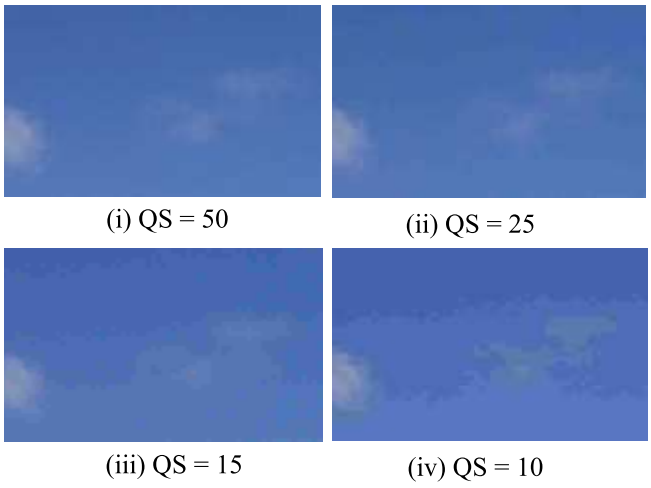
This chapter begins with an analysis of the design requirements followed by a detailed description of design and development procedures of my objective image quality assessment model. My algorithm consists of two parts: first, finding image artifacts in a stereo pair, and, second disparity estimation for the image pair. I am taking blockiness and blur as major image artifacts because these are the most visible distortions in compressed images. The preprocessing stage of my method is block-based segmentation which is discussed first and then the details of my algorithm are given in the following sections.

3.1 Design Motivation

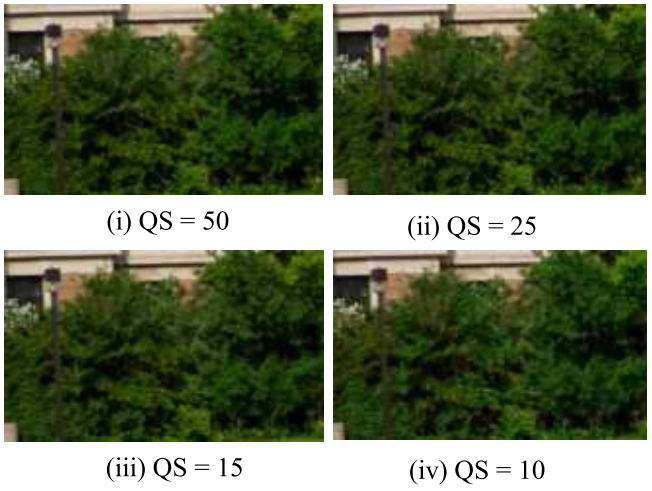
Image compression creates artifacts which degrade the visual quality of an image [38]. In theory, the visual artifacts of an image increase with an increased rate of compression (i.e., decreasing bit rate, the number of bits per pixel (bpp) used

during image compression). However, the relationship between the artifacts and the level of compressions (i.e., different bit rates) depends on the texture contents of an image [26]. In order to verify this relationship, I ran several experiments to analyze the degradation of images which cause visual difficulty i.e., appearance of image artifacts for various compression and texture levels within the images. For instance, JPEG coding introduces blockiness within the "University of Manitoba" image (see Figure 3.1 (a)) that contains a variety of textures such as uniform and non-uniform

(a) Reference image; Image taken from [1] (b) Compressed image; QS = 10



(c) A small portion of an uniform area



(d) A small portion of a non-uniform area

Figure 3.1: Variation of Perceived Distortion (uniform and non-uniform areas)

areas. Out of all edge and non-edge areas in Figure 3.1(a). I analyze a small portion of uniform and non-uniform areas which are represented by the top-right rectangular box and the bottom-right rectangular box (dotted line), respectively. A high level of compression (i.e., lowest bit rate, QS10) is applied to the image. See Figure 3.1(b). The result shows the blocking artifacts is more visible to uniform areas compared to that the non-uniform areas (see the corresponding areas in the compressed image) even though the level of compression is equal. In order to study the relationship more extensively, I apply four levels of compression (QS: 50, 25, 15, and 10) to the image and consider expanded views of the portions of non-uniform and uniform areas (consider the rectangular box areas) for each level of compression which are shown in Figure 3.1(c) and 3.1(d), respectively. The range of QS is between 0 to 100 (i.e., bit rate range from 0.01 bpp to 2 bpp). More specifically QS10, 25, and 60 indicate 0.5, 0.75, and 1 bpp, respectively [26]. QS0 and QS100 indicate the worst (i.e, very high level of compression) and the best quality (i.e, very low level of compression) of am image, respectively. These two figures indicate that perceived distortions for these areas are not similar even if the compression levels are equal. In details, blocking artifacts are more visible in uniform areas compared to the non-uniform areas (see Figure 3.1 (c)(iii) and (d)(iii), and also Figure 3.1 (c)(iv) and (d)(iv)). Similarly, the blur artifacts are more visible in the non-uniform areas compared to uniform areas see Figures (c)(iii) and (d)(iii), and also Figures 3.1 (c)(iv) and (d)(iv)). Therefore, the results indicate that visibility of image artifacts is strongly depended on the texture similarity of the area. Thus, I studied the effect of overall image quality (subjective quality score) not only for various level of compression but also for

symmetric/asymmetric coded stereoscopic images.

3.1.1 Effects of Bit rates on Image Quality

In order to explain image quality for various bit rates, I took two stereo images e.g., Car and Saboten (stereo image pairs from the MICT database [25]) to present a comparative analysis of MOS versus compression levels i.e., different bit rates. Coding levels versus overall image quality for the two sample pairs of images are shown in Figures 3.2, and 3.3. The figures indicate that the image quality (MOS score) increases with an increasing bit rate for both images with a few exceptions. At the points (37-55, and 37-79) of Figure 3.2 and (55-27, and 55-37) of Figure 3.3, one view(left or right) of the stereoscopic image pairs are coded with high JPEG compression level of QS37 and/or QS27. Therefore, the corresponding views contain large artifacts, primarily the blockiness and blur artifacts, can not be reduced by using high quality view for the rest of a pair. Moreover, the Car and Saboten images contain some uniform and non-uniform areas where blockiness and blur artifacts are more visible for uniform and non-uniform areas, respectively. Consequently, even though the average bit rate is increasing for those image pairs, subjective quality score remains fixed. Therefore, the results suggest that quality score not only depend on the level of compressions but also the texture contents of an image has large impact on perceived image quality, especially in the range of medium to high bit rate (i.e., coding level, QS37 and above). However, the quality trend increases with increasing of bit rate specifically at low bit rate range (i.e., coding level up to less than QS37).

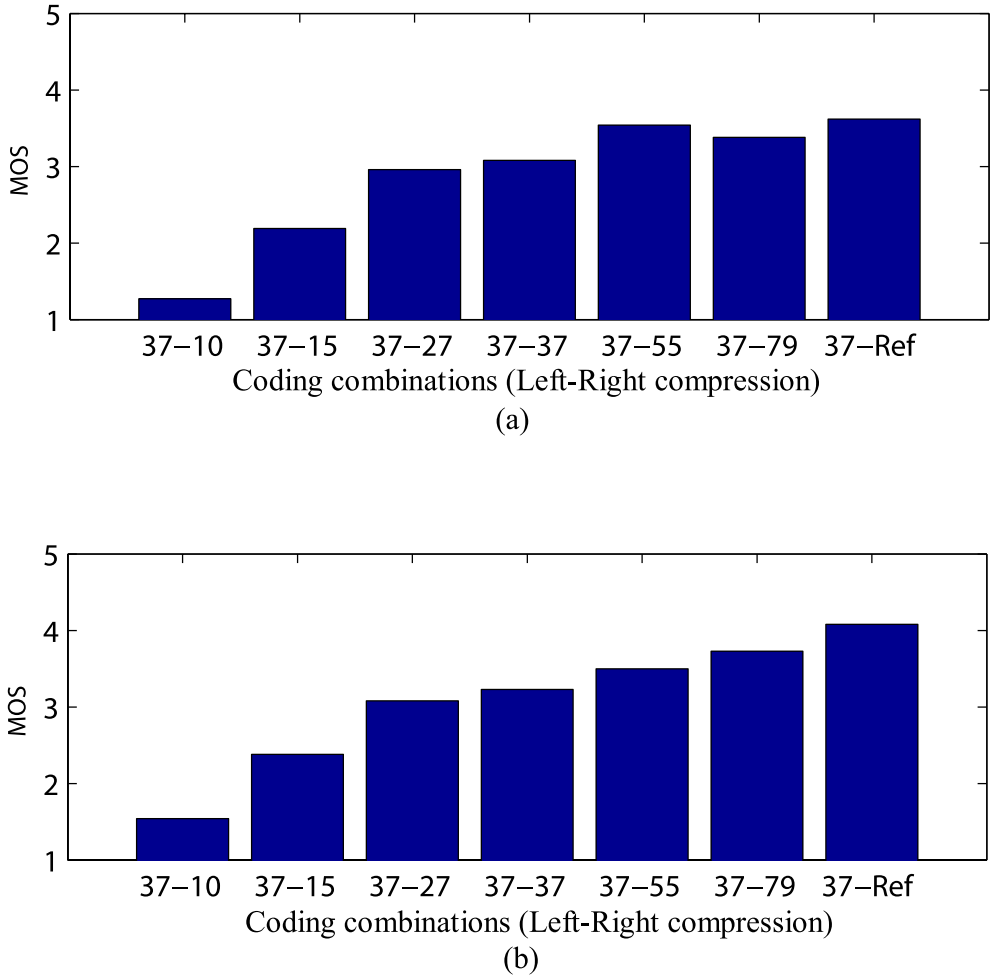


Figure 3.2: MOS score for different coding combinations: (a) Car (b) Saboten

3.1.2 Effect of Symmetric/asymmetric Compression on Image Quality

I investigated symmetric and asymmetric coded image pairs , to study the effect of subjective quality score. I use Cattle and Women stereo pairs (stereo image pairs from the MICT database [25]) for my description. Figures 3.4, and 3.5 show overall image quality scores and the coding levels of different symmetric and asymmetric

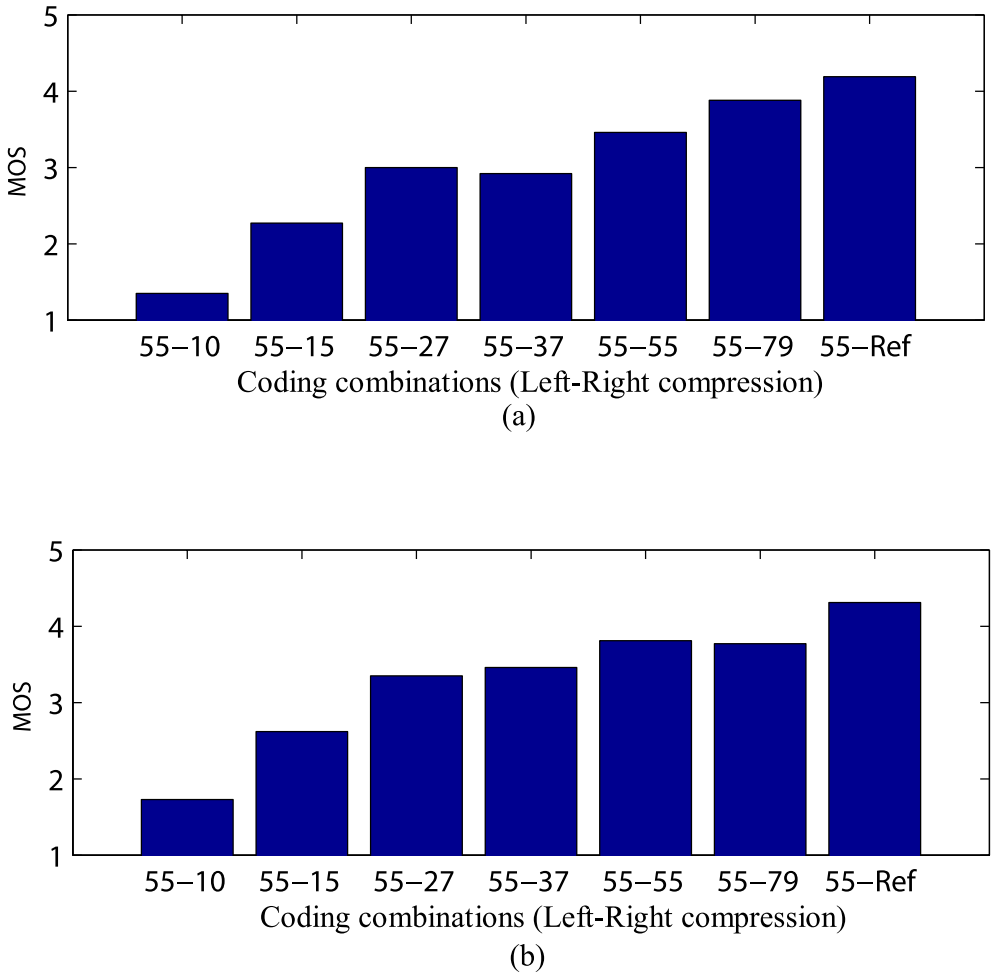


Figure 3.3: MOS score for different coding combinations: (a) Car (b) Saboten

pairs for the stereo images Cattle and Women. These two figures indicate that the relationship between the perceived quality and the average bit rate is not linear for stereoscopic images. In Figure 3.4, the quality scores of the asymmetric coded pair (37-27) is higher compare to that of the symmetric coded pair (37-37), even though the average bit rate for asymmetric pair (37-27) is lower than the symmetric pair (37-37). Also in Figure 3.5, the quality scores of the asymmetric coded pair (55-37) is higher compare to the symmetric coded pair (55-55), even though the

average bit rate for the asymmetric pair (55-37) is lower than the symmetric pair (55-55). The results presented in [33] shows that the perceived image quality of a stereo pair is approximately the average of high and low quality scores for that pair and the overall perceived quality follows the low quality score of a stereo pair. However, the MICT Lab's subjective data do not support that statement. Rather the subjective score of a stereo pair strongly depends on the texture contents of an image pair. Moreover, the overall image quality of a stereo pair follows a low score when the the level of compressions (i.e., in low bit rate range) is very high. The above explanation shows the deviance. It also shows that asymmetric coding is better compared to the symmetric for coded stereoscopic image pairs to reduce average bit rate without compromising the perceptual quality of an image. Thus the symmetric (79-79)/asymmetric (79-55) scores are approximately of equal quality for both Cattle and Women images (see Figure 3.4 and 3.5). Although the result presented in [33] suggests that symmetric coding is better for coded stereoscopic image pairs, the MICT's data suggest that asymmetric coding is better compared to symmetric pairs for coded stereoscopic images.

It is clear from the above study that perceived image artifacts are strongly dependent on local features such as uniform and non-uniform areas. Moreover, it has already been established that the primary function of the HVS is to extract structural or edge information from the viewing field, and the HVS is highly adapted for edge detection [31]. Thus, I believe that the human perception is strongly dependent on objects, structures or textures of stereo image content. Therefore, the purpose of my thesis is to develop a new stereoscopic image quality metric both for symmetric

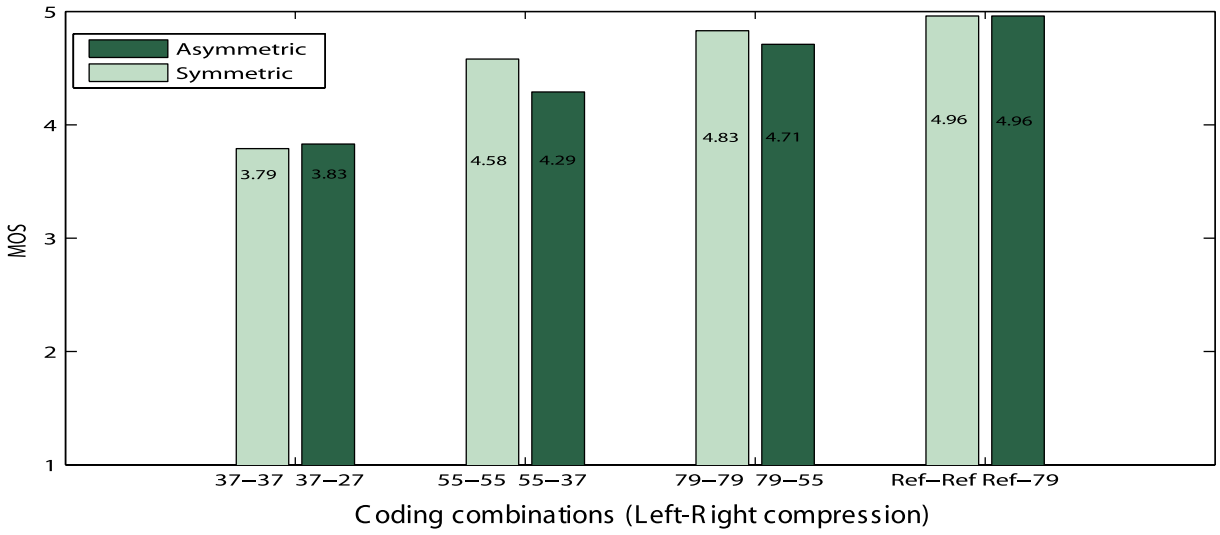


Figure 3.4: MOS score for different symmetric and asymmetric coding combinations of Cattle images

and asymmetric coded images by considering artifacts/distortion measures with the incorporation of HVS characteristics. A NR objective stereoscopic image quality assessment method is proposed based on artifacts and disparity measures distinctly for uniform and non-uniform areas. My thesis is an important extension of the research in [2]. It presents a modified version of the algorithm not only by reducing the number of parameters but also reducing the chance of over-training the metric by using too many parameters. These two important factors will reduce the computational cost significantly and make my algorithm more robust. The block diagram of my proposed model is shown in Figure 3.6.

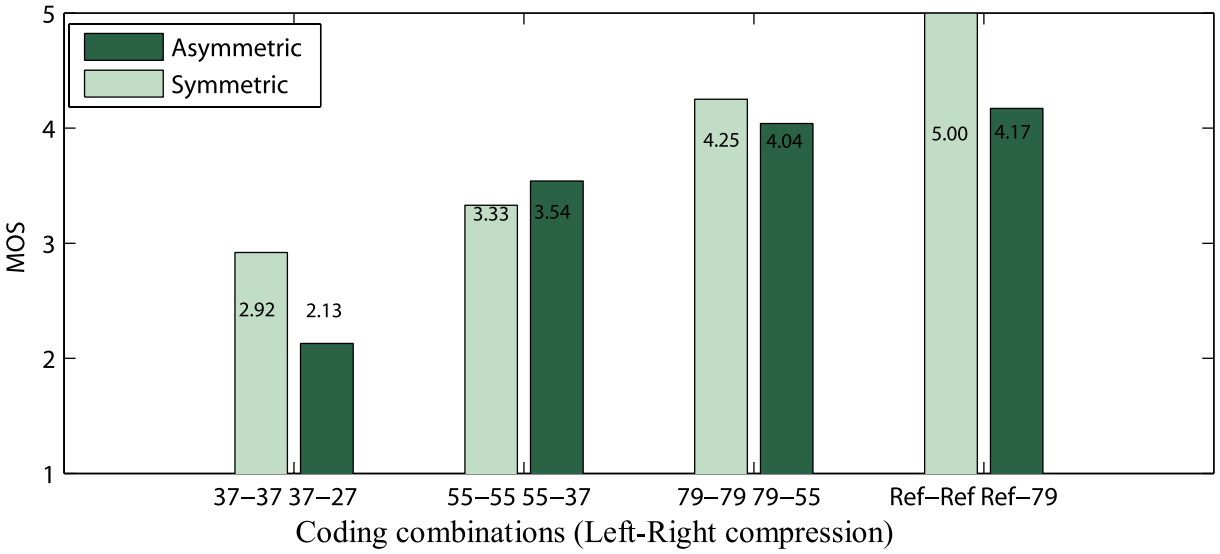


Figure 3.5: MOS score for different symmetric and asymmetric coding combinations of Women images

3.2 Metric Design and Development

The goal is to predict the perceived quality of stereoscopic images without human viewers. A stereoscopic image consists of two views: left and right. An efficient 2D compression technique, such as JPEG coder, can be applied independently both on left and right images to compress the stereo pair. Since JPEG is a block-based DCT coding technique, both blocking and blurring artifacts may be created during the quantization of the DCT coefficients. Blur increases significantly when the level of compression is high (i.e., low bit rate per second). Thus, the higher the blur, the smoother the image signal is which causes the reduction of signal edge points. Consequently, the average edge point detection of a block give more insight about the relative blur in an image. Here, zero-crossing [17]; an edge detecting technique is used to detect every pixel of an image into edge or non-edge pixel. The edge detection

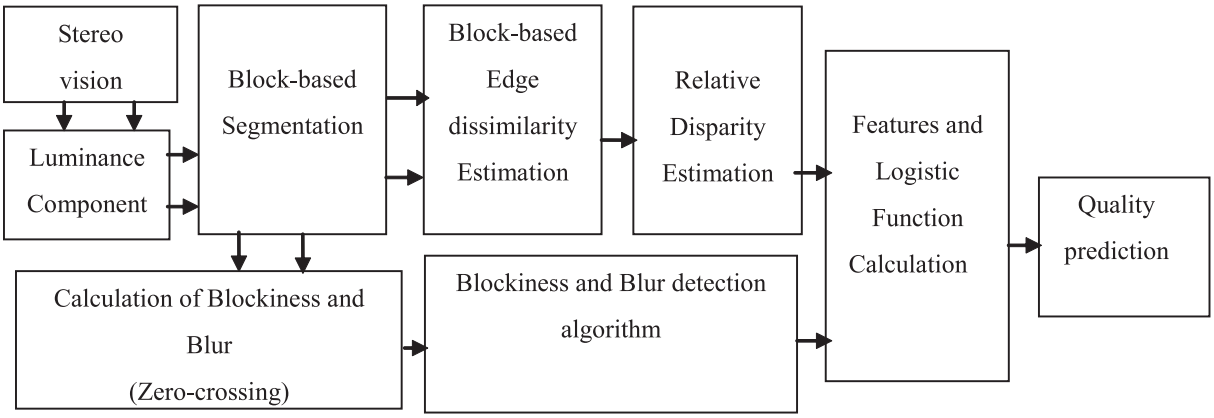


Figure 3.6: NR Objective Quality Evaluation Model

is a technique to identify the changes of image brightness that can be grouped into two categories: gradient and Laplacian. The gradient detects edges by looking at the maximum and minimum values of the first derivative of an image, whereas the Laplacian searches for zero-crossings in the second derivative to find edges. In order to explain the effect of number of zero-crossing on compression levels, I consider two images, Cattle and Women. The total number of zero-crossing points (i.e., the total number of pixels where zero-crossing occurs) of the two images for different coding levels are shown in Figure 3.7. The results show that zero-crossing points decrease with a decreasing bit rate (i.e., coding levels, QS 55 to 10). This result suggest that blur artifacts of an image can be estimated by the detection of zero-crossing points of the image. Although the effect of overall image coding artifacts for an asymmetric image pair depends on the visual appearance of the artifact, where blockiness are more disturbing than blur [31], I take the maximum blockiness and blur values between the left and right views of a stereo pair. Therefore, I consider the higher blockiness

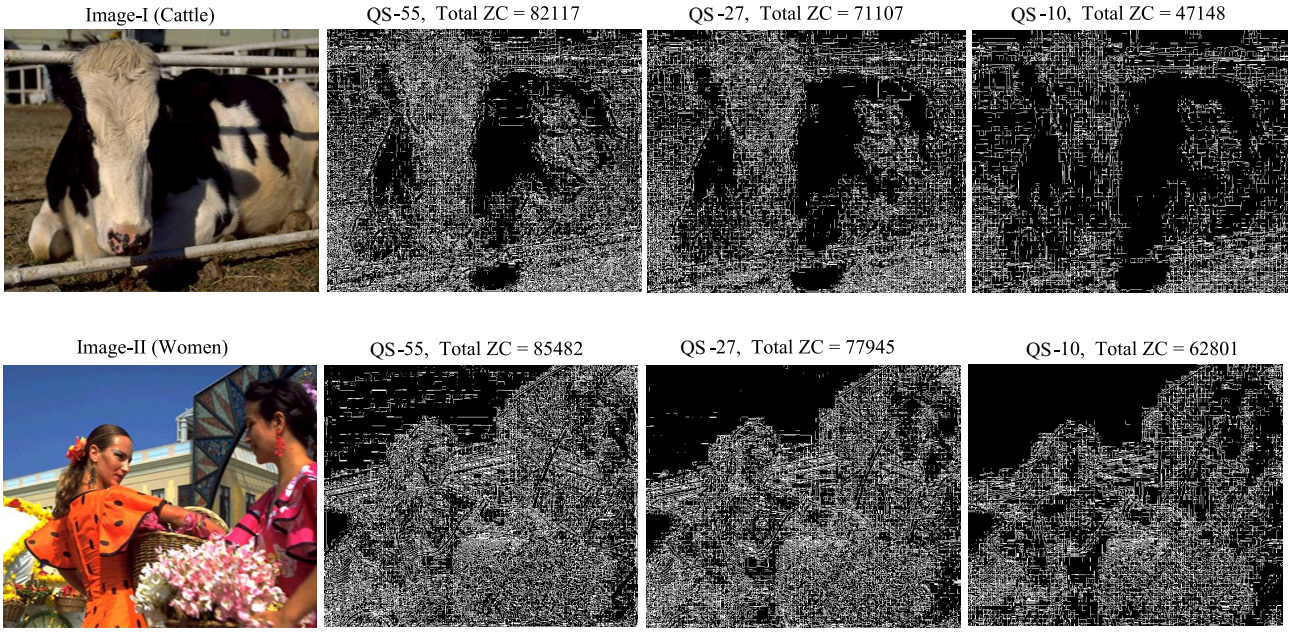


Figure 3.7: Effect of number of zero-crossing on coding levels for Cattle and Women images. Reference images from the MICT database [25]

and the lower zero-crossing values between the two views. For simplicity, only the luminance component is considered to make quality prediction of the color stereo images. The details of my segmentation algorithm to classify an image into uniform and non-uniform areas is discussed in the following section. Subsequently, the artifacts and disparity measures are described in the later sections.

3.2.1 Block-based Segmentation

A block-based segmentation [43; 44] algorithm is used to classify the edge and non-edge areas of an image. At first, I design a simple pixel classification algorithm to classify every pixel of the image into edge, or non-edge pixel. Initially, the standard deviation (STD) of pixel values at each location is estimated within its 3×3 and

5×5 neighbourhood. For all corners pixels, I take into account only available pixels for the measures. A large window increases the probability of sufficient intensity variation, but it reduces the selectivity of finding correct disparities, especially in areas with depth discontinuities. The adaptive window technique [27], adjust the disparity of points near the target over the window size. This may solve the issue but the computational cost goes high. Therefore, to reduce computational cost, I am considering 3×3 and 5×5 neighbourhoods of pixels only. Let $STD_{3\times3}(m,n)$, and $STD_{5\times5}(m,n)$ be the standard deviated images of 3×3 and 5×5 neighbourhoods, respectively. I then calculate absolute difference, $D_a(m,n)$ by the following equation:

$$D_a(m,n) = |STD_{3\times3}(m,n) - STD_{5\times5}(m,n)|, \tag{3.1}$$

where m = 1, 2,....M and N = 1,2,....N. Here, M and N are the total number of rows and column i.e., image size. Subsequently, I calculate STD of $D_a(m,n)$ by

$$D = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M\times N} (\overline{D_a} - D_{a_i})^2} \tag{3.2}$$

where D_{a_i} is the D_a value for each pixel or image sample and $\overline{D_a}$ is the average absolute difference of D_a. I then use the following algorithm to classify edge and non-edge pixels of the image.

$$P\,(m,n) = \begin{cases} 1 & \text{if } D_a(m,n) >= \text{D} \\ 0 & \text{otherwise} \end{cases}$$

where "1" and "0" denote edge and non-edge pixels respectively. Secondly, we classify each block (8×8) of the image into either uniform or non-uniform block by using the following equation:

Let

$$Sum = n_e + n_n \tag{3.3}$$

where n_e and n_n are respectively the number of edge and non-edge pixels per (8×8) block. Therefore, the "Sum" is the total number of pixels per block.

if $\left(\frac{n_e}{Sum} > th_n\right)$ *then the block is "*$uniform$*"*
else the block is "$non-uniform$*"*

where "th_n" is the algorithmic threshold. The threshold is estimated empirically. Details can be found in Section 4.

3.2.2 Artifacts Measure

Perceived quality of a coded image could be degraded by blockiness and blur. I consider both blockiness and blur in the spatial domain (i.e., the image space, in which a change in position in pixels directly corresponds to a change in position in the scene). Here, the "blockiness" of a block (i.e., 8×8 pixels) is calculated by averaging the absolute difference across a block boundary whereas blur is measured by the averaging the edge points within that block. Blur smooths the image signal within a block causing the reduction of edge points. Therefore, the higher the blur the smoother the image signal is. Consequently, average edge point detection within a block gives more insight about the relative blur of that image. Thus, the zero-crossing technique can be used to detect edge points of an image to measure blur of an image. To measure these two major artifacts the following steps are followed:

- At first, I calculate the blockiness and zero-crossing of each block of the stereo image pair separately (left and right images).

- In the second, I apply the block (8×8) based segmentation algorithm to the left and right images separately to classify uniform, and non-uniform blocks within the images.

- I calculate the average value of blockiness and zero-crossing separately for the edge, and non-edge blocks (for uniform and non-uniform areas) for each of the stereo pair.

- The total blockiness and zero-crossing of a stereo image pair is cacluated by taking the maximum blockiness and the minimum zero-crossing between the left and right of a stereo image pair. A detailed explanation of choosing the maximum blockiness and the minimum zero-crossing value is given in the evaluation chapter.

- Finally, I update the total blockiness and zero-crossing values with weighting factors by using the PSO algorithm.

The mathematical features, blockiness and zero-crossing within a block of an image are calculated both in horizontal and in vertical direction, respectively. For horizontal direction: Let the test image signal be *I(m, n)* for m ∈ [1, M] and n ∈ [1, N], image signal difference for in horizontal direction is calculated by

$$d_h(m,n) = I(m,n+1) - I(m,n), \tag{3.4}$$

n ∈ [1, N-1] and m ∈ [1, M]

Blockiness of a block (8×8) in the horizontal direction is estimated by

$$B_{bh} = \frac{1}{8}\sum_{i=1}^{8} |d_h(i,8j)| \tag{3.5}$$

where "i" and "8j" are respectively number of row and column position, and j = 1, 2, 3, ...(N/8).

For the horizontal zero-crossing (ZC):

$$d_{h-sign}(m,n) = \begin{cases} 1 & \text{if } d_h(m,n) > 0 \\ -1 & \text{if } d_h(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.6}$$

Where $d_{h-sign}(m,n)$ is the sign identification of $d_h(m,n)$.

$$\begin{aligned} & d_{h-mul}(m,n) \\ &= d_{h-sign}(m,n) \times d_{h-sign}(m,n+1) \end{aligned} \tag{3.7}$$

Where $d_{h-mul}(m,n)$ is the multiplication between two adjacent column of the sign identification signal $d_{h-sign}(m,n)$. I define for n $\in$ [1, N-2]:

$$z_h(m,n) = \begin{cases} 1 & \text{if } d_{h-mul}(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.8}$$

Here $z_h(m,n)$ indicates horizontal zero-crossing for a pixel (m,n). The value "1" indicates zero-crossing and "0" indicates no zero-crossing. The size of $z_h(m,n)$ is $M \times (N-2)$.

The horizontal zero-crossing of a block (8×8), ZC_{bh}, is calculated as follows:

$$ZC_{bh} = \sum_{i=1}^{8} \sum_{j=1}^{8} z_h(i,j) \tag{3.9}$$

Thus, I can calculate blockiness and zero-crossing of each available block of the left and right images.

For the vertical direction: I calculate a difference signal along each vertical line:

$$d_v(m,n) = I(m+1,n) - I(m,n), \tag{3.10}$$

$$\text{n} \in [1, \text{N}] \quad \text{and} \quad \text{m} \in [1, \text{M-1}]$$

Similarly, the vertical features of blockiness (B_{bv}) and zero-crossing (ZC_{bv}) of the block are calculated. Therefore, the overall features B_b and ZC_b per block are given by:

$$B_b = \frac{B_{bh} + B_{bv}}{2}, ZC_b = \frac{ZC_{bh} + ZC_{bv}}{2} \tag{3.11}$$

Consequently, the average blockiness value of uniform, and non-uniform areas of the left image are calculated by:

$$Bl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} B_{be} \tag{3.12}$$

$$Bl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} B_{bn} \tag{3.13}$$

where N_e, and N_n are the number of uniform, and non-uniform blocks of the image, respectively. Similarly, the average blockiness values of Br_e, and Br_n for the right image are calculated.

Accordingly, the average zero-crossing values of ZCl_e, and ZCl_n for the left image are estimated by:

$$ZCl_e = \frac{1}{N_e} \sum_{b=1}^{N_e} ZC_{be} \tag{3.14}$$

$$ZCl_n = \frac{1}{N_n} \sum_{b=1}^{N_n} ZC_{bn} \tag{3.15}$$

Similarly, the average zero-crossing values of ZCr_e, and ZCr_n for the right image are calculated. I then calculate the total blockiness and zero-crossing features of uniform, and non-uniform areas of the stereo image. For the total blockiness features (B_e, and

B_n) of a stereo image, I consider the maximum values for a pair by using following algorithm:

$$B_{e/n}\,(Bl, Br) = max(Bl, Br) \tag{3.16}$$

However for zero-crossing features (ZC_e, and ZC_n), I estimate the minimum values of a stereo image pair by using the following equation:

$$ZC_{e/n}\,(ZCl, ZCr) = min(ZCl, ZCr) \tag{3.17}$$

Finally, the overall blockiness and zero-crossings of each stereo image pair are calculated by

$$B = {B_e}^{w_1} \cdot {B_n}^{w_2} \tag{3.18}$$

$$Z = {ZC_e}^{w_3} \cdot {ZC_n}^{w_4} \tag{3.19}$$

where w_1, and w_2 are the weighting factors for the blockiness of uniform, and non-uniform areas and also w_3, and w_4 are the weighting factors for zero-crossing which are calculated by optimization algorithm.

3.2.3 Disparity Estimation

To measure disparity, I use a simple fixed size block matching approach to segment the target image into fixed size of blocks and then finding each block of the left image to that the corresponding area of the right image for that stereo image pair.

In other words, the displacement between x-coordinates in the left and right views is referred to as disparity in this algorithm. The stereoscopic images in the database that I considered are epipolar rectified images. The displacement between the left and

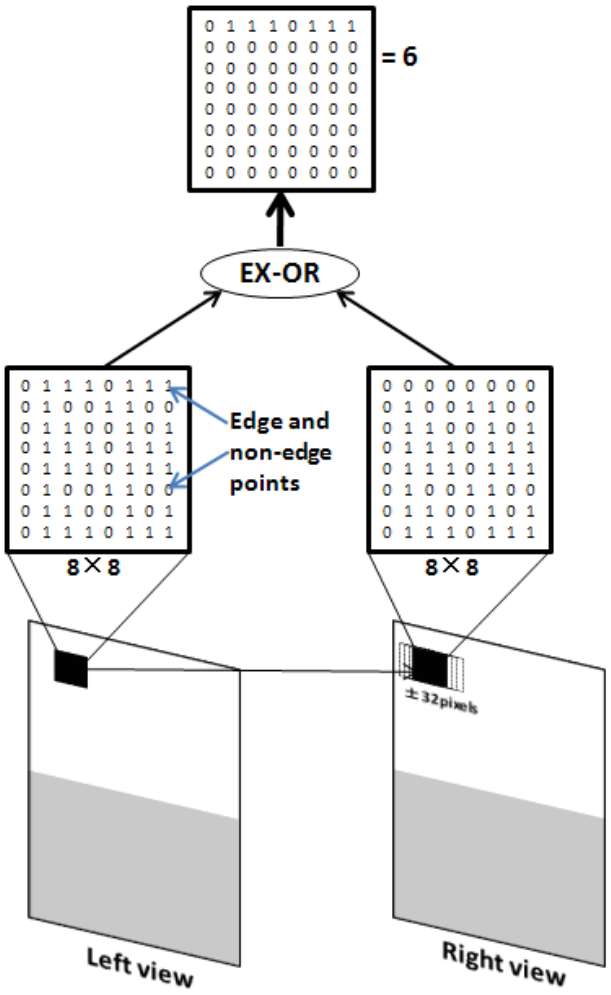


Figure 3.8: Disparity estimation approach

right view of a stereo pair is restricted in the horizontal direction only. A *block-based edge dissimilarity* measurement approach, i.e., the edge differences within the search block of a stereo pair, is proposed for relative disparity estimation. The principle of my disparity estimation is to divide the left image into non overlapping fixed-sized (8×8) blocks with classification of uniform and non-uniform blocks. A disparity value for each block of the left image is estimated by searching the minimum difference zero-

crossing (DZC) value for that corresponding block in the right image. To reduce the computational cost, the searching is restricted within ±128 pixels and one dimensional (horizontal direction) only. The disparity estimation approach is shown in Figure 3.8. The figure shows the value of minimum DZC of a block is 6. Here zero-crossing (horizontal and vertical) of a block is estimated according to Section 3.2.2 where, "1" and "0" indicate zero-crossing (edge) and not zero-crossing (non-edge) points, respectively. Consequently, the relative disparities in my algorithm for non-uniform (edge) and uniform (non-edge) areas are evaluated by averaging the DZC values.

In order to measure disparity, firstly, the segmentation algorithm is applied to left image only to classify non-edge and edge blocks. Secondly, block-based DZC is estimated in the two corresponding blocks between the left and right images. Thirdly, I average the DZC rate values separately for non-edge and edge blocks. Finally, the values are updated with some weighting factors. ZCl, and ZCr are the zero-crossing of a block in the left image and the corresponding block in the right image, respectively. The DZC of the block can be estimated by the following equation:

$$DZC = ZCl \oplus ZCr \tag{3.20}$$

Subsequently, the DZC rate (DZCR) is calculated by:

$$DZCR = \frac{1}{8 \times 8} \sum DZC \tag{3.21}$$

For the horizontal direction: let ZCl_h, and ZCr_h be the horizontal zero-crossing of a block in the left image and the corresponding search block in the right image, respectively. The DZC_h of the block is estimated by the following equation:

$$DZC_h = ZCl_h \oplus ZCr_h \tag{3.22}$$

where the symbol $\oplus$ indicates logical Exclusive-OR operation. Thus, we can calculate DZC_h rate ($DZCR_h$) of the 8×8 block by

$$DZCR_h = \frac{1}{8 \times 8} \sum DZC_h \tag{3.23}$$

Therefore, the average $DZCR_h$ (AZC_h) for uniform, and non-uniform blocks of the left image are calculated by

$$AZC_{h_e} = \frac{1}{N_e} \sum_{e=1}^{N_e} DZCR_{h_e} \tag{3.24}$$

$$AZC_{h_n} = \frac{1}{N_n} \sum_{e=1}^{N_n} DZCR_{h_n} \tag{3.25}$$

where N_e, and N_n are respectively the number of uniform, and non-uniform blocks of the left image.

For vertical direction: similarly, we can calculate AZC_{v_e}, and AZC_{v_n}. Subsequently, the total disparity features for uniform, AZC_e and non-uniform, AZC_n areas are estimated by the following equation:

$$AZC_e = \frac{AZC_{h_e} + AZC_{v_e}}{2}, AZC_n = \frac{AZC_{h_n} + AZC_{v_n}}{2} \tag{3.26}$$

Finally, the overall disparity feature is estimated by

$$DZ \quad = \quad {AZC_e}^{w_5} \cdot {AZC_n}^{w_6} \tag{3.27}$$

$\cdot$ is the multiplication of two values where w_5, and w_6 are respectively the weighting factors of the disparity features for uniform, and non-uniform areas.

3.2.4 Features Combination

Finally, in my algorithm all features, both artifacts and disparity, are combined to construct a stereoscopic quality prediction model. The following equation is used to combine the artifacts and disparity features in my proposed NR stereo quality assessment metric:

$$S = \alpha(DZ) + \beta B \cdot Z \tag{3.28}$$

where α, and β are the model parameters. The model parameters and weighting factors (w_1 to w_6) must be estimated by an optimization algorithm with the subjective test data. The proposed method performance is also studied without disparity by the following equation:

$$S = \alpha + \beta B \cdot Z \tag{3.29}$$

I consider a logistic function according to VQEG recommendation as a non-linear property between the human perception and the physical features [35]. Finally, the obtained MOS prediction, MOS_p, is given by the following equation.

$$MOS_p = \frac{4}{1 + exp[-1.0217(S-3)]} + 1 \tag{3.30}$$

Here, the PSO algorithm is used to optimize the model's parameters and the weighting factors [19].

Chapter 4

Evaluation

To evaluate the performance of my proposed quality assessment model, I followed the standard performance evaluation procedures of VQEG for FR-TV Phase II testing [35]. The standard was developed for calculating the prediction error between a mathematical model and subjective scores (i.e., human viewers opinion). A short description of the standard performance evaluation procedures of VQEG is given at the beginning of this chapter. I consider the following important criteria for the evaluation of my proposed model:

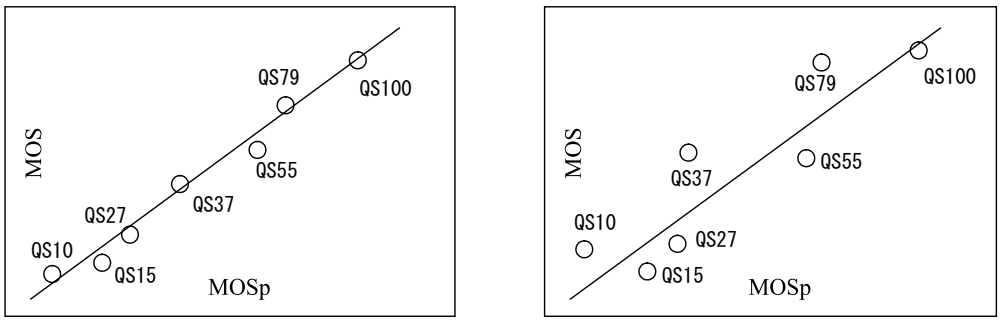
- Threshold estimation of my proposed block-based segmentation algorithm.
- Significance of considering the maximum blockiness between left and right images of a stereo pair.
- Significance of considering the minimum zero-crossing between the left and right images of a stereo pair.

- Effect of different block sizes and search areas on disparity estimation.
- Performance evaluation according to VQEG.
- Comparison with other methods.

4.1 VQEG's Standard Performance Evaluation Procedures

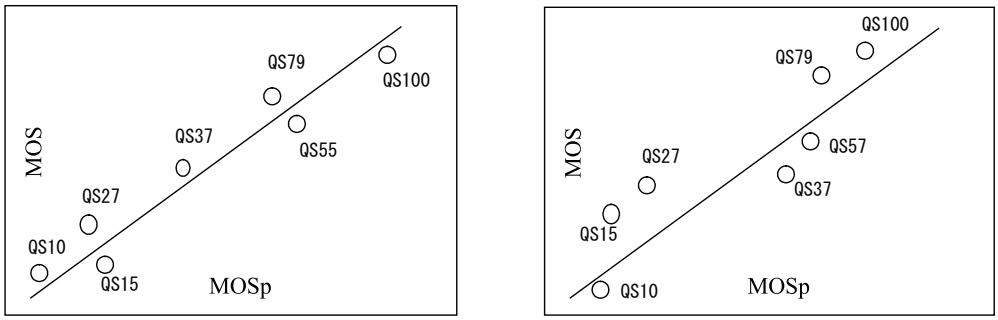
According to the VQEG [35], the performance of an objective quality model is characterized by three prediction attributes:

- Accuracy — is the ability to predict the distortions between MOS and MOSp. In an ideal case, the relationship between the MOS and MOSp is expected to be linear. Figure 4.1 illustrates the hypothetical relationships between the MOS and the MOSp for two models. Model-I is more accurate than the Model-II because most of the images evaluations are reasonably closer to the straight line.
- Monotonicity — is the degree to which the model's predictions agree with the relative magnitudes of subjective quality ratings. The prediction monotonicity is the extent of agreement between the subjective test and the objective model for variations in picture quality. As an example, viewers rank image A for many different levels of compressions where it implies the picture quality gets better when the level of compression is minimal. A monotonic objective model should give the same result, but it does not follow the trend even though they are



Model-I (accurate) Model-II (not accurate)

Figure 4.1: Two hypothetical models with different prediction accuracy [30]

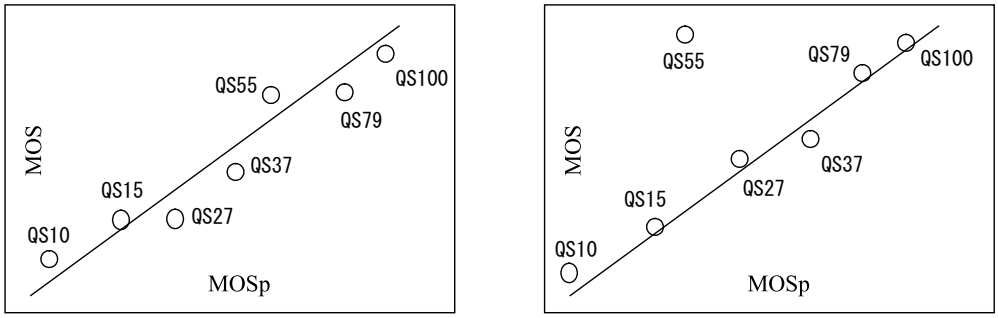


Model-I (non monotonic) Model-II (monotonic)

Figure 4.2: Two hypothetical models with different prediction monotonicity [30]

mathematically equivalent. Figure 4.2 illustrates the hypothetical relationships between the MOS and the MOSp for two models. Model-I has a better Pearson correlation than Model-II, but it falsely predicts degradation in picture quality in two events when the assessors actually see an improvement in picture quality. Therefore, in terms of monotonicity, Model-II is better than Model-I.

- Consistency — is the degree to which the model maintains prediction accuracy over the range of all types of images or for a subset of images. An objective



Model-I (consistent) Model-II (inconsistent)

Figure 4.3: Two hypothetical models with different prediction consistency [30].

model should perform well over a wide range of test images with minimum prediction error. Figure 4.3 shows two hypothetical models with MOS and the MOSp, and in terms of consistency, model-I is more consistent than model-II.

The followings are the performance evaluation metrics recommended by VQEG for objective quality assessment model:

Metric 1: Pearson Correlation Coefficient (CC) between objective (MOSp) and subjective (MOS) scores. MOSp is the mean opinion score prediction that is the output of objective (i.e., mathematical) model and whereas MOS is the mean opinion score of human assessments. This metric provides an evaluation of *prediction accuracy* which can be defined as:

$$CC = \frac{\sum_{i=1}^{N}\left(MOS(i) - \overline{MOS}\right)\left(MOS_p(i) - \overline{MOS_p}\right)}{\sqrt{\sum_{i=1}^{N}\left(MOS(i) - \overline{MOS}\right)^2}\sqrt{\sum_{i=1}^{N}\left(MOS_p(i) - \overline{MOS_p}\right)^2}} \tag{4.1}$$

where the index i denotes the image sample and N denotes the total number of samples.

Metric 2: Spearman Rank Order Correlation Coefficient (SROCC) between objective (MOSp) and subjective (MOS) scores. It is considered as a measure of *prediction monotonicity* and it is defined by:

$$SROCC = 1 - \frac{6\sum_{i=1}^{N}\left(MOS(i) - MOS_p(i)\right)^2}{N(N^2-1)} \tag{4.2}$$

where 6 is a constant (it is always used in the formula).

The prediction monotonicity is the extent of agreement between the subject test and the objective model in terms of the sign of change in picture quality.

Metric 3: Outlier Ratio (OR) represents number of "outlier-points" to the total points N. It is considered as a measure of *prediction consistency*, which can be defined by the following equation:

$$OR = \frac{(total\ number\ of\ outliers)}{N} \tag{4.3}$$

where an outlier is a point for: $|MOS(i) - MOS_p(i)| > 2\times\sigma(\text{MOS(i)})$,

where σ(MOS(i)) represents the standard deviation of the individual scores associated with the image sample i. The individual scores are approximately normally distributed and therefore twice the σ value represents the 95% confidence interval. Thus, $2\times$ σ(MOS(i)) value represents a good threshold for defining an outlier point. It is desirable that an objective model performs well over a wide range of test conditions. In other words, besides having the prediction error as small as possible, consistency in the magnitude of the prediction error is also preferable.

Metric 4: Average absolute prediction error (AAE) between objective (MOSp) and subjective (MOS) scores is defined by:

$$AAE = \frac{1}{N}\sum_{i=1}^{N}|MOS(i) - MOS_p(i)| \tag{4.4}$$

Metric 5: Root Mean Square Error (RMSE) between objective (MOSp) and subjective (MOS) scores is defined by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(MOS(i) - MOS_p(i)\right)^2} \tag{4.5}$$

Metric 4, and 5 are also considered as a measure of *prediction accuracy*.

An excellent objective model should exhibit good accuracy, monotonicity, and consistency in predictions. The measurement of prediction accuracy and monotonicity can be measured by Pearson correlation and Spearman rank order correlation metrics, whereas the consistency can be evaluated by the number of outlier points.

4.2 Threshold Estimation for Block-based Segmentation

Table 4.1: Number of edge pixels and n_e/Sum

n_e	6	7	8	9	10	11	12	13	14	15
n_e/Sum	0.094	0.109	0.125	0.141	0.156	0.172	0.186	0.203	0.219	0.234
th_n 3.2.1	–	0.10	0.10	0.13	0.13	0.16	0.16	0.19	0.19	0.22

The threshold value ("th_n") of my block-based segmentation algorithm is estimated empirically within the range of 0 and 1. I empirically selected five different "th_n"s (0.10, 0.13, 0.16, 0.19, and 0.22) for all images. Table 4.1 helps to describe the relationship between the number of edge pixels and the threshold values. The

table shows the total number of edge pixels, n_e and value of n_e/Sum for a block size of (8×8). According to the block-based segmentation algorithm 3.2.1, the threshold,"th_n" = 0.10 indicates that if a block contains 7 or more edge pixels, the block is a non-uniform block otherwise it is uniform. Similarly, thresholds, "th_n" = 0.13 and 0.16 indicate a block is non-uniform if the total number of edge pixels is more than 8 and 10, respectively. Using these threshold values ("th_n") the performances of block-based segmentation algorithm for two reference images (Cattle and Women from the MICT database [25]) are shown in Figures 4.4 and 4.5. The dark and the white blocks represent uniform and non-uniform areas, respectively, in the segmented images. The figures indicate sufficient segmentation performance. These figures also indicate that low level pixel variations areas are uniform areas (dark blocks) and the other areas are non-uniform areas (white blocks). The segmentations performance are empirically similar for each of the five different thresholds as can be seen in Figures 4.4 and 4.5. I estimated the best suitable threshold that helped to increase quality prediction. The optimum threshold for a metric can be estimated by comparing the performances of that algorithm for the different threshold values where the value of AAE and OR between MOS and MOSp is minimum [35]. In order to identify the best suitable algorithmic threshold between the five values, I consider the MICT's subjective stereo images database and divided the database into two parts for training and testing. The training database consists of five randomly selected reference stereo pairs (from the total ten) and all of their different combinations of symmetric/asymmetric coded stereo images (245 stereo pairs). The testing database consists of the other five reference stereo pairs and their symmetric/asymmetric coded

Table 4.2: Threshold performances based on AAE and OR

th_n	*Training*		*Testing*	
	AAE	OR	AAE	OR
0.07	0.297	0.065	0.338	0.049
0.10	0.289	0.061	0.339	0.040
0.13	**0.276**	**0.061**	**0.334**	**0.040**
0.16	0.291	0.069	0.347	0.049
0.19	0.317	0.078	0.350	0.057

versions (245 stereo pairs). There is no overlap between training and testing. A ten-fold cross validation [20] was used to generate the training and testing datasets. I compared the performance of my algorithm for all these different threshold values.

Since the goal of the segmentation is to improve quality prediction, I have estimated all weighting factors (w_1 to w_6) and method parameters (α, and β) for each threshold value separately by an optimization algorithm within the training images database. The optimization algorithm is used to find the best suitable value from the sets of available alternatives. Here the Particle Swarm Optimization (PSO) [19] algorithm was used to optimize the parameters and weighting factors. It is one of the popular optimization techniques. Though genetic algorithms have proven to be a useful method of optimization for difficult and discontinuous multidimensional engineering problems, in many cases PSO is able to accomplish the same goal as genetic algorithm within a shorter calculation time and with stable convergence characteristics [8]. Like genetic algorithms, PSO must also have a fitness evaluation function that takes the agent's position and assigns to it a fitness value. Subsequently, I have calculated the AAE and OR for all training and testing images that are shown in Table 4.2. Therefore, for all cases (in Table 4.2), the best performance was obtained for

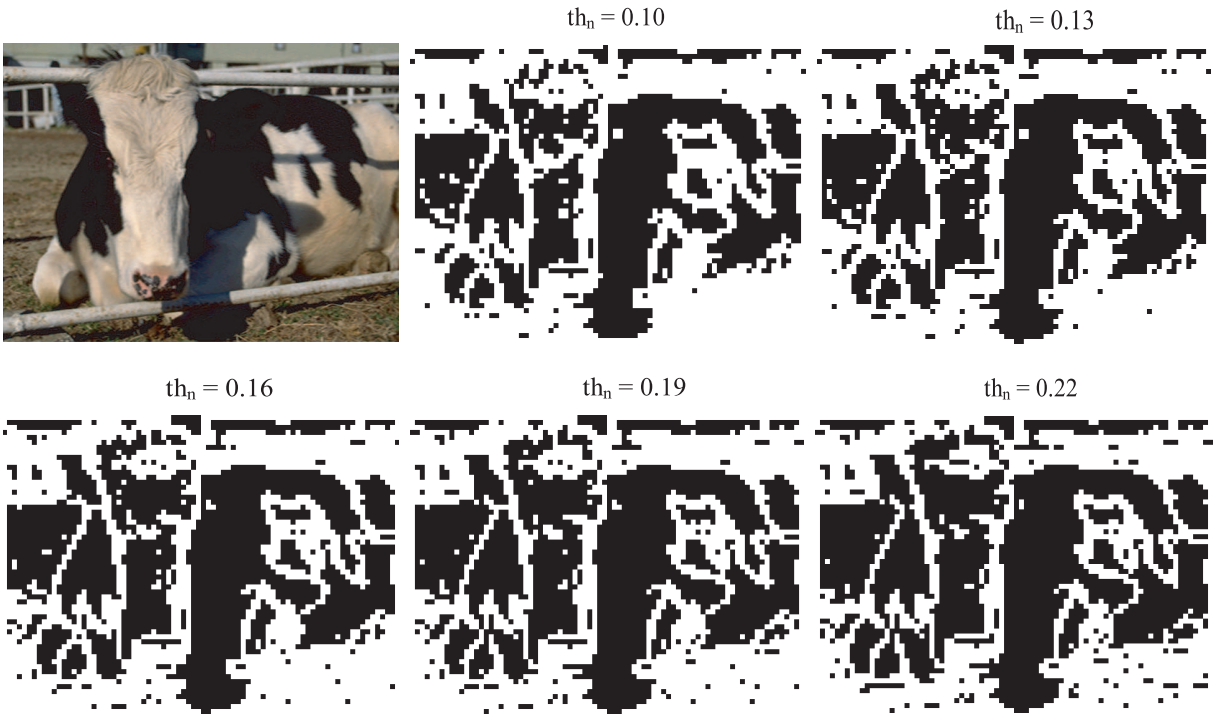


Figure 4.4: Segmented Cattle image for various thresholds. Reference Cattle image from the MICT database [25]

a threshold value of "th_n" = 0.13. The obtained model's parameters and weighting factors by the PSO algorithm for the threshold value of my training database (MOS scale, 1-5) are shown in Table 4.3. I selected the threshold "th_n" = 0.13 based on MOSp performances of the training and testing image databases which indicated the highest quality prediction performance (see Table 4.2).

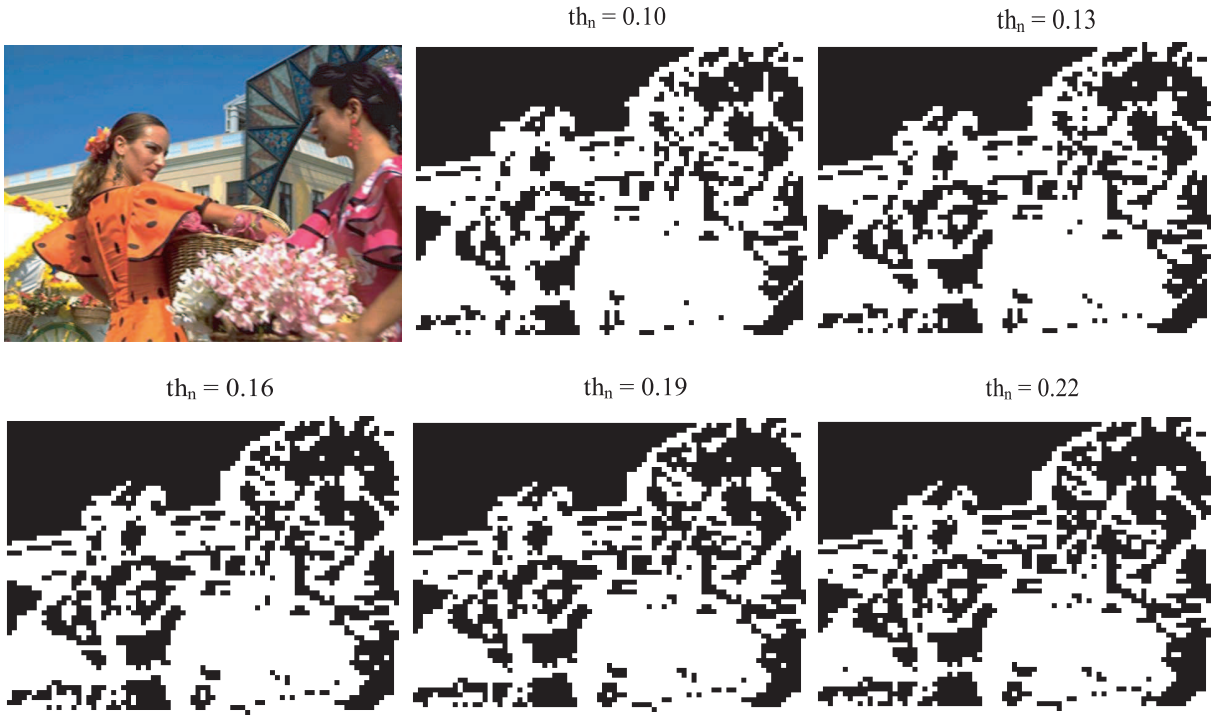


Figure 4.5: Segmented Women image for various thresholds. Reference Women image from the MICT database [25]

4.3 Significance of considering the maximum blockiness of a stereo pair

In this section, I discuss the reason for choosing the maximum blockiness of a stereo pair of my objective model. The goal is to measure the maximum possible blockiness within a stereo pair so, the metric can correlate well with human viewers perception without actual human input. Moreover, the model is developed both for symmetric and asymmetric images. In order to count the highest degradation, I consider the maximum blockiness between the left and the right views. In order to explain the consideration of the maximum blockiness, I took three stereo images,

Cattle, Women, and Saboten (stereo images from the MICT database [25]) for my description. Coding levels versus blockiness of the three types of stereo images are shown in Figures 4.6, 4.7, and 4.8. I consider both the highest and average blockiness between the two views. These Figures show variations of blockiness with the increasing of bit rate. The highest and the average blockiness are considered between the two views. The results show that the blockiness variation is higher in case of maximum number of blockiness compared to that of the average blockiness for increasing of bit rate. The normalized MOS (NMOS) versus blockiness (N-blockiness) with increasing bit rate for two types of stereo images are shown in Figures 4.9, and 4.10. The Coding levels (L,R: 79-10, 79-15, 79-27, 79-37, 79-55, 79-79, 79-Ref), and (L,R: Ref-10, Ref-15, Ref-27, Ref-37, Ref-55, Ref-79, Ref-Ref) in the Figures 4.9, and 4.10 indicate increasing bit rate. Although NMOS scores show an increasing trend with decreasing N-blockiness, the consideration of maximum blockiness correlates better with NMOS compared to average blockiness. Therefore, the results above suggest that consideration of the maximum blockiness is more justified than the average blockiness for developing an objective model.

4.4 Significance of Considering the minimum zero-crossing of a Stereo Pair

An analysis for choosing the minimum zero-crossing value between the left and the right of a stereo pair is given in this section. Earlier in this chapter (see Section 3 for artifacts measure), it has been discussed that the average edge point detection

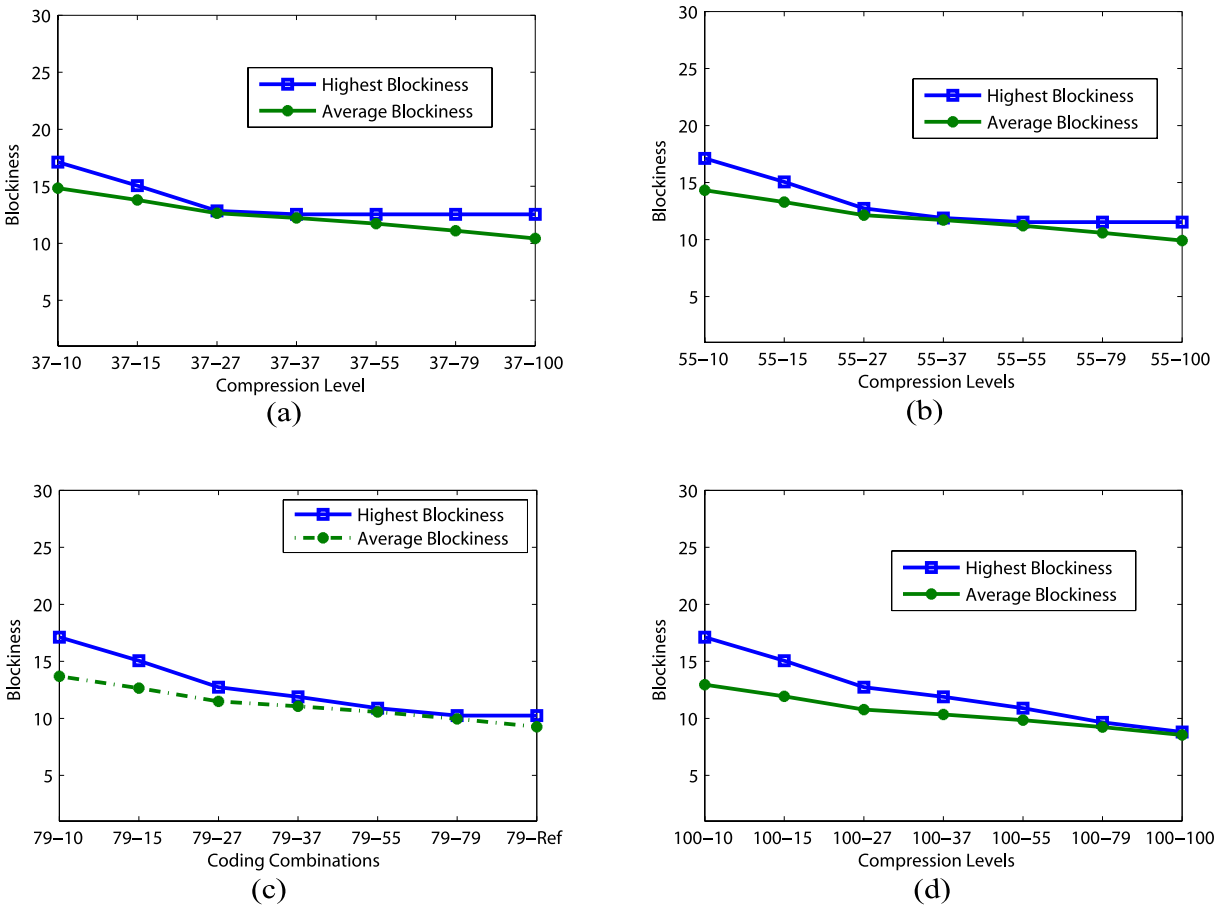


Figure 4.6: Blockiness versus different coding levels for Cattle image pairs

within the image blocks give better insight of blur measurement within an image. Consequently, the zero-crossing scores show a decreasing (i.e., increasing blur) trend with the increasing compression rate. Therefore, there is a direct relationship with the transition of zero-crossing and the over-all blur within an image. In order to study the relationship, I took three types of stereo images, Cattle, Women, and Saboten (stereo images from the MICT database [25]). Normalized MOS (NMOS) versus zero-crossing (Nzero-crossing) of the three types of stereo images are shown in Figures 4.11, 4.12,

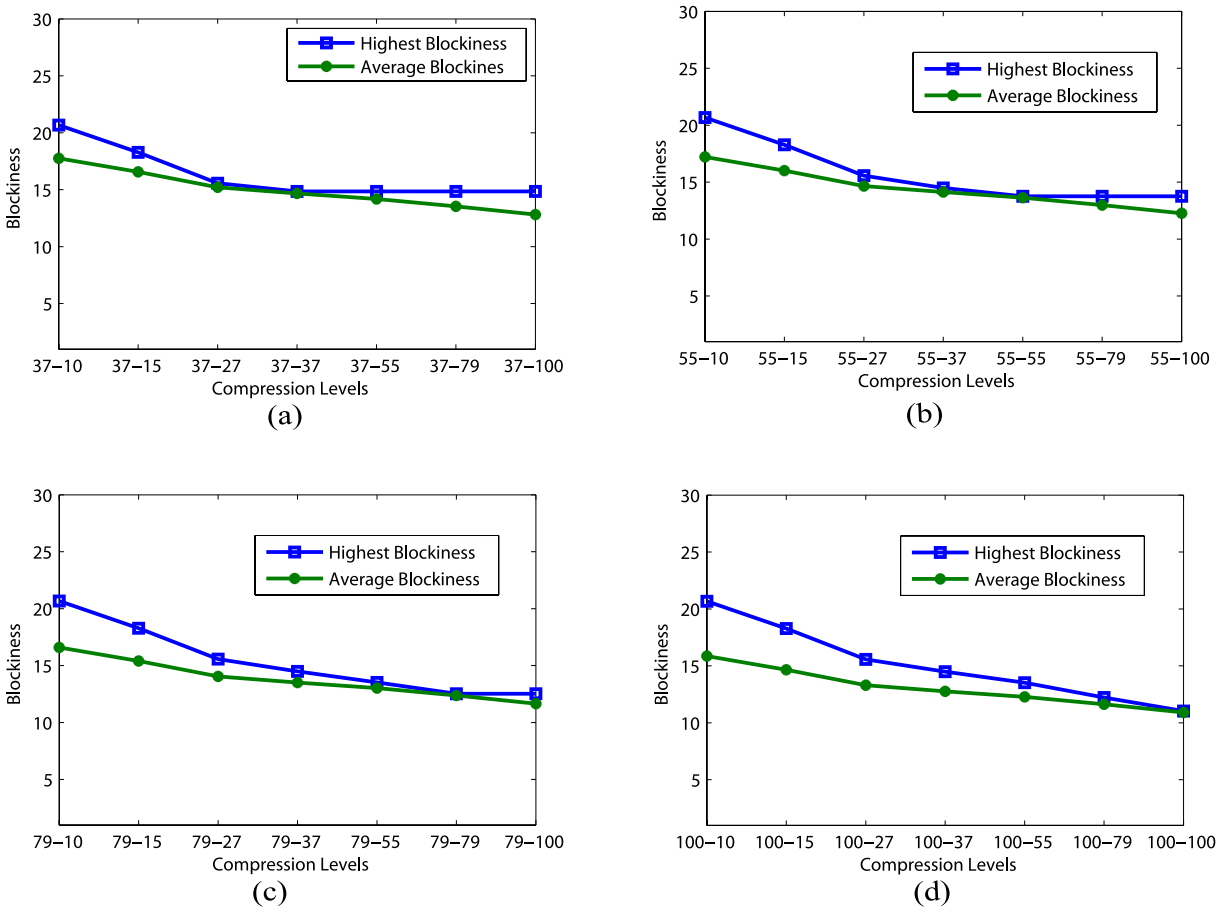


Figure 4.7: Blockiness versus different coding levels for Women image pairs

and 4.13. I considered both the minimum and the average zero-crossing value of the stereo pairs. These Figures show that the minimum zero-crossing measures are closer to the NMOS score compared to that of the average zero-crossing. In addition, the zero-crossing scores show an increasing trend for increasing bit-rate (i.e., high compression ratio). Therefore, the results indicate that the minimum zero-crossing is more justified than the average zero-crossing to develop the prediction metrics.

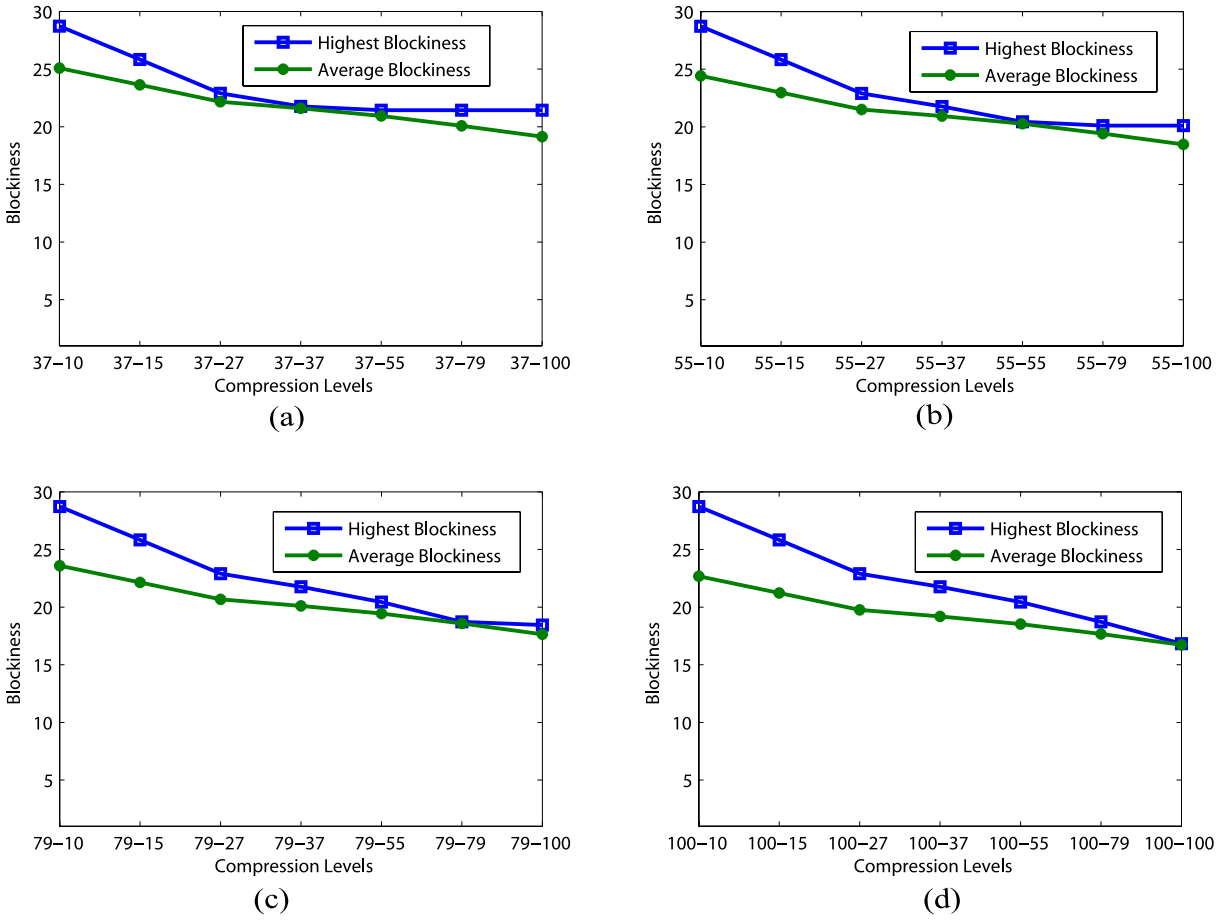


Figure 4.8: Blockiness versus different coding levels for Saboten image pairs.

4.5 Effect of different block size and search areas on disparity estimation

In order to identify the most suitable block-size and search area for my block-based disparity estimation, I compared the effect of different block sizes and search areas on disparity estimation. I considered three different block sizes, 4×4, 8×8, and 16×16, and three different search areas, ±32 pixels, ±64 pixels, and ±128 pixels for

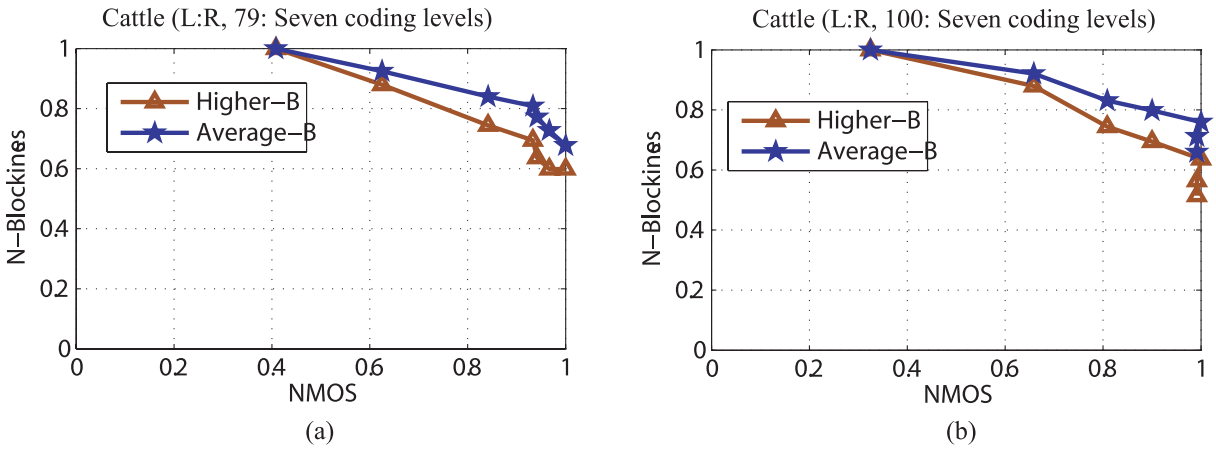


Figure 4.9: Normalized MOS versus blockiness for different Cattle image pairs

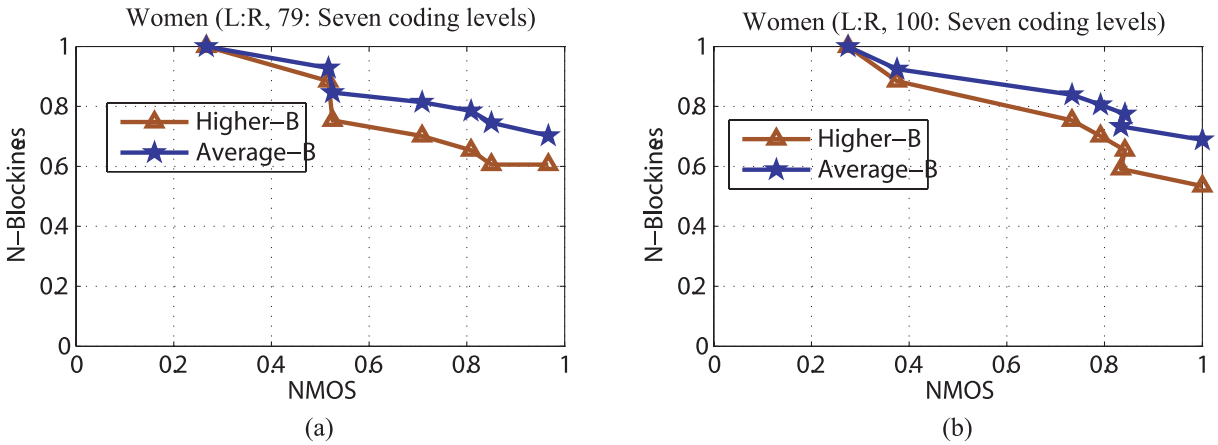


Figure 4.10: Normalized MOS versus blockiness for different Women image pairs

my objective model evaluation. I used the Cattle and Peacock stereo image pairs in my explanation (stereo images from the MICT database [25]). The disparity maps of reference Cattle pair for block size 4×4, 8×8, and 16×16 with all three search areas are shown in Figures 4.14, 4.15, and 4.16, respectively. Similarly for the Peacock pair, the disparity maps are shown in Figures 4.17, 4.18, and 4.19, respectively. The grey

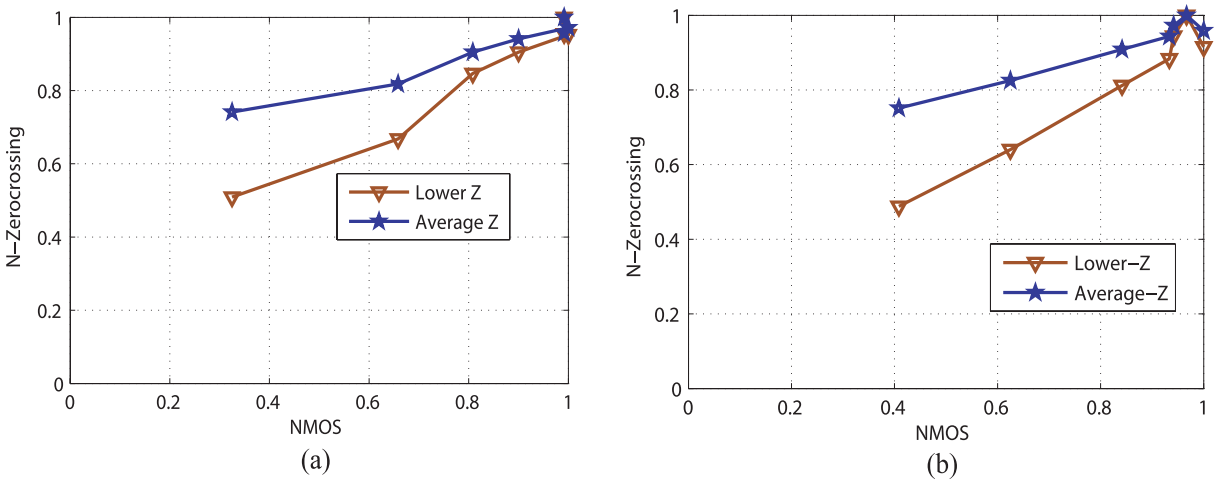


Figure 4.11: Normalized MOS versus zero-crossing of different Cattle image pairs for coding levels: (a) L,R: Ref-10, Ref-15, Ref-27, Ref-37, Ref-55, Ref-79, Ref-Ref, (b) L,R: 79-10, 79-15, 79-27, 79-37, 79-55, 79-79, 79-Ref

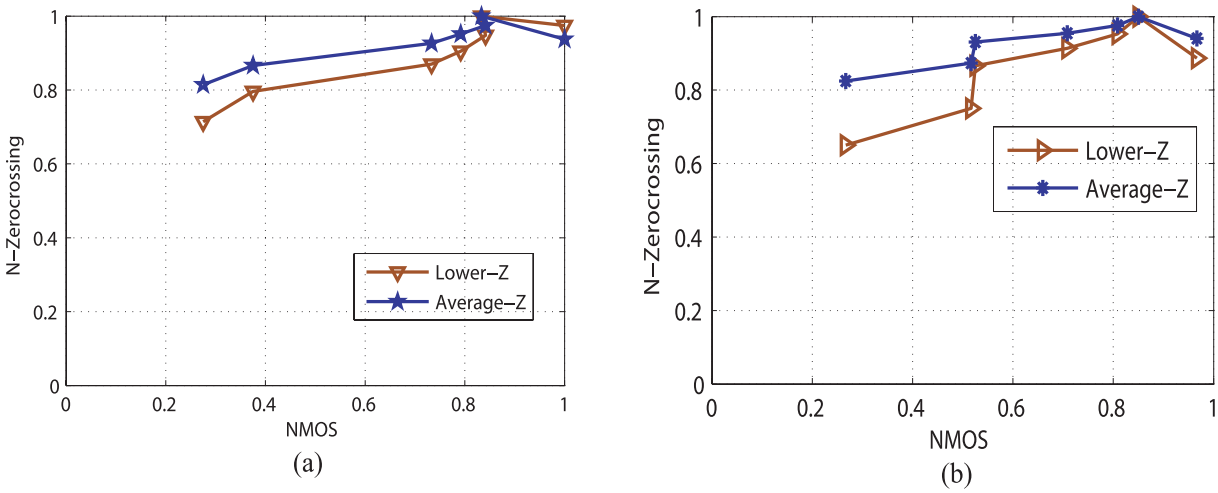


Figure 4.12: Normalized MOS versus zero-crossing of different Women image pairs for coding levels: (a) L,R: Ref-10, Ref-15, Ref-27, Ref-37, Ref-55, Ref-79, Ref-Ref, (b) L,R: 79-10, 79-15, 79-27, 79-37, 79-55, 79-79, 79-Ref

colors in the disparity maps are indicated by vertical bars in the right side for each of the figures and estimated depths of the image pairs. Here suitable block sizes and search area are estimated empirically. These Figures show that the performance of

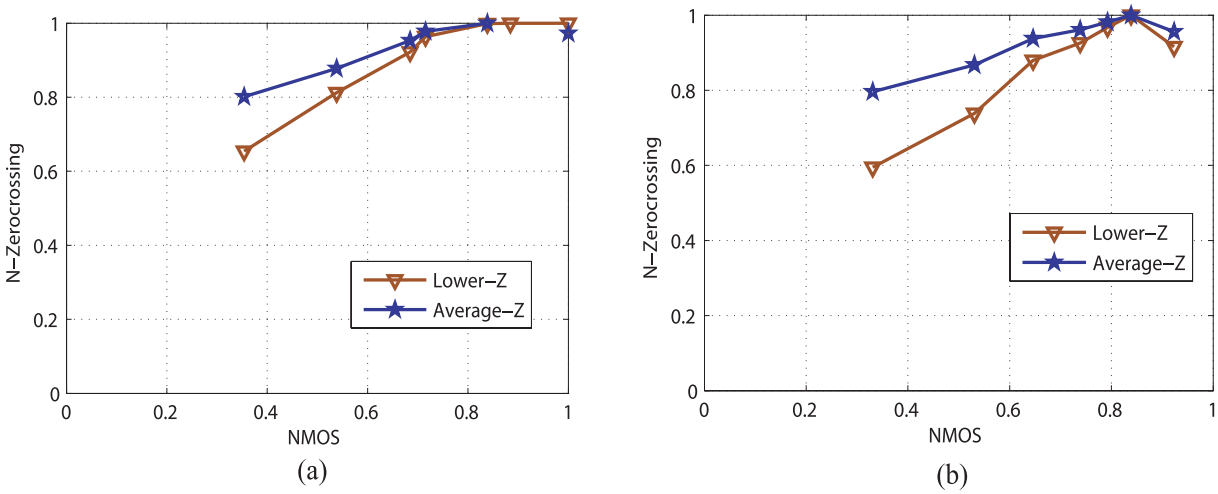


Figure 4.13: Normalized MOS versus zero-crossing of different Saboten image pairs for coding levels: (a) L,R: Ref-10, Ref-15, Ref-27, Ref-37, Ref-55, Ref-79, Ref-Ref, (b) L,R: 79-10, 79-15, 79-27, 79-37, 79-55, 79-79, 79-Ref

the disparity algorithm is sufficient for the block size 8×8 with search area of ±128 pixels. The disparity maps of symmetric and asymmetric coded stereoscopic pairs for Cattle images are respectively shown in Figures 4.20, 4.21, A.10, and 4.23. Figures 4.20, and 4.21 show that the performance of disparity estimation is degraded with increasing compression levels (i.e., decreasing in bit rate) for symmetric coded pairs. These two figures indicate that the performance of symmetric pair (10-10) is lowest compared to the pairs (27-27), (55-55), and (Ref-Ref). Figures 4.20, 4.21, A.10, and 4.23 are also indicated that disparity estimation performance is better for symmetric pairs compared to asymmetric pairs. Figures A.10, and 4.23 also show that estimation performance is degraded for increasing level of compression levels. For example, in Figure A.10 the disparity estimation performance of the asymmetric pair (Ref-79) is better than the asymmetric pair (55-37).

Although the block-based disparity algorithm is simple but the problem arises to

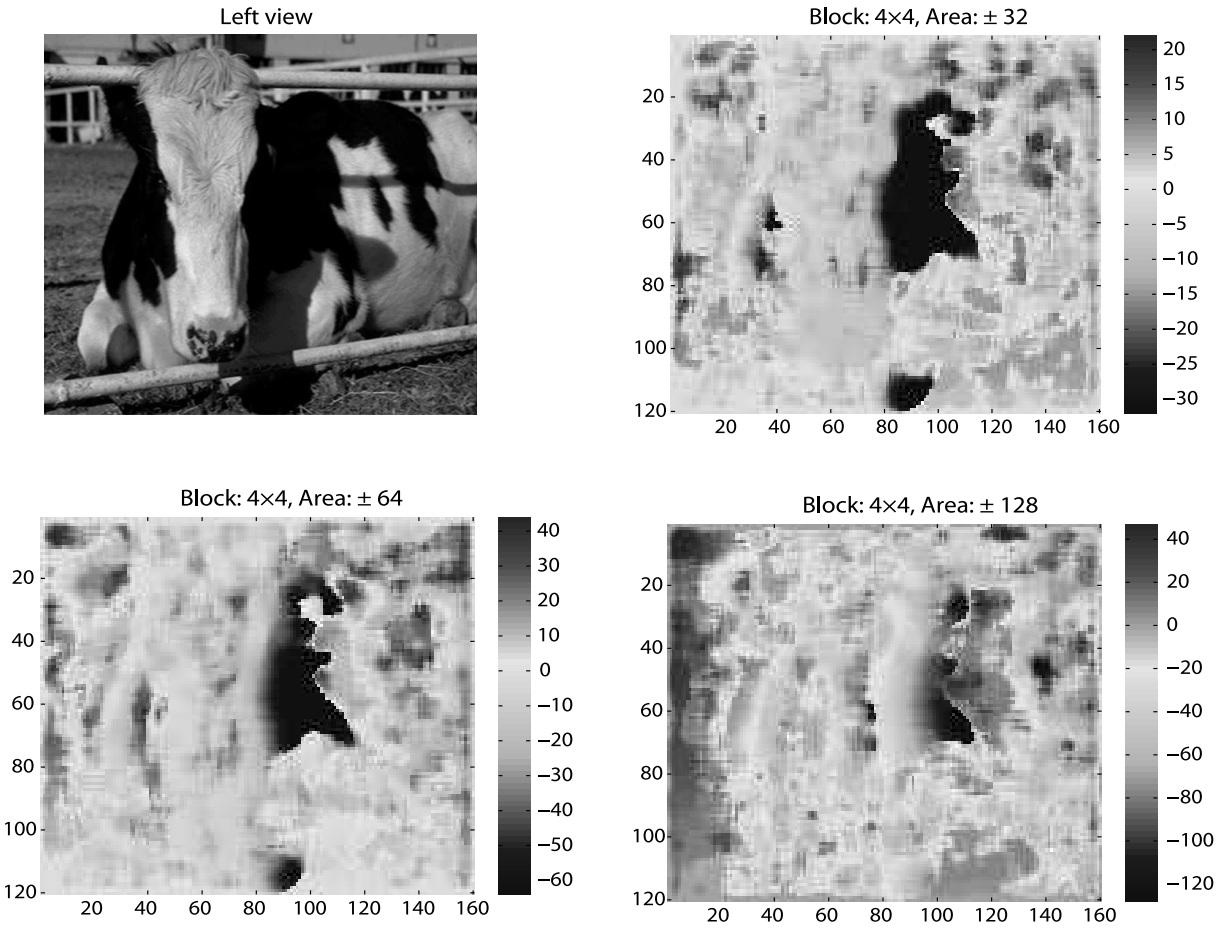


Figure 4.14: Disparity map of reference Cattle image (symmetric pair) for block size, 4×4 with the three different search areas. Cattle image form the MICT database [25]

localize the disparities accurately. The large search space increases the probability of sufficient intensity variation but introduces ambiguities and mismatches. Therefore, it reduces the selectivity of finding correct disparities, especially in the areas with depth discontinuities. Furthermore, the edge points detection may be a more suitable measure irrespective of search area but this idea might not be useful in real-time disparity estimation. On the other hand, my algorithm in unstructured regions (e.g., uniform areas) may lead to ambiguity and mismatches because the same region in the

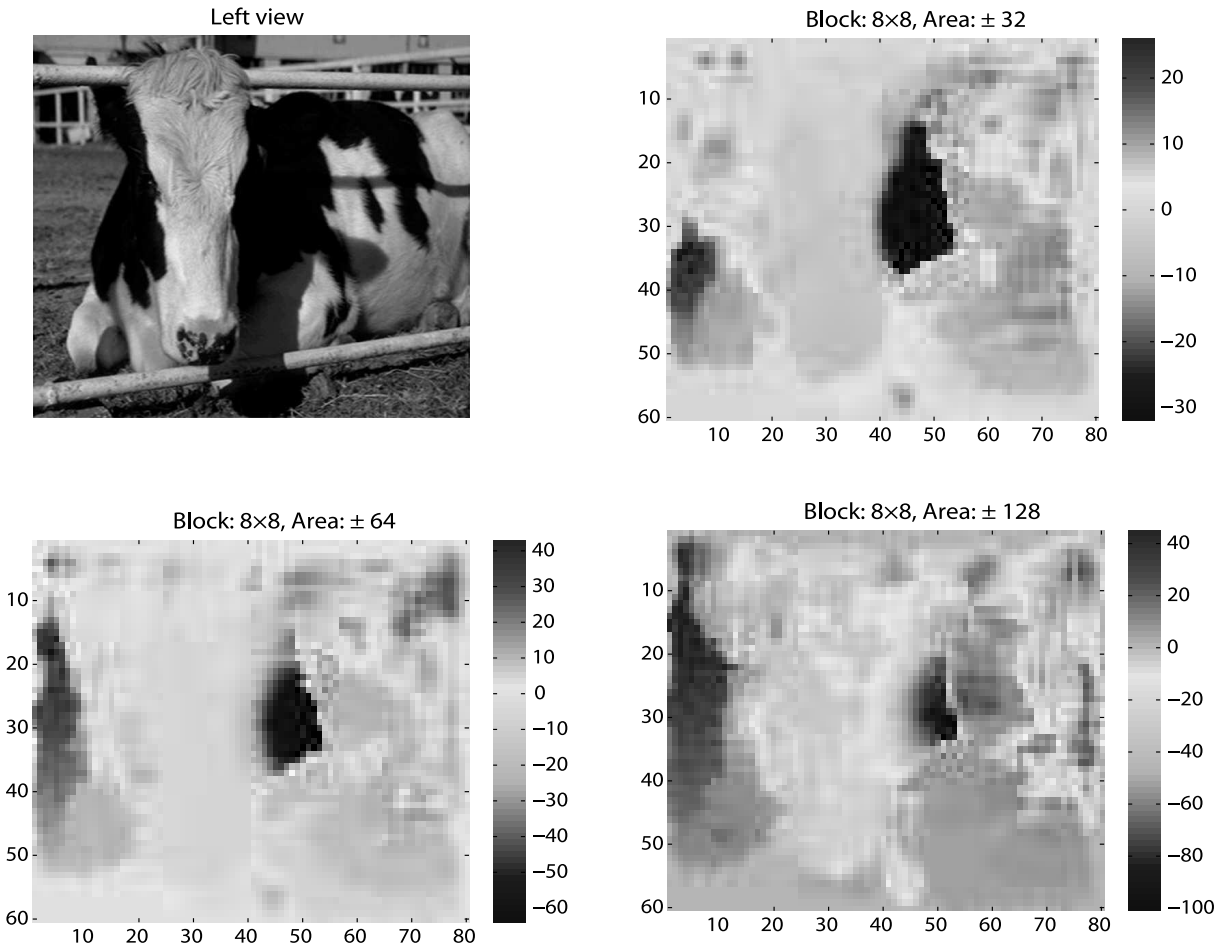


Figure 4.15: Disparity map of reference Cattle image (symmetric pair) for block size, 8×8 with the three different search areas. Cattle image form the MICT database [25]

left image will match equally well with many other similar blocks in the right image.

4.6 Model Performance Evaluation According to VQEG

To measure the prediction performance of my objective model qualitatively, I follow the standard performance evaluation procedure recommended in VQEG [35],

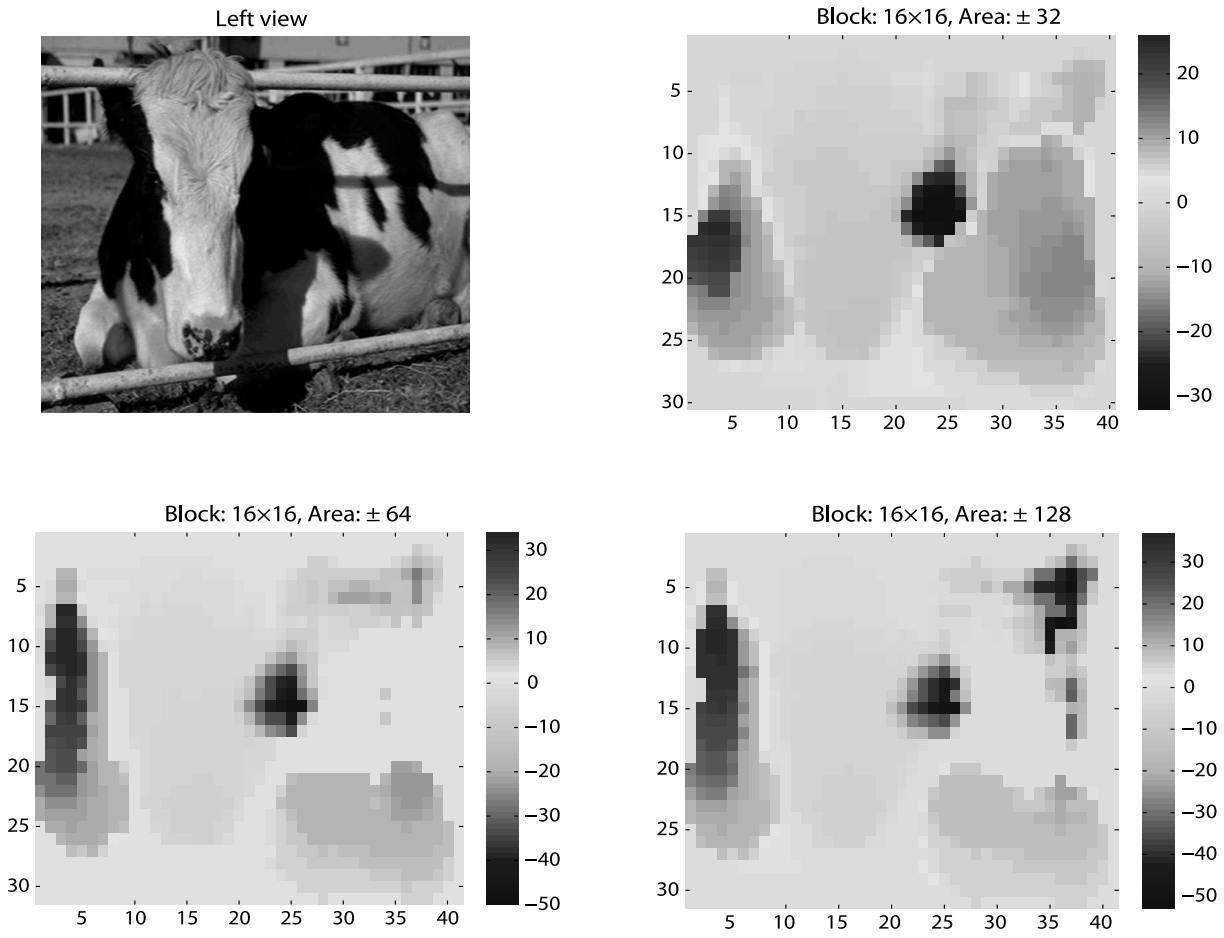


Figure 4.16: Disparity map of reference Cattle image (symmetric pair) for block size, 16×16 with the three different search areas.

where mainly linear correlation coefficient (CC), average absolute prediction error (AAE), root mean square prediction error (RMSE), and outlier Ratio (OR) between predicted objective scores (MOSp) and subjective scores (MOS) were used for evaluation. In order to verify the permanence of my model, I consider the MICT stereoscopic image database. The database is divided into two parts for training and testing (see Section 4.2). The method's parameters and weighting factors are obtained for the quality scales (scale, 1-5) by using the PSO algorithm for all of the training images

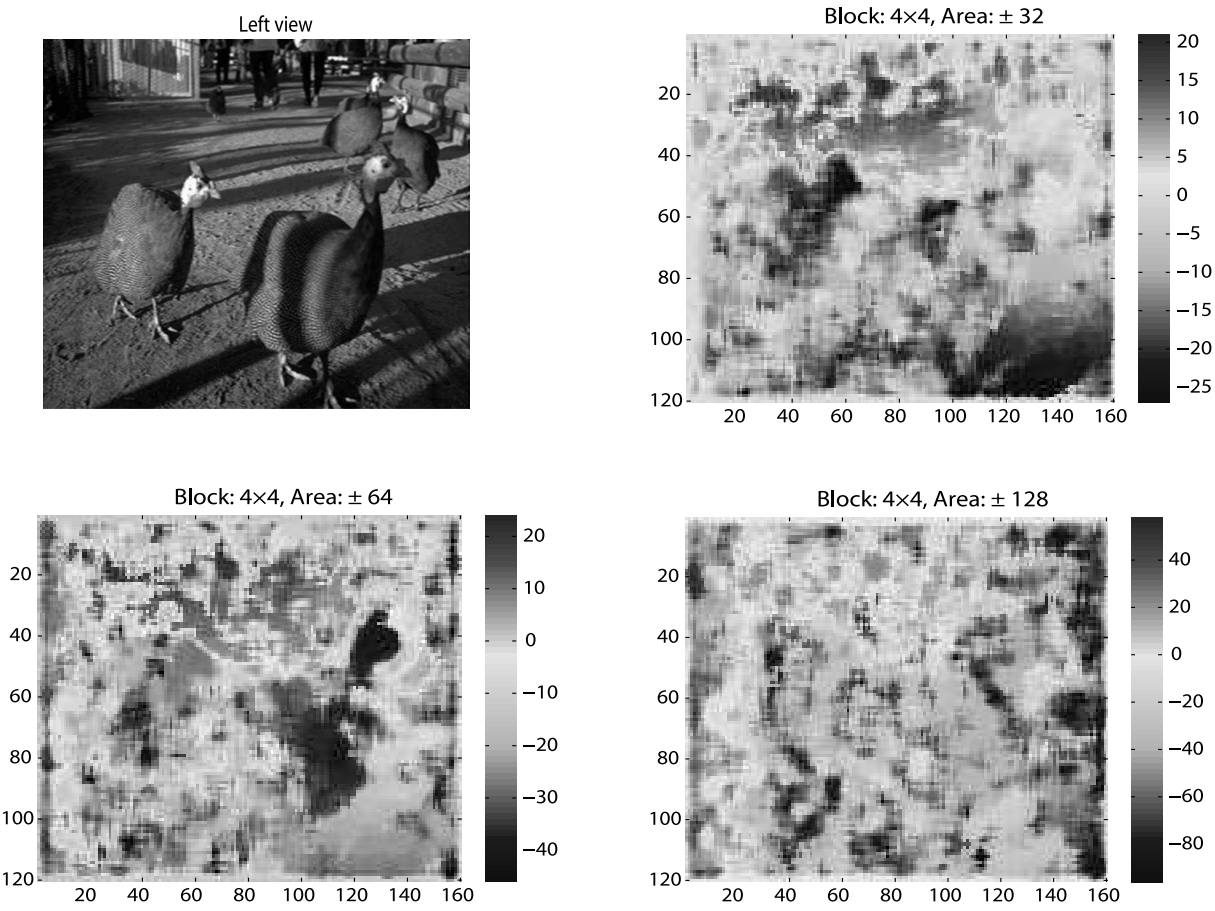


Figure 4.17: Disparity map of reference Peacock image (symmetric pair) for block size, 4×4 with the three different search areas. Peacock image form the MICT database [25]

Table 4.3: Method parameters and weighting factors for quality scale, 1-5

α = 58.064452	β = -51.026118	
w_1 = 0.036062	w_2 = 0.00513	w_3 = 0.010634
w_4 = -0.026979	w_5 = -0.017522	w_6 = 0.013169

shown in Table 4.3. Same parameters are chosen for the rest of the pairs within the image database for testing. In order to measure the performance of my proposed model quantitatively as well as justification of the estimated image features in the model, I consider the following prediction performances:

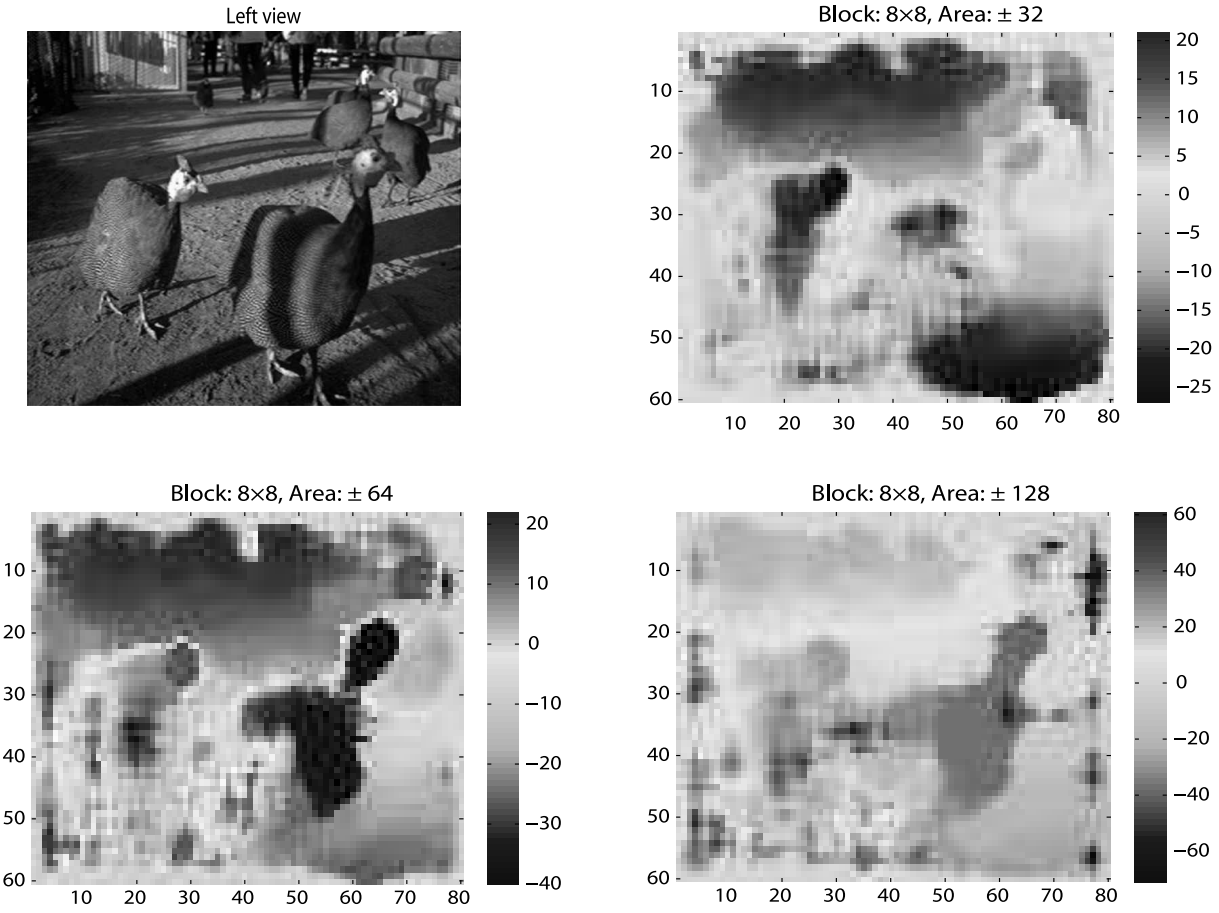


Figure 4.18: Disparity map of reference Peacock image (symmetric pair) for block size, 8×8 with the three different search areas.

- Method with disparity
  - Proposed model (i.e., considering blockiness, zero-crossing, and disparity) using the features combining Equation 3.28
  - Method considering only blockiness and disparity using the following equation to combine the individual features:

$$S = \alpha(DZ) + \beta B \tag{4.6}$$

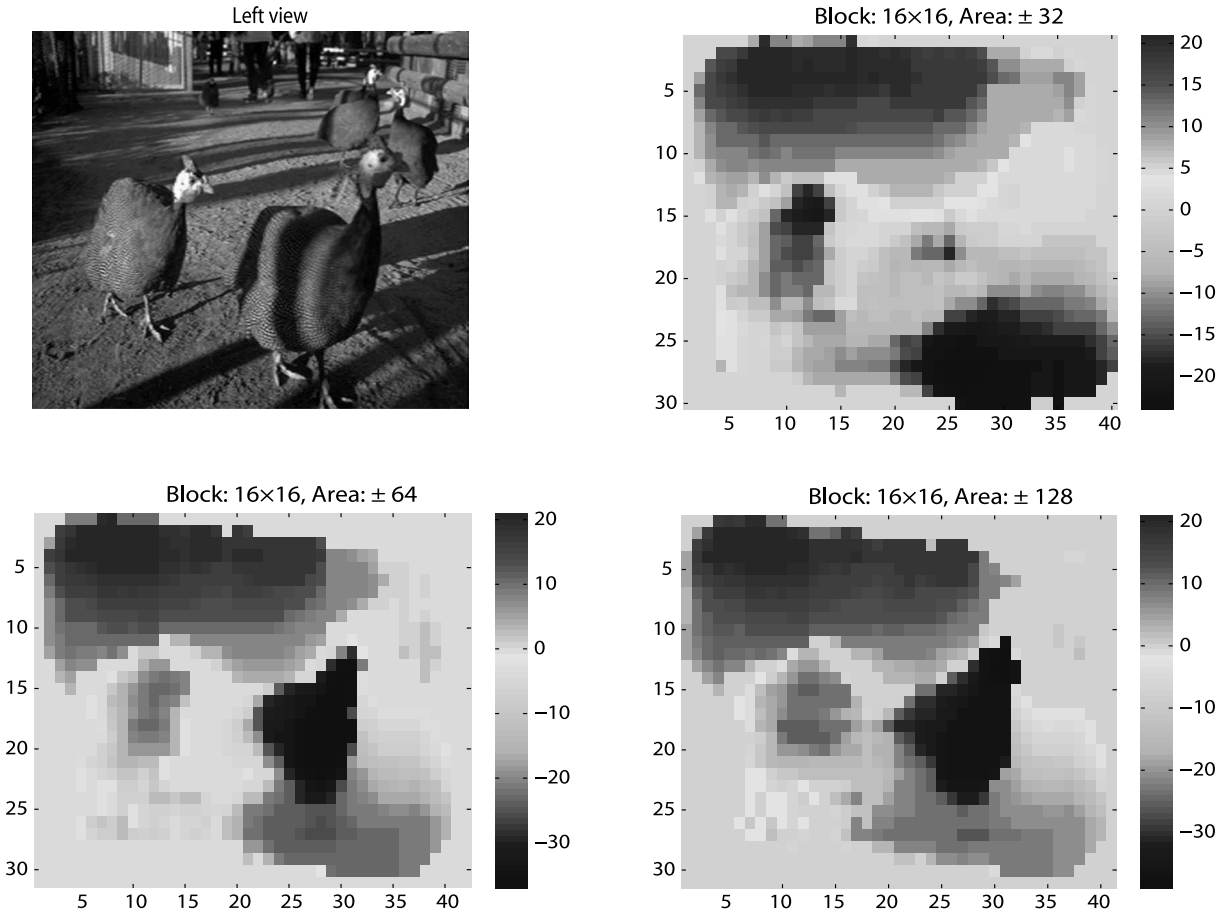


Figure 4.19: Disparity map of reference Peacock image (symmetric pair) for block size, 16×16 with the three different search areas.

– Method considering only zero-crossing and disparity using the following equation to combine the individual features:

$$S = \alpha(DZ) + \beta Z \tag{4.7}$$

- Method without disparity

– Method considering blockiness, and zero-crossing using the following equation to combine the individual features 3.29

Method considering only blockiness by using the following features com-

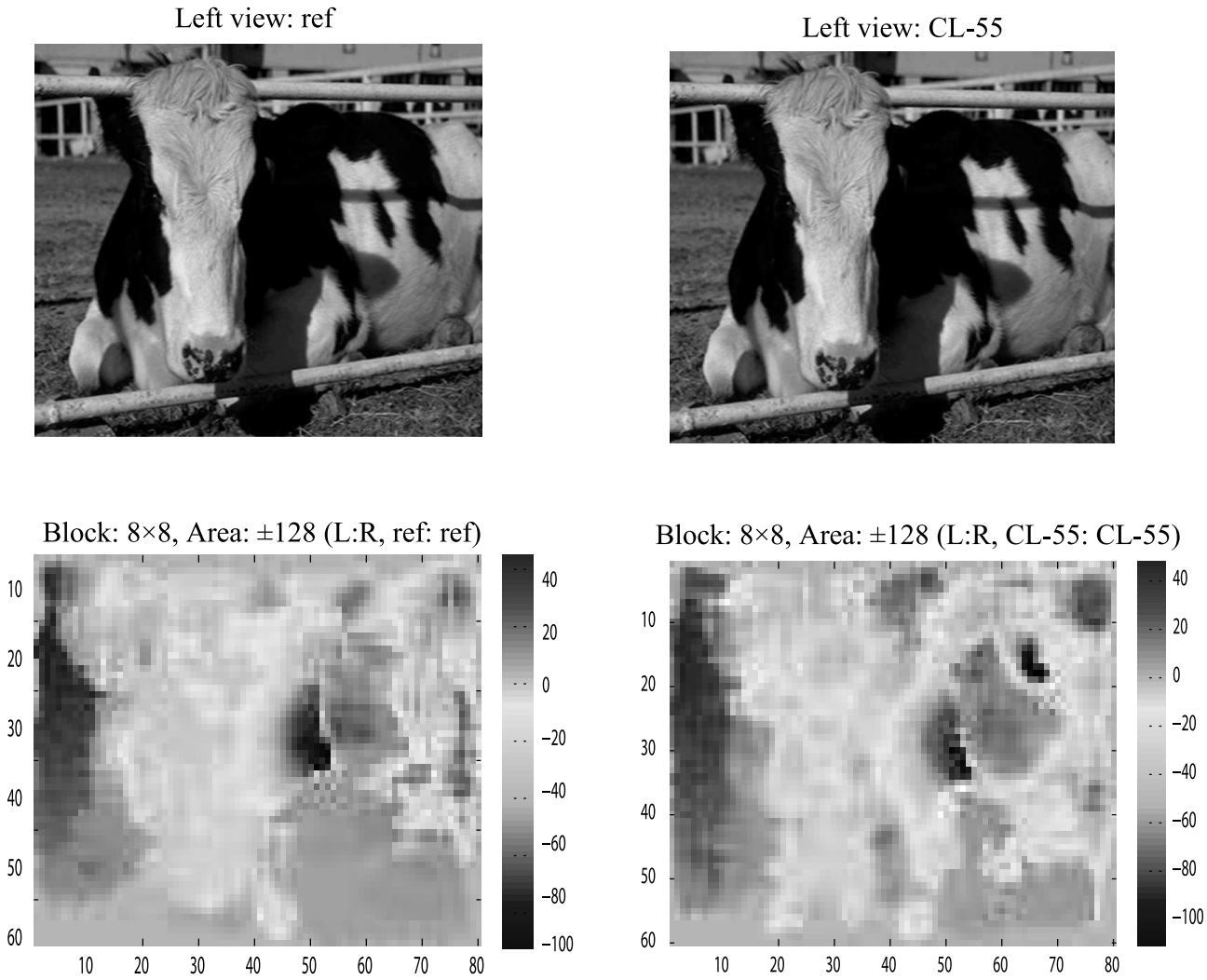


Figure 4.20: Disparity map of reference and coded Cattle image (symmetric pairs) for block size, 8×8 and search areas, ±128 pixels.

bining equation:

$$S = \alpha + \beta B \tag{4.8}$$

– Method considering only zero-crossing using the following equation to combine the individual features:

$$S = \alpha + \beta Z \tag{4.9}$$

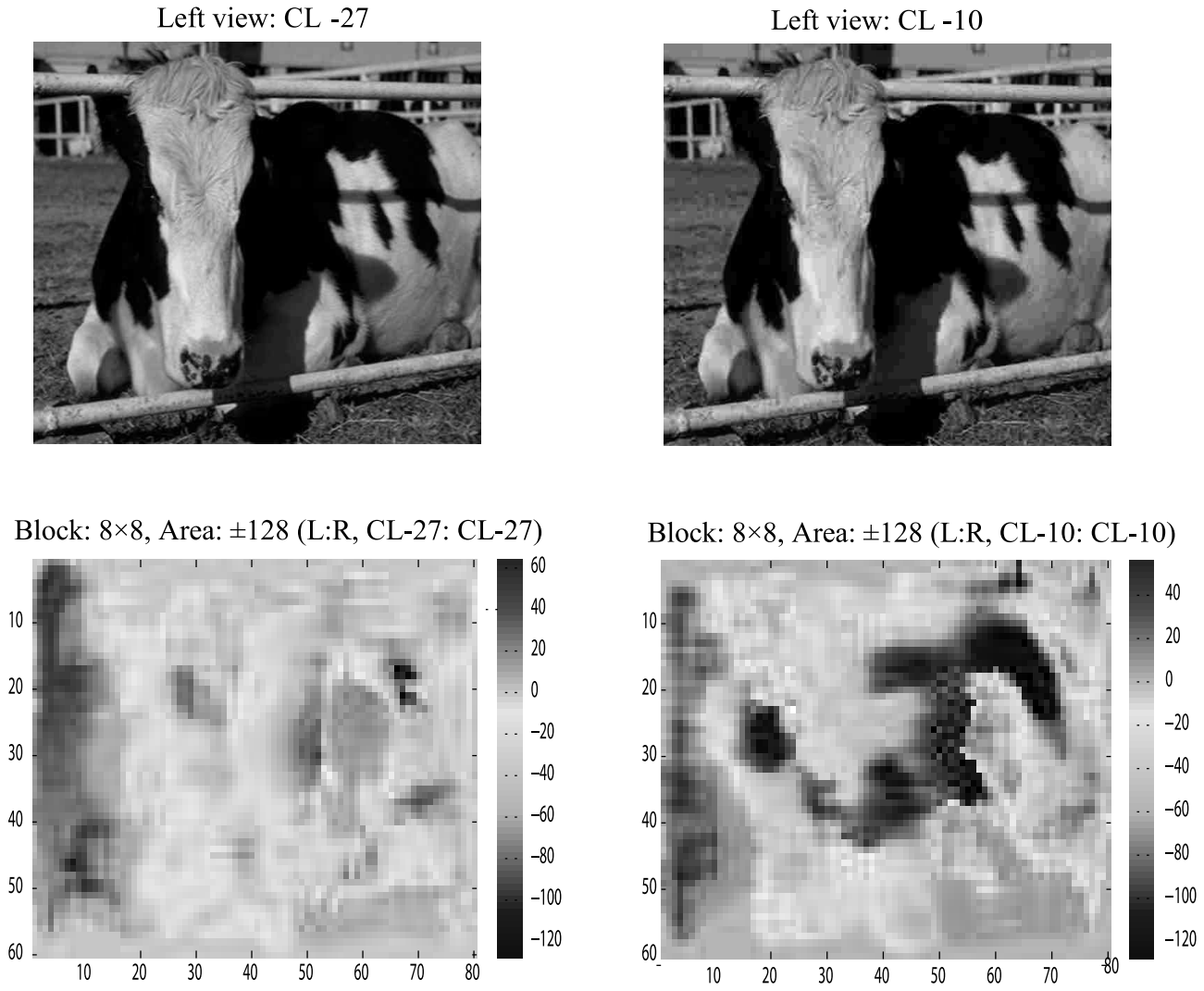


Figure 4.21: Disparity map of coded Cattle image (symmetric pairs) for block size, 8×8 and search areas, ±128 pixels.

- Other methods

  – Method considering blockiness, zero-crossing, and disparity with linear weighting (i.e., linear weighting approach) by the following equation to combine the individual features:

$$S = \alpha(DZ) + \beta B \cdot Z \tag{4.10}$$

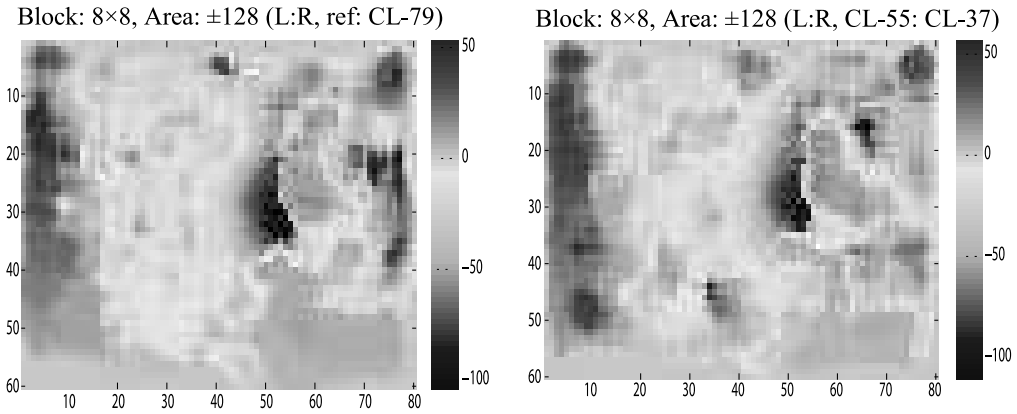


Figure 4.22: Disparity map of coded Cattle image (asymmetric pairs) for block size, 8×8 and search areas, ±128 pixels.

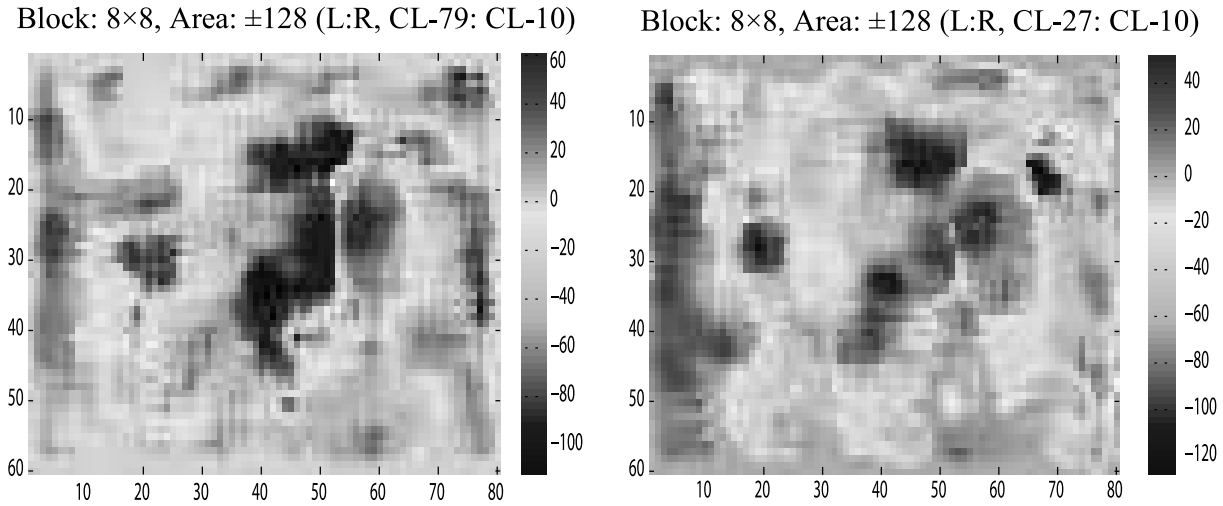


Figure 4.23: Disparity map of coded Cattle image (asymmetric pairs) for block size, 8×8 and search areas, ±128 pixels.

where,

$$B = w_1 \cdot B_e + w_2 \cdot B_n \tag{4.11}$$

$$Z = w_3 \cdot ZC_e + w_4 \cdot ZC_n \tag{4.12}$$

and

$$DZ = w_5 \cdot AZC_e + w_6 \cdot AZC_n \tag{4.13}$$

– Method considering the blockiness and zero-crossing of a stereo pair and measure the quality score of the left and the right views independently, and average them without disparity, "2D quality mean" [13]

The evaluation results of all above mentioned methods are summarized in Tables 4.4, 4.5, and 4.6. The Table shows that the proposed model's performances for every one of the evaluation metrics are quite sufficient both for the training and the testing datasets. It has also been observed from the Table 4.4 that the proposed method provides sufficient prediction accuracy (higher CC), and sufficient prediction consistency (lower OR). Tables 4.4 and 4.5 also show that the model performances are superior compared to without disparity. Whereas, "2D quality mean" performance is not sufficient even when compared to without the disparity approach (i.e., considering only blockiness and zero-crossing) (see Tables 4.5, and 4.6).

Table 4.4: Methods' evaluation results for training and testing (Scale, 1-5) with disparity

Methods	*Training*			
	CC	AAE	RMSE	OR
Proposed model	0.964	0.276	0.336	0.061
Only blockiness with disparity	0.867	0.529	0.664	0.086
Only zero-crossing with disparity	0.897	0.484	0.578	0.110
	Testing			
Proposed model	0.940	0.339	0.413	0.040
Only blockiness with disparity	0.833	0.537	0.724	0.106
Only zero-crossing with disparity	0.804	0.578	0.707	0.102

Table 4.5: Methods' evaluation results for training and testing (Scale, 1-5) without disparity

Methods	*Training*			
	CC	AAE	RMSE	OR
Blockiness and zero-crossing without disparity	0.953	0.322	0.401	0.074
Only blockiness without disparity	0.705	1.037	1.199	0.367
Only zero-crossing without disparity	0.883	0.515	0.610	0.110
	Testing			
Blockiness and zero-crossing without disparity	0.932	0.349	0.432	0.053
Only blockiness without disparity	0.705	1.041	1.196	0.359
Only zero-crossing without disparity	0.814	0.595	0.719	0.114

Table 4.6: Other methods' evaluation for training and testing (Scale, 1-5)

Methods	*Training*			
	CC	AAE	RMSE	OR
Linear weighting with disparity	0.843	0.580	0.705	0.089
Linear weighting without disparity	0.830	0.561	0.656	0.114
2D quality mean	0.912	0.432	0.55	0.078
	Testing			
Linear weighting with disparity	0.804	0.578	0.707	0.102
Linear weighting without disparity	0.765	0.649	0.828	0.183
2D quality mean	0.89	0.40	0.534	0.057

Although, the incorporation of disparities measure to the FR stereo quality method [5] indicate poor results, my proposed method (with relative disparity) indicate better results when compared to without disparity (i.e., considering both blockiness and zero-crossing). It is clear from Tables 4.4, and 4.5 that all methods performances with disparity are superior compared to without disparity. Therefore, the relative disparity measure which is considered in my proposed method can be a significant measure for 3D quality prediction. In order to understand the significance of estimated image features (i.e., blockiness and zero-crossing), I consider the above mentioned methods

Table 4.7: Evaluation results comparison (Scale, 1-5)

Method	*Training*			
	CC	AAE	RMSE	OR
Proposed, NR	0.964	0.276	0.336	0.061
Method, NR[31]	0.966	0.292	0.367	0.069
Method, FR[5]	0.945	0.310	0.381	0.065
2D quality mean, FR (by using [22]	0.917	0.410	0.521	0.089
	Testing			
Proposed, NR	0.94	0.33	0.41	0.040
Method, NR[31]	0.935	0.350	0.421	0.065
Method, FR[5]	0.929	0.370	0.441	0.082
2D quality mean, FR (by using [22])	0.913	0.450	0.521	0.102

which used both features, blockiness and zero-crossing individually with and without disparity. It is clear from Tables 4.4, and 4.5 that the performance of the method considering only zero-crossing is better when compared to the method considering only blockiness both for with and without disparity. Therefore, the zero-crossing feature is more significant compared to blockiness feature for quality prediction. Proposed model's weighting factors also show the deviance. Weighting factors (w_3, and w_4) of zero-crossing are higher compared to weighting factors (w_1, and w_2) of blockiness (see Table 4.3). Although my proposed model's features combined equation use exponential weighting (see Equations 3.18, 3.27, and 3.28) I consider linear weighting (see Equations 4.10, 4.11, and 4.13) and and study the performance in order to understand the justification of exponential weighting. It is clear from Tables 4.4, and 4.6 that the proposed model performance with exponential weighting is far better compared to linear weighting. Therefore, using an exponential function for combining features gives superior performance.

The MOS versus MOSp of my proposed model for training and testing images are respectively shown in Figures 4.24 (a), and 4.24 (b). The symbols $*$ and $\pm$ respectively indicate MOSp points for the databases of training and testing. The MOSp points $*$ and the error bars of ± 2 standard deviation intervals of each different stereo images are shown in Figures 4.25, 4.6, and 4.6. Error bars show the ± 2 standard deviation interval of the MOS. The Figures indicate the predictions consistently performed well in almost similar nature on a variety of image contents.

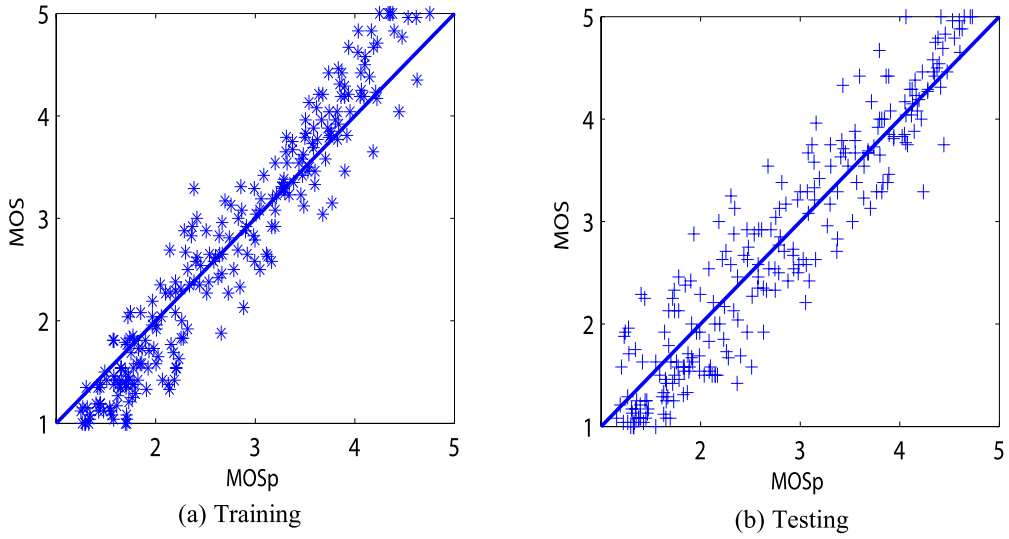


Figure 4.24: MOS versus MOSp of my proposed model.

4.7 Performance Comparison

In this section, I compare the performance of my proposed model against the recently published NR model [31]. The method in [31] used three local features (edge, flat, and texture) and the MICT database. My proposed model's evaluation results

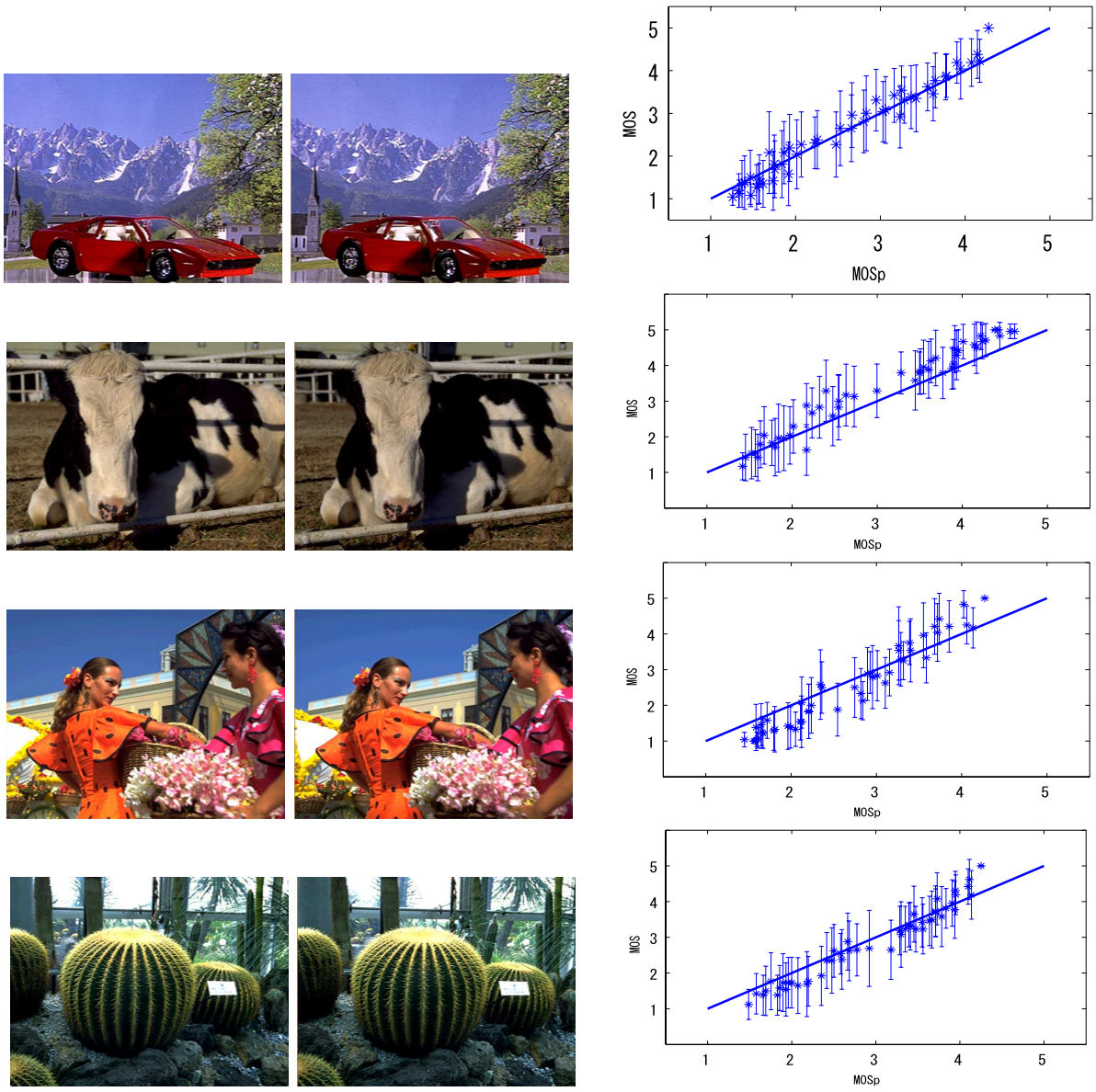


Figure 4.25: The MOSp performances on texture variety of stereo pairs over the quality range. The predictions points $*$ and $\pm$ 2 standard deviation intervals are shown for each stereo pair.

on the same database are shown in Table 4.7. Table 4.7 shows that the performance of my proposed model is superior compared to the published method both for the training and testing databases. As a comparison, I can compare the performance of

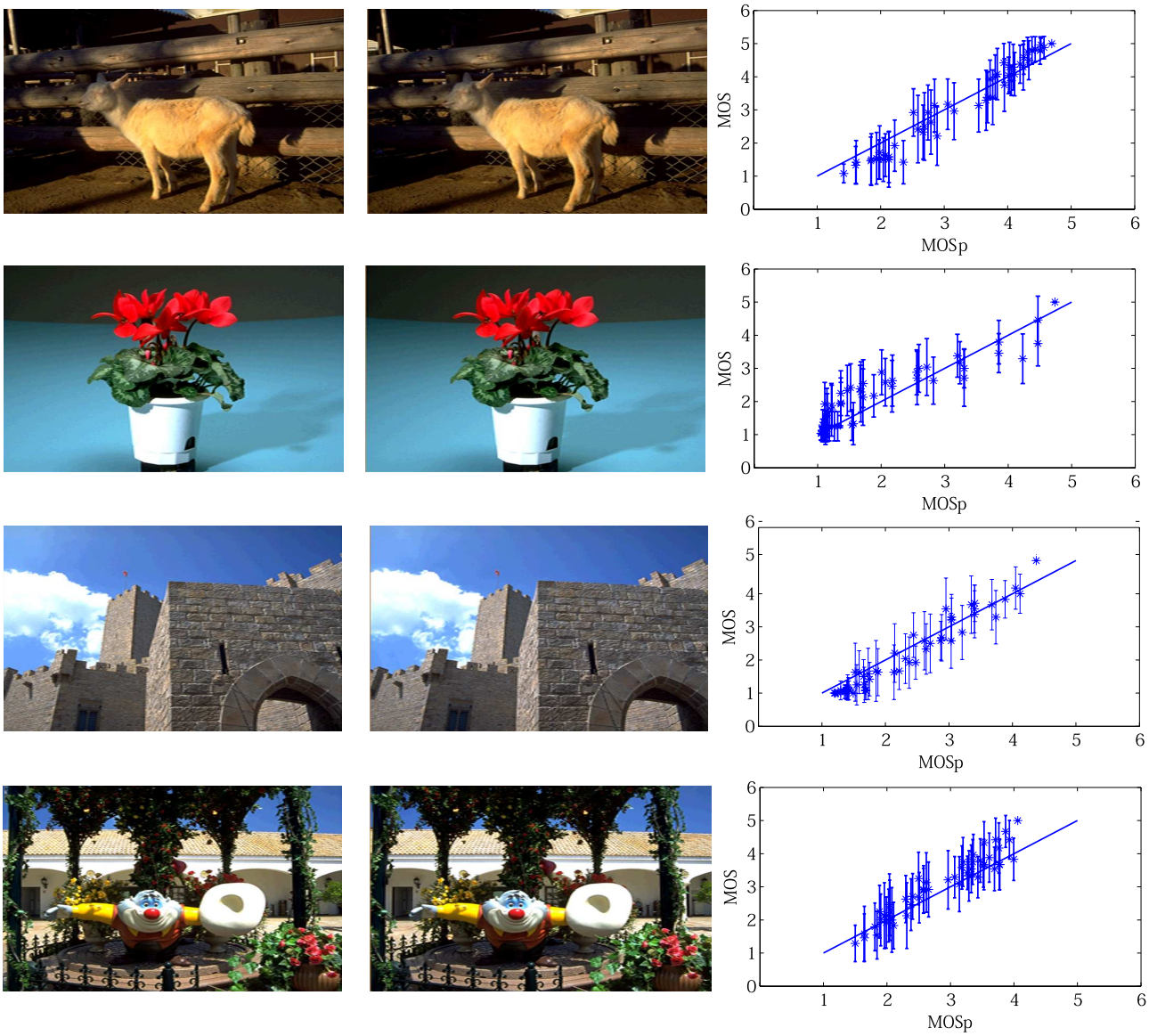


Figure 4.26: The MOSp performances on texture variety of stereo pairs over the quality range. The predictions points $*$ and $\pm$ 2 standard deviation intervals are shown for each stereo pair. Stereo image pairs from the MICT database [25]

my proposed method against the recently published FR method presented in [5]. I evaluate the performance of the method with the same database (MICT database). Table 4.7 shows that the performance of my proposed model is better even compared

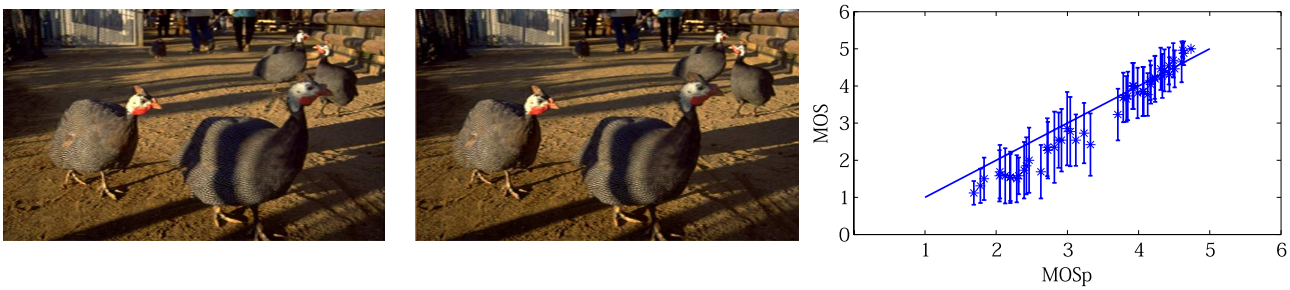
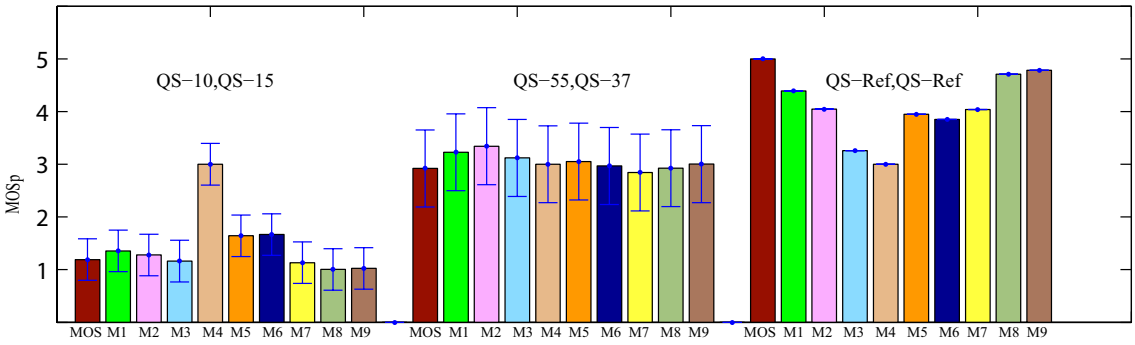


Figure 4.27: The MOSp performances on texture variety of stereo pairs over the quality range. The predictions points $*$ and $\pm$ 2 standard deviation intervals are shown for each stereo pair. Stereo image pairs from the MICT database [25]

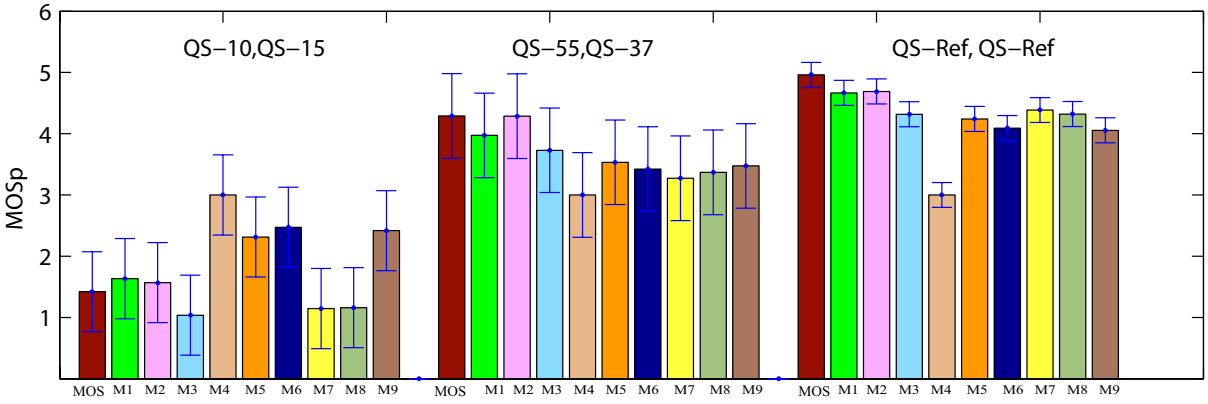
to the FR method [5]. Some researchers claim that 3D stereo quality measure of a stereo pair can be estimated by using a 2D objective quality method (i.e., averaging 2D quality of the left and right views). According to this idea, I compared the performance of my proposed method against the popular FR, objective method for 2D quality assessment method in [22]. I also evaluated the performance of the method on the same database (MICT database). Table 4.7 shows that the performance of my proposed model is better compare to the method in [22]. It is clear from this result that a 2D quality metric is not enough good for 3D quality prediction. Therefore, it is clear from Figures 4.24 and Table 4.4 that my proposed method performance is sufficient.

In order to compare the prediction performance of my proposed method with other methods (see Section 4.6), I considered three different levels of compressions for each of the reference stereo pairs. The levels of compression are high, medium and low, which are compression levels (QS10-QS15), (QS55-QS37) and (Ref-Ref) respectively. Figures 4.28– 4.37 show the MOSp performances of various methods for all reference pairs. The MOS values and the error bar of $\pm$ standard deviation intervals for



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

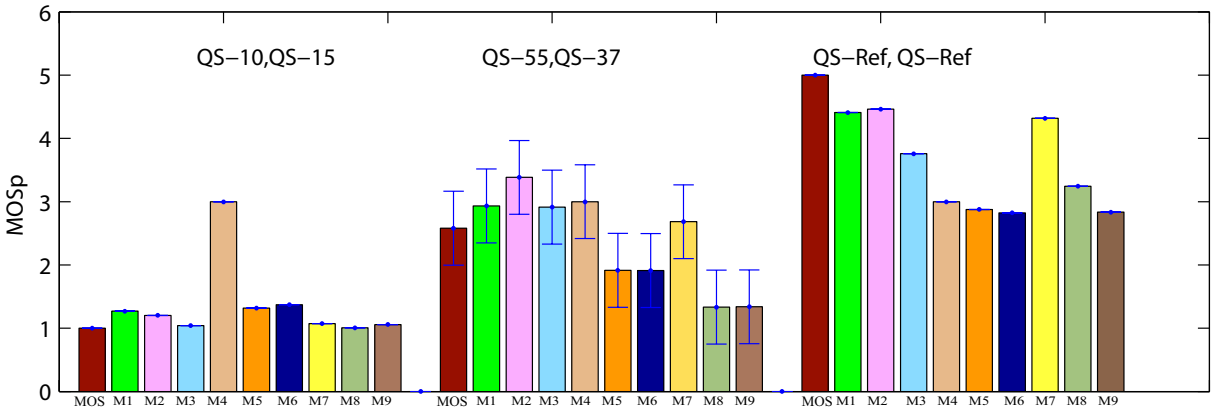
Figure 4.28: Performance of various methods for Car image with ± 2 standard deviation intervals.



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

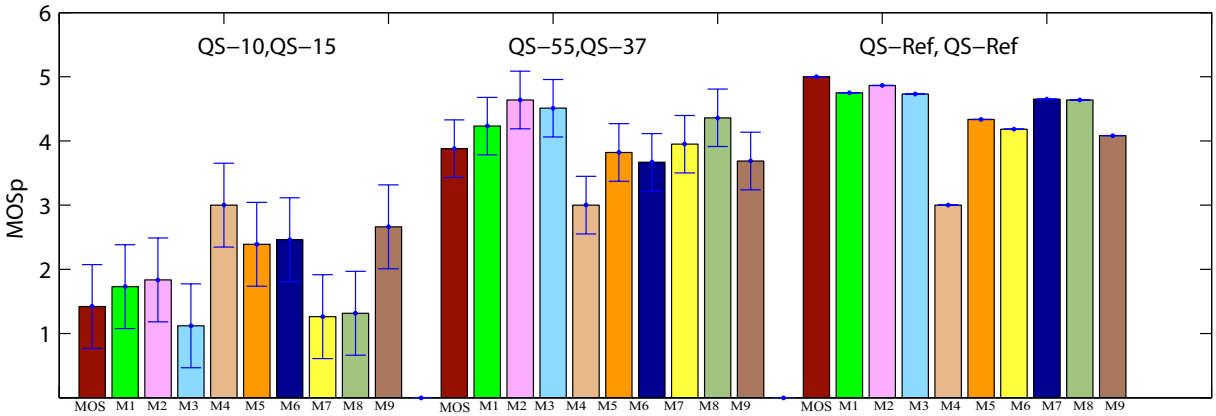
Figure 4.29: Performance of various methods for Cattle image with ± 2 standard deviation intervals.

each compression level are also shown in these figures. The figures indicate that the performance of my proposed method is sufficient and better compared to the other evaluated methods. They also indicate that my method is consistent with not only



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

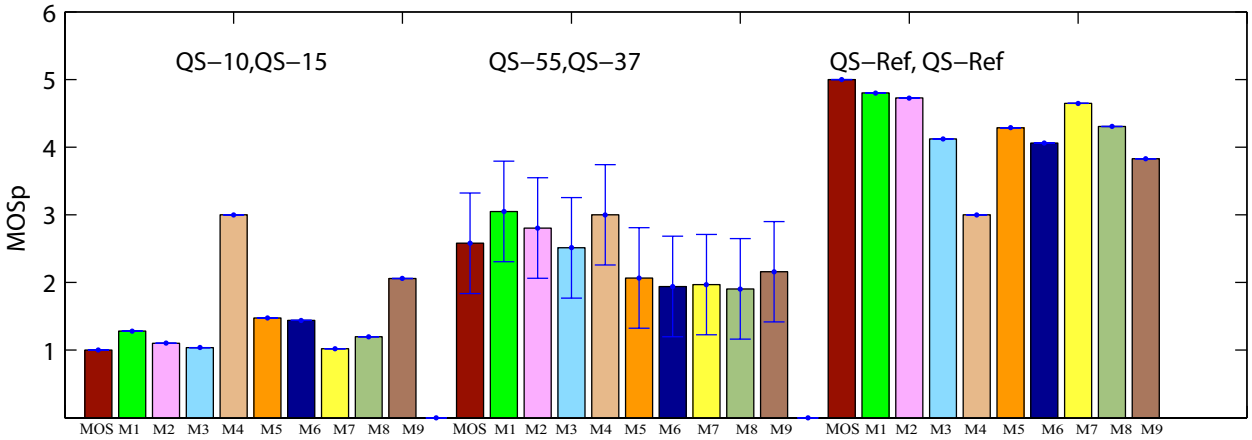
Figure 4.30: Performance of various methods for Monguy image with ± 2 standard deviation intervals.



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

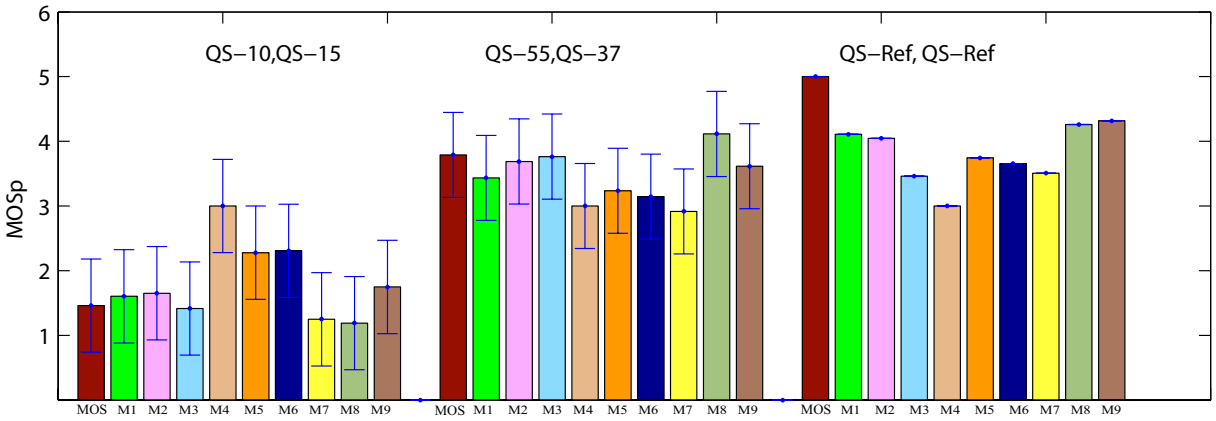
Figure 4.31: Performance of various methods for Goat image with ± 2 standard deviation intervals.

different type of images but also various level of compressions.



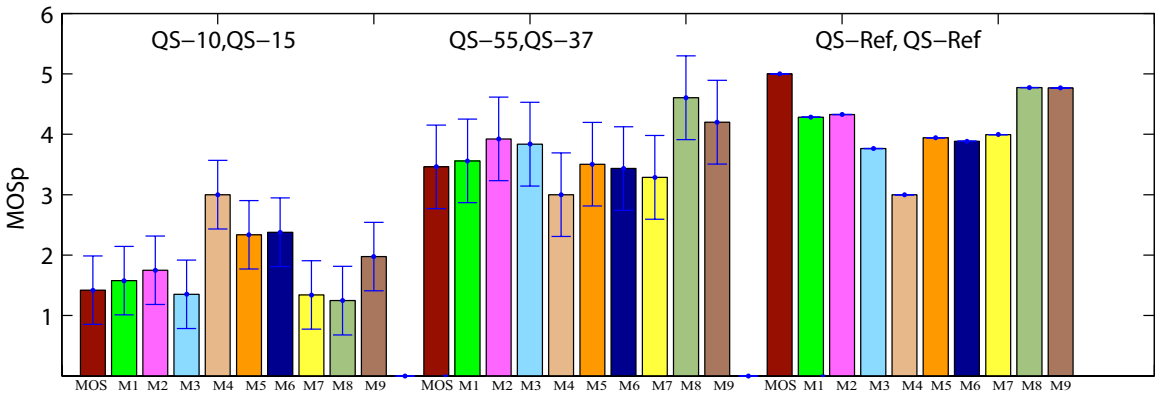
MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.32: Performance of various methods for Flower image with ± 2 standard deviation intervals.



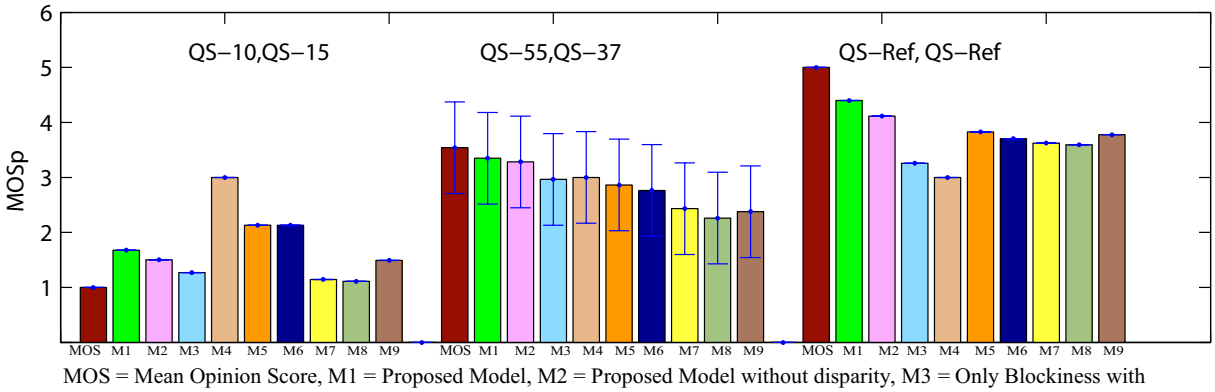
MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.33: Performance of various methods for Ningyo image with ± 2 standard deviation intervals.



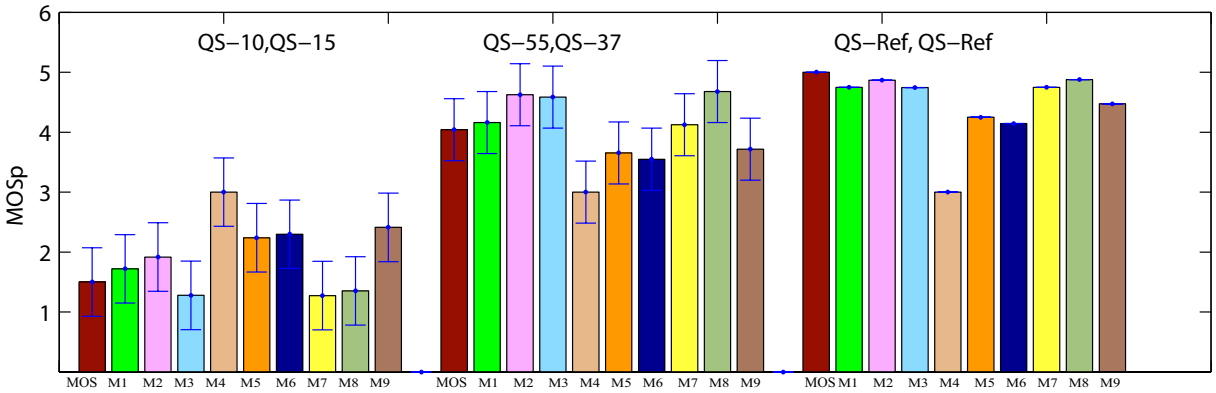
MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.34: Performance of various methods for Saboten image with ± 2 standard deviation intervals.



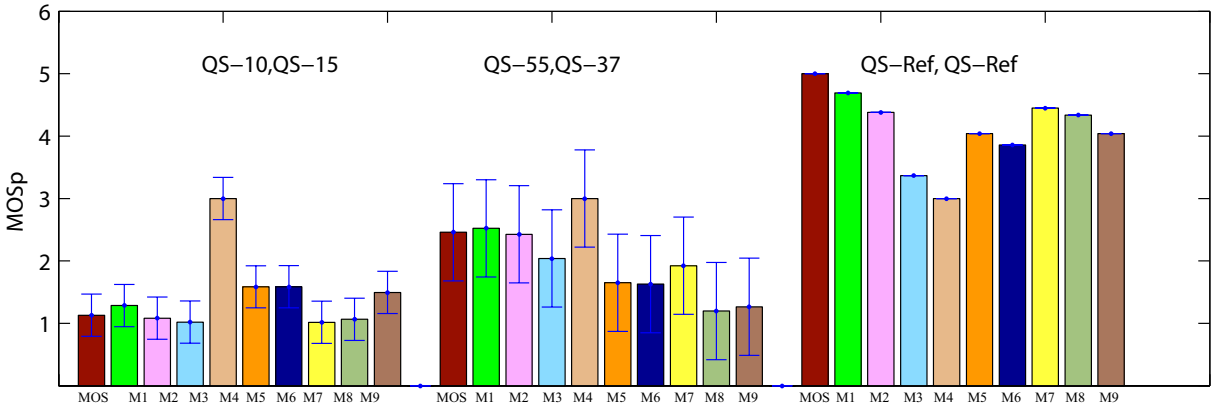
MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.35: Performance of various methods for Women image with ± 2 standard deviation intervals.



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.36: Performance of various methods for Peacock image with ± 2 standard deviation intervals.



MOS = Mean Opinion Score, M1 = Proposed Model, M2 = Proposed Model without disparity, M3 = Only Blockiness with disparity M4 = Only Blockiness without disparity, M5 = Only Zero-crossing with disparity, M6 = Only Zero-crossing without disparity, M7 = 2D Mean, M8 = Linear weighting, M9 = Linear weighting without disparity.

Figure 4.37: Performance of various methods for Flower2 image with ± 2 standard deviation intervals.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This thesis provides a description and analysis of various methods for objective image quality assessment. In my related work, I have reviewed several objective quality assessment metrics and discussed their limitations for not being consistent with human visual perception. I have reviewed the methods that consider the human visual characteristics and discussed their limitations. In order to verify the relationship of image features and HVS, I ran several experiments to analyze the degradation of images for different levels of compression and various contents of images. By considering the limitations of these existing methods, I proposed a new metric for quality assessment based on image features: edge and non-edge areas within the image, which have promising advantages over traditional image quality assessment approaches.

The main contribution of this thesis is the development of a NR objective quality assessment metric, which can be used not only for blind image quality assessment but

also for real time disparity estimation. Two major artifacts in compressed images, blockiness and blur are addressed in this thesis and the test results of my proposed method illustrates the sufficient consistency with human visual perception. The evaluation also shows that my method is very compatible with human assent scores for all different types of images and various level of compressions. In summary, I have made the following important contributions:

- I have developed a fully automated NR image quality assessment metric which can assess the quality of processed images without human intervention. My proposed method exhibits excellent consistency and compatibility with subjective assessment.

- My metric can blindly estimate the quality of images of the distortion during image processing or transmission such as 3DTV or HDTV streaming or broadcasting.

- My metric can also be used to identify two major artifacts, blockiness and blur of compressed images. Therefore, my metric can be embedded into JPEG or JPEG2000 coders.

- My algorithm can also be used in real time disparity estimation.

The scarcity in NR stereoscopic image quality assessment and the necessity of 3D depth estimation influenced me to develop such an algorithm. I strongly believe that my research would be recognized as an important contribution in the field of 3D imaging and it can be used for benchmark image-processing system.

5.2 Future Work

There are many possibilities for the development of innovative metrics for NR quality assessment. Due to the limited knowledge of HVS for predicting the quality of images, the quality assessment metric is still validated by subjective databases. A major improvement can be done by ensuring the good quality subjective assessment. In this thesis, I have proposed a method to estimate image artifacts or distortions by using image features. I have pointed out that the blockiness and blur as a major artifacts because they are the most common artifacts or JPEG or JPEG2000 coder. In future, this approach can be applied for any other coded images irrespective of image artifacts or compression techniques.

The improved approach may also include color information which may lead to better quality prediction accuracy. In the next step, 3D video quality assessment is possible by incorporation of the temporal dependency between adjacent images (frames) of the video. Another direction for the future research is to evaluate the disparity in real time application of the enhanced image. Furthermore great improvements can be achieved by studying my existing method for real time depth estimation in 3D scene recognition.

Appendix A

Supporting Graphs

A.1 PSNR vs MOS Graphs

A.2 Disparity Map

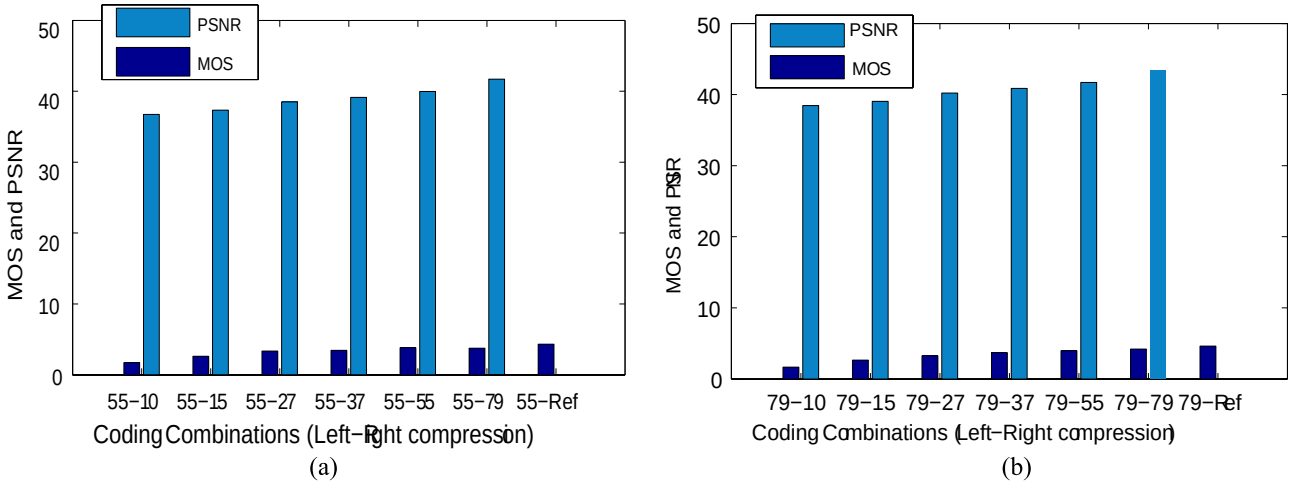


Figure A.1: PSNR versus MOS

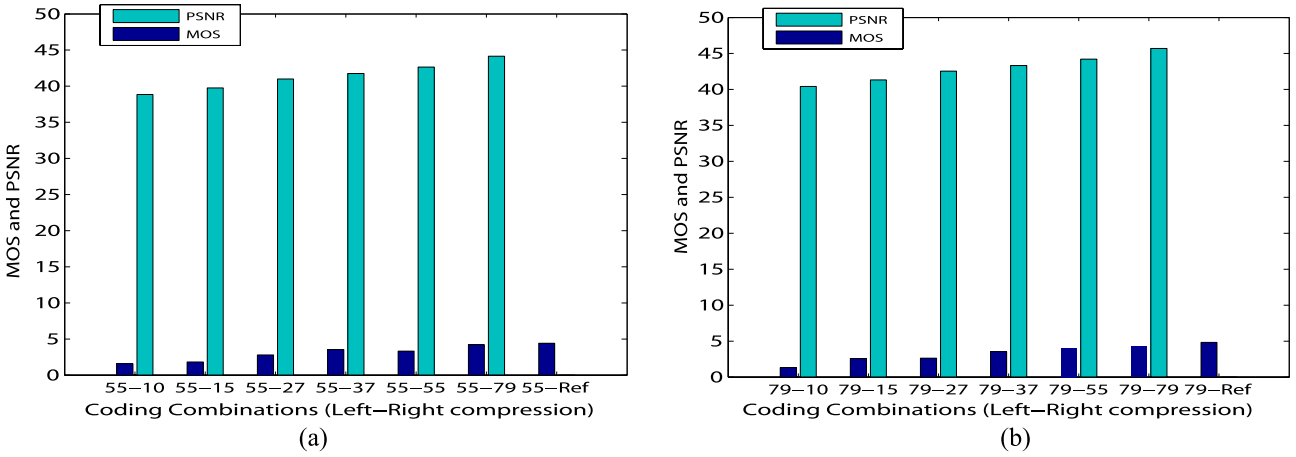


Figure A.2: PSNR versus MOS

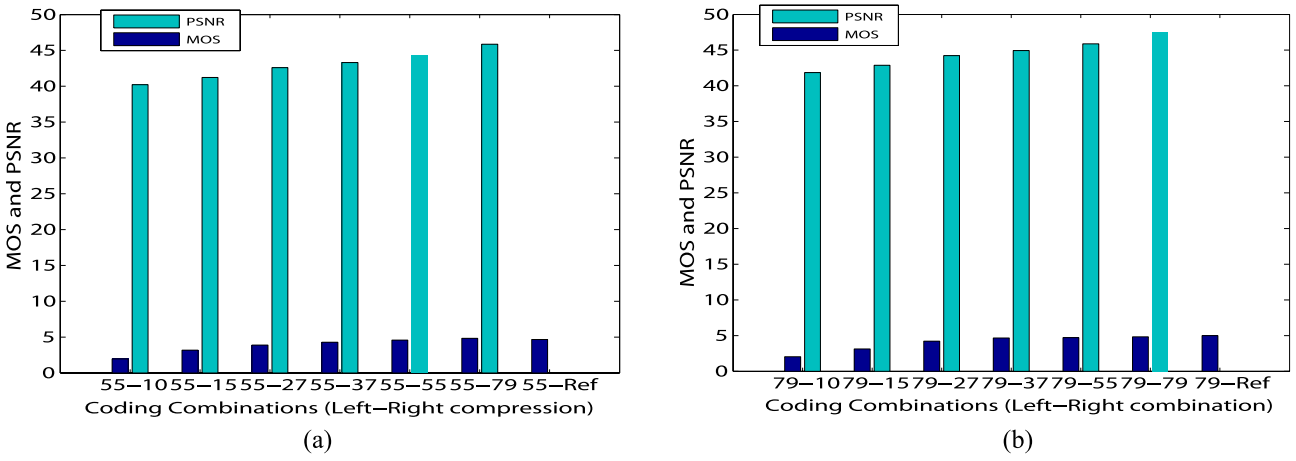


Figure A.3: PSNR versus MOS

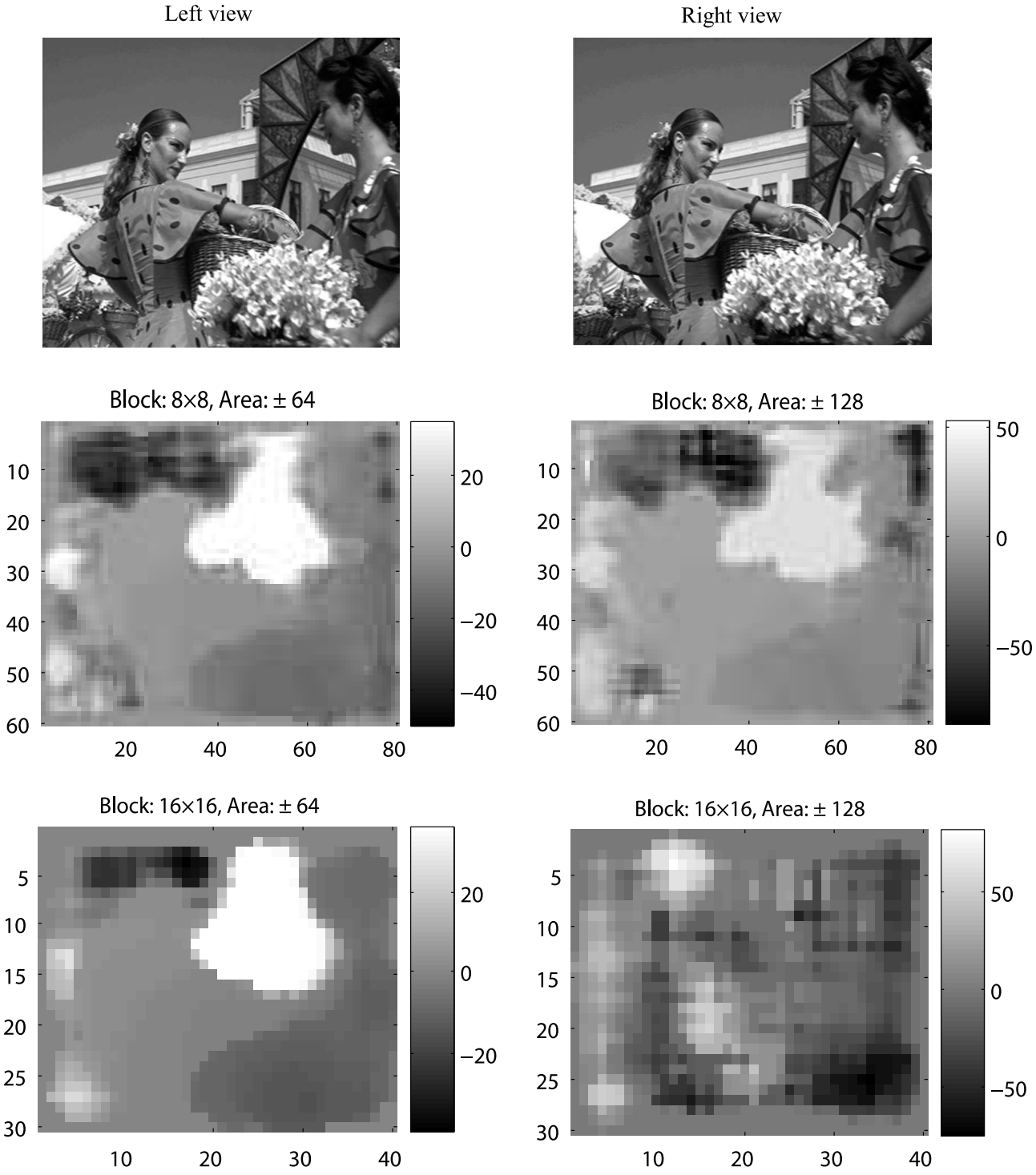


Figure A.4: Disparity map of Cattle image pairs with various block-sizes and search areas. Image pair from the MICT database [25]

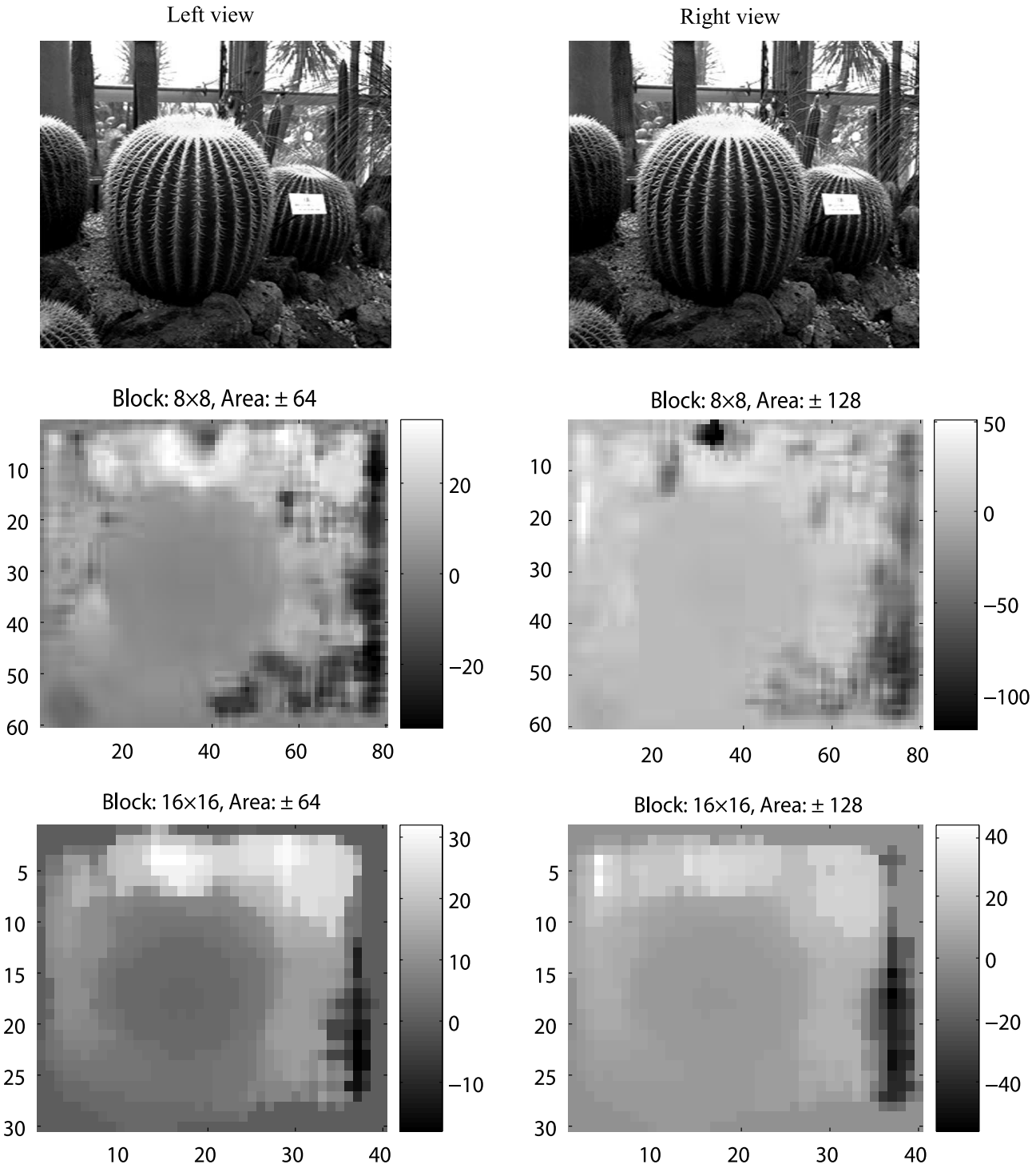


Figure A.5: Disparity map of Saboten image pair with various block-sizes and search areas. Image pair from the MICT database [25]

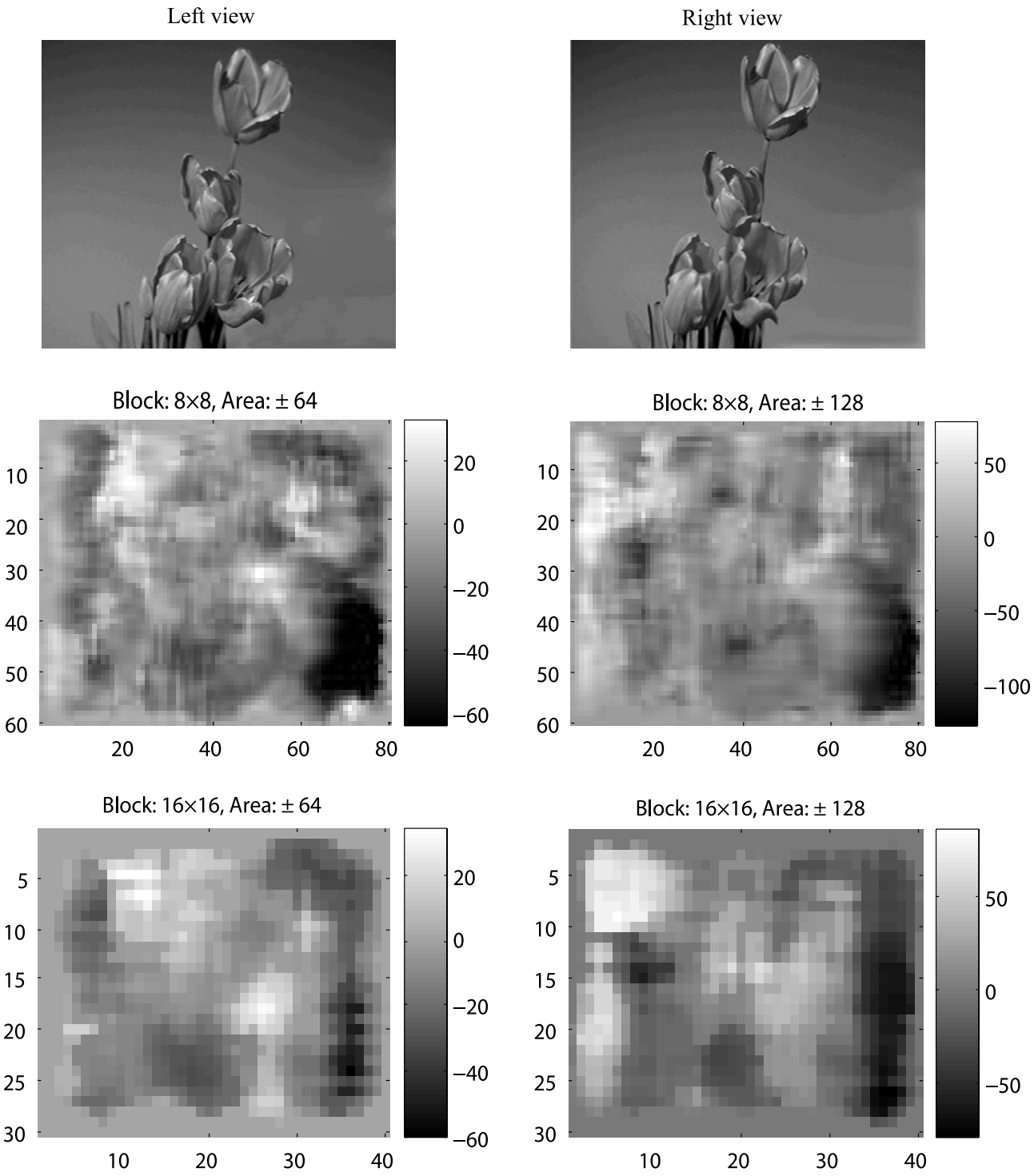


Figure A.6: Disparity map of Flower image pair with various block-sizes and search areas. Image pair from the MICT database [25]

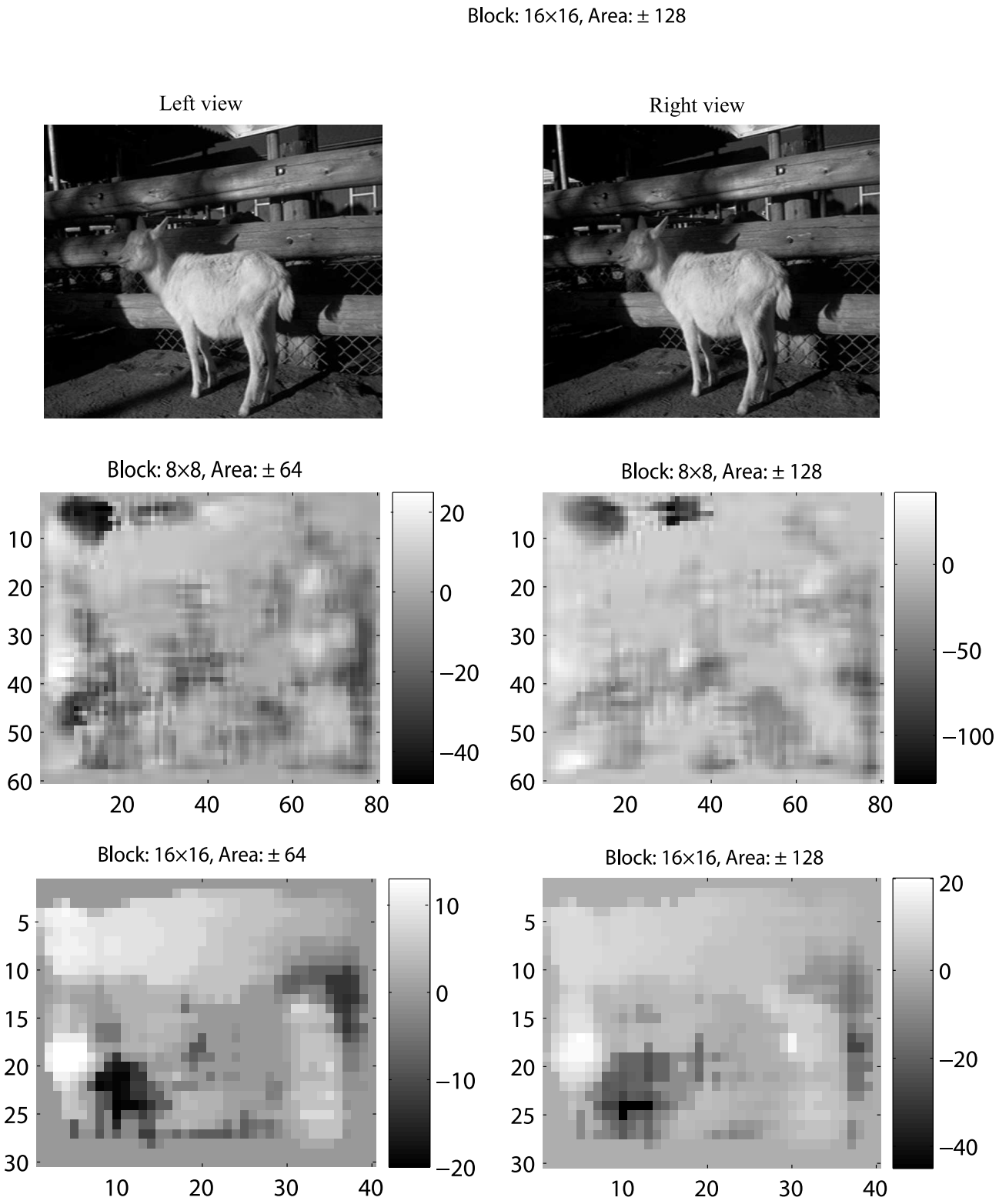


Figure A.7: Disparity map of Goat image pair with various block-sizes and search areas. Image pair from the MICT database [25]

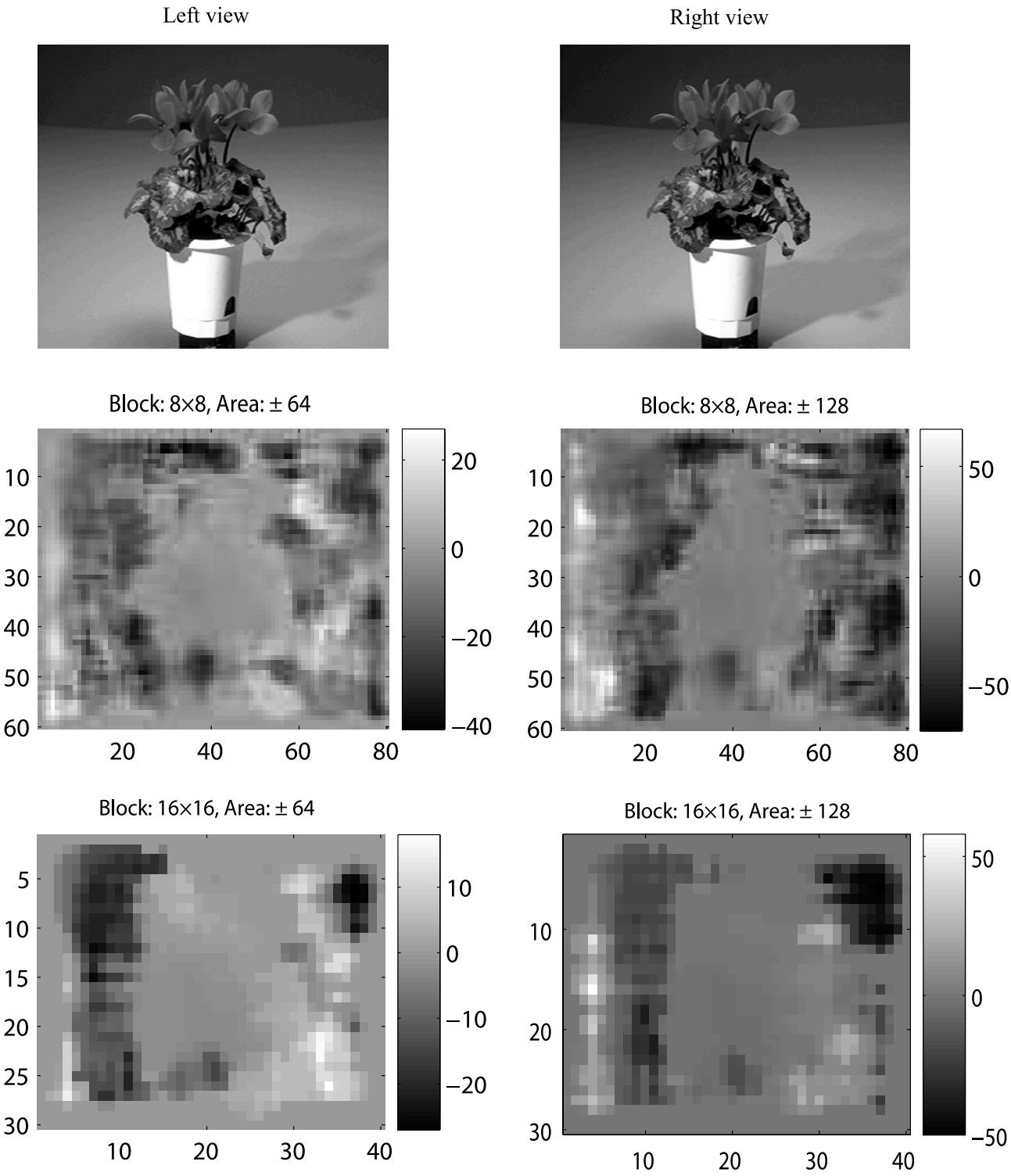


Figure A.8: Disparity map of Flower2 image pair with various block-sizes and search areas. Image from the MICT database [25]

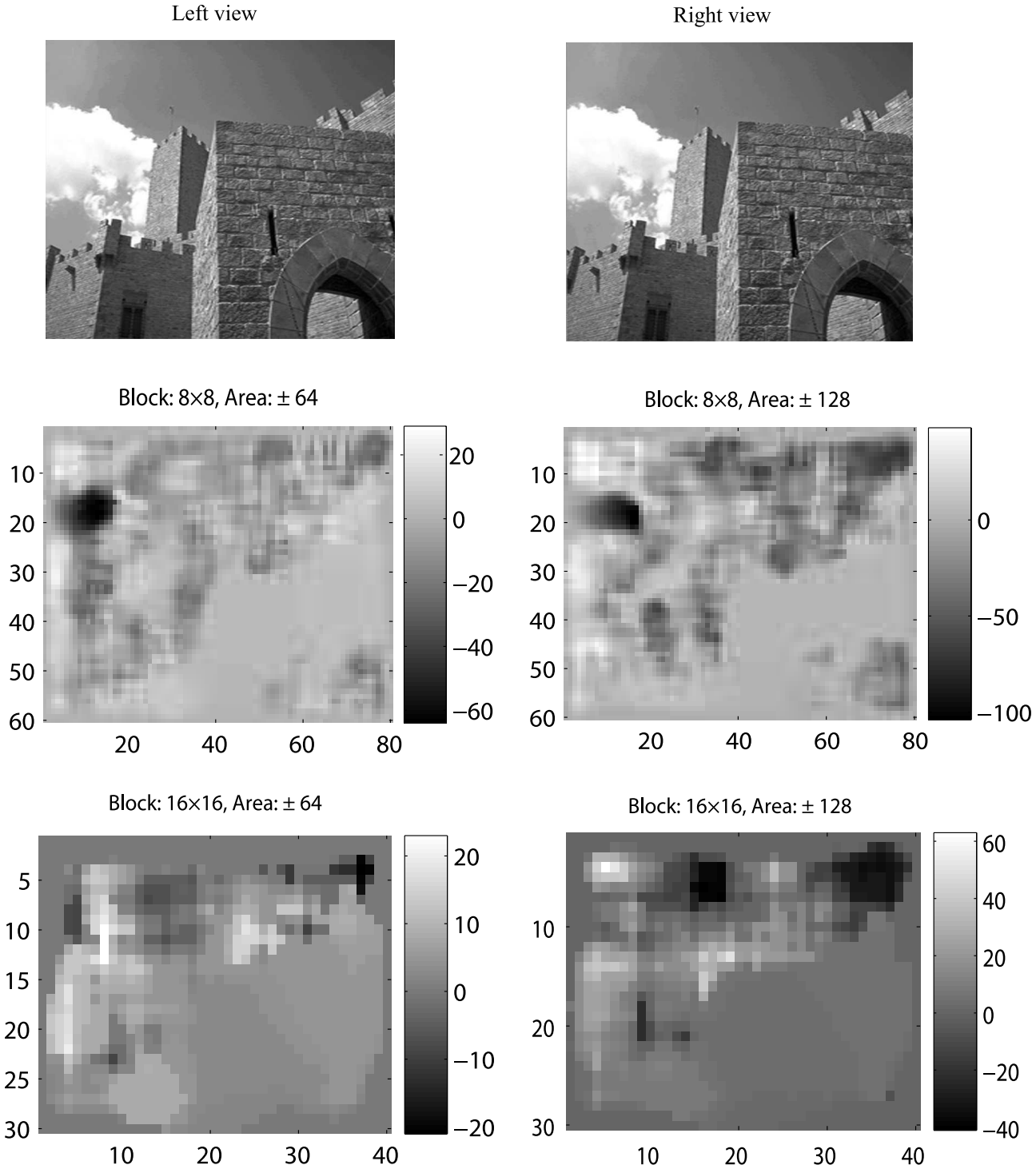


Figure A.9: Disparity map of Mongyu image pair with various block-sizes and search areas. Image from the MICT database [25]

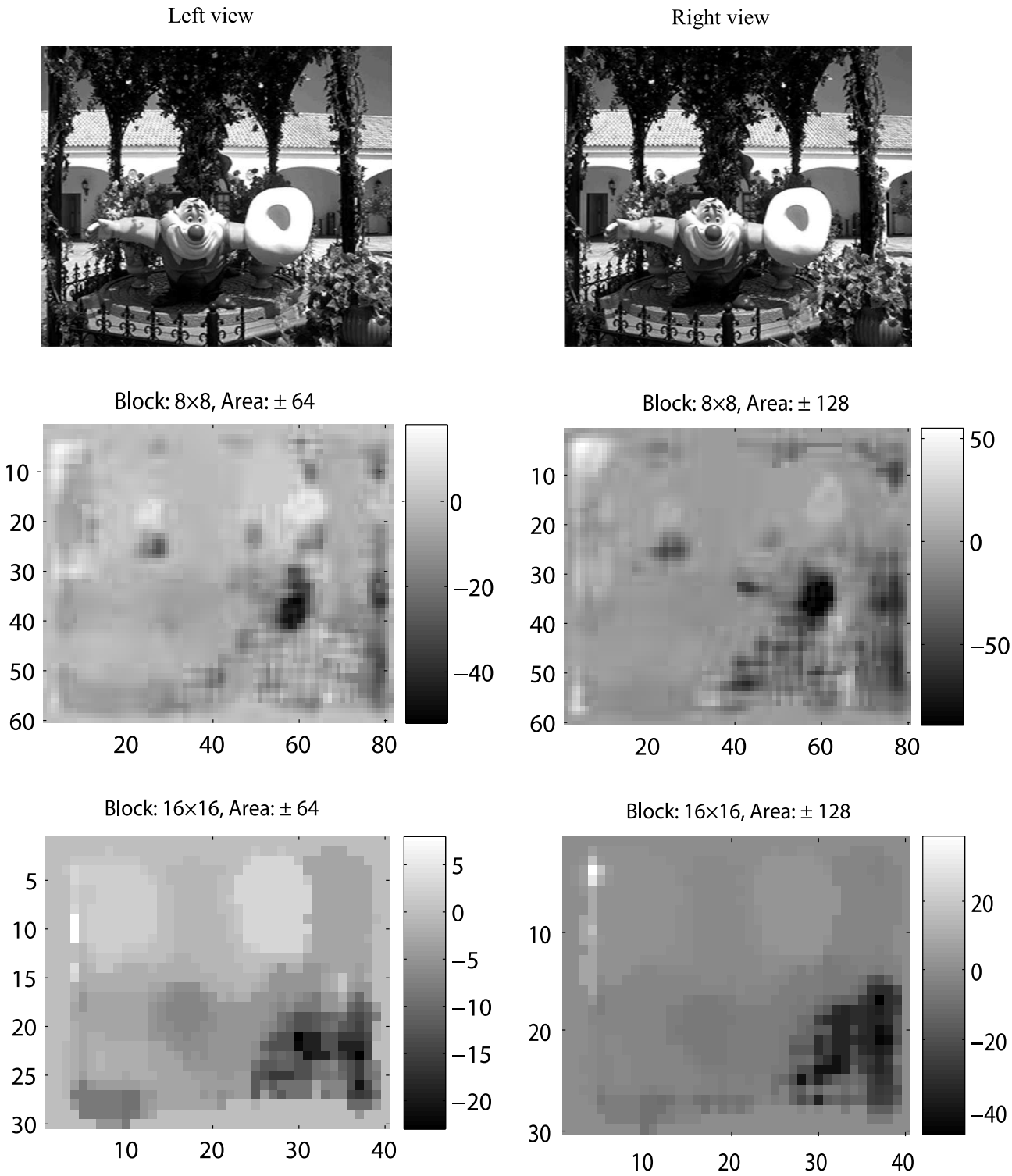


Figure A.10: Disparity map of Ningyo image for various block-sizes and search areas. Image from the MICT database [25]

Appendix B

Subjective Experiment: VQEG Description

In my thesis, I consider subjective assessment data to evaluate the performance of my mathematical (objective) method. The subjective experiment was conducted in the Media Information and Communication Technology (MICT) laboratory [25], the University of Toyama, Japan on 24 bit/pixel RGB color stereoscopic images. A double stimulus impairment scale (DSIS) method was used in the subjective experiment. The test format is shown in Figure B.1. Both distorted (Ai, Bi) and reference (Ar, Br) images were displayed sequentially in the method. At the end of the presentation, the subject was asked to assess the annoyance he/she felt over all perceptual quality on the distorted stereo image with respect to the reference stereo one. The impairment scale contained five categories marked with adjectives and numbers as follows: "Imperceptible =5", "Perceptible but not annoying =4", "Slightly annoying =3", "Annoying =2" and "Very annoying =1". Twenty-four non-expert subjects (12

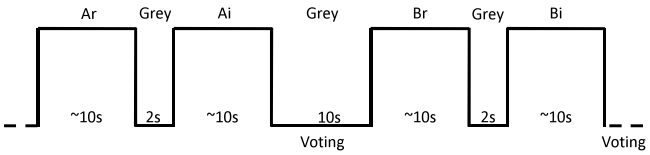


Figure B.1: Double Stimulus Impairment Scale (DSIS) test format [16]

males and 12 females, age range: 19-32 years) were shown the database; most of them were college/university student. A 10-inch auto stereoscopic, LCD (SANYO) display (resolution: 640 × 480) was used in this experiment to display the stereoscopic images and the subjects were instructed about the limited horizontal viewing angle to perceive 3D image correctly. Mean opinion scores (MOSs) were then computed for each stereo image after the screening of post-experiment results according to ITU-R Rec. 500-10 [16]. The subjective test conditions and parameters are summarized in Table B.1.

Table B.1: Subjective test conditions and parameters [25]

Method	DSIS
Evaluation scales	5 Grades (Impairment scales)
Reference stereo pairs	10, (640×480 pixels) 24-bit/pixel, RGB
Coder	JPEG
Coding parameters	QS: 10, 15, 27, 37, 55, and 75
Subjects	24 (Non expert, students)
Display	10-inch, LCD 3D Auto stereoscopic
Display resolution	640×480 pixels (LR: 320 × 480 pixels)
Display Technology	Image-Splitter
Input signals	RGB (Switching 2D/3D)
Viewing distance	4H (H = Picture height)
Room illumination	Dark

Appendix C

Glossary

AAE Average absolute prediction error.

CC Correlation Coefficient

DCT Discrete Cosine Transformation presents data as a sum of cosine functions of various magnitudes and frequencies.

DSIS Double Stimulus Impairment Scale. A method for subjective image quality assessment.

Edge can be defined as image features which identify the variation of images. Mathematically, an ideal edge is a discontinuity of the spatial grey value of the image plane.

HVS model A model used for image processing and computer vision experts to deal with biological and psychological for simplifying the behaviours of a complex human visual characteristic.

ITU International Telecommunication Union.

JPEG encoder A software module for JPEG image compression based on discrete cosine transformation.

Image Segmentation is the process to distinguish image features from other objects and from its background. It identifies each individual pixel to see whether it belongs to an object of interest or not.

Lossy Image Compression A method which reduces a file size by removing information from the original source for lower transmission bandwidth and/or storage capacity.

MOS Mean Opinion Score. The mean opinion score is the human assessment of perceived image quality after compression, which expressed as a single score from 1 to 5. The quality scale (1 to 5) indicates form the lowest to highest quality measurement of an image.

MICT Media Information and Communication Technology laboratory, University of Toyoma, Japan.

MOSp Mean Opinion Score Predicted. MOS score given by a mathematical model without actual human viewers.

MSE Mean Square Error. MSE is the cumulative squared error between the compressed and the original image [37].

OR Outlier Ratio.

PSNR Peak Signal to Noise Ratio. PSNR measures how closely a distorted image resembles a reference image.

PSO Particle Swarm Optimization. An optimization technique inspired by social analogy without detailed knowledge of the problem to be optimized. In PSO algorithm, there is a swarm of particles moving in an n-dimensional problem space, where each particle represents a potential solution and adjusts its position according to experience and neighbours.

QS Quality Scale. A measurement for the quality evaluation of an image features by human observer, rather than the mathematical differences of each pixel.

SROCC Spearman Rank Order Correlation Coefficient.

VQEG Video Quality Experts Group.

Zero-crossing an edge detector that looks for the variation of pixels where the intensity of the image changes.

Bibliography

[1] University of Manitoba. `http://umanitoba.ca/`, Accessed on September 23, 2009.

[2] Roushain Akhter, Z. M. Parvez Sazzad, Yuukou Horita, and Jacky Baltes. No-reference stereoscopic image quality assessment. In *Proceedings of SPIE: Stereoscopic Displays and Applications XXI*, volume 7524, San Jose, CA, USA, January 2010.

[3] Reijo Asikainen. *Quality Analysis of a Printed Natural Reference Image*. PhD thesis, Aalto University, Department of Electrical and Communications Engineering, May 2010.

[4] Haliik Aydinoglu, Faouzi Kossentini, Qin Jiang, and Monson H. Hayes. Region-based stereo image coding. In *IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 57–60, Washington D.C, October 1995.

[5] Alexandre Benoit, Patrick Le Callet, Patrizio Campisi, and Romain Cousseau. Quality assessment of stereoscopic images. *EURASIP Journal on Image and Video Processing*, 2008, October 2008.

[6] Myron Z. Brown, Darius Burschka, and Gregory D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 25(8):993–1008, August 2003.

[7] Neil A. Dodgson. Autostereoscopic 3D displays. *Computer*, 38(8):31–36, 2005.

[8] Russell C. Eberhart and Yuhui Shi. Comparison between genetic algorithms and particle swarm optimization. In *Evolutionary Programming VII*, volume 1447, pages 611–616. Springer, 1998.

[9] Mohammad Eslamia and Farah Torkamani Azar. Block-based disparity estimation by Partial Finite Ridgelet Distortion Search (PFRDS). *Optics and Lasers in Engineering*, 48(1–2):125–131, January 2010.

[10] Pedro F. Felzenszwalb and Daniel P. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision (IJCV)*, 70(1):41–54, October 2006.

[11] Paolo Gastaldo, Rodolfo Zunino, Ingrid Heynderickx, and Elena Vicario. Objective quality assessment of displayed images by using neural networks. *Signal Processing: Image Communication*, 20(7):643–661, August 2005.

[12] Paul Gorley and Nick Holliman. Stereoscopic image quality metrics and compression. In *Proceedings of SPIE: Stereoscopic Displays and Applications XIX*, volume 6803, January 2008.

[13] Chaminda Thushara Hewage, Stewart T. Worrall, Safak Dogan, and Ahmet

Kondoz. Prediction of stereoscopic video quality using objective quality models of 2-D video. *Electronics letters*, 44(16):963–965, 2008.

[14] Yuukou Horita, Yuukou Kawai, Yamaguchi Minami, and Toshiaki Murai. Quality evaluation model of coded stereoscopic color image. In *Proceedings of SPIE: Visual Communications and Image Processing (VCIP)*, volume 4067, pages 389–398, Perth, Australia, June 2000.

[15] Wijnand A. IJsselsteijn, Huib de Ridder, and Joyce Vliegen. Subjective evaluation of stereoscopic images: Effects of camera parameters and display duration. *IEEE Transactions on Circuits and Systems for Video Technology*, 10(2):255–233, March 2000.

[16] International Telecommunication Union (ITU). Methodology for the subjective assessment of the quality of television pictures. `http://www.itu.int/en/pages/default.aspx`, Accessed on April 12, 2010. Recommendation ITU-R BT.500-11.

[17] Anil K Jain. *Fundamentals of Digital Image Processing*. Prentice Hall, NJ, USA, October 1988.

[18] JPEGclub.org. cjpeg, Compress JPEG. `http://jpegclub.org/`.

[19] James Kennedy and Russell Eberhart. Particle swarm optimization. In *IEEE International Conference on Neural Networks(ICNN)*, volume 4, pages 1942–1948, Perth, Australia, 1995.

[20] Ron Kohavi. A study of cross-validation and bootstrap for accuracy estimation

and model selection. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1137–1143, Montreal, Quebec, Canada, August 1995.

[21] Akira Kubota, Aljoscha Smolic, Marcus Magnor, Masayuki Tanimoto, and Tsuhan Che. Multiview imaging and 3DTV. *IEEE Signal Processing Magazine*, 24(6):10–21, 2007.

[22] Ligang Lu, Zhou Wang, Alan C. Bovik, and Jack Kouloheris. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, April 2004.

[23] Pina Marziliano, Frederic Dufaux, Stefan Winkler, and Touradj Ebrahimi. Perceptual blur and ringing metrics: Application to JPEG2000. *Signal Processing: Image Communication*, 19(2):163–172, February 2004.

[24] Brain P. McKinnon. Point, line segment, and region-based stereo matching for mobile robotics. Master's thesis, University of Manitoba, Computer Science, August 2009.

[25] Media Information and Communication Technology (MICT) Laboratory. MICT image quality evaluation database. `http://mict.eng.u-toyama.ac.jp/mict/index2.html`, Accessed on February 24, 2009.

[26] Lydia Maria Johanna Meesters. *Predicted and perceived quality of bit-reduced gray-scale still images*. PhD thesis, Eindhoven University, IPO, Center for Research User-System Interaction, April 2002.

[27] Masatoshi Okutomi and Takeo Kanade. A stereo matching algorithm with an adaptive window: Theory and experiment. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 2, pages 1088–1095, Sacramento, CA, USA, April 1991.

[28] Thrasyvoulos N. Pappas and Robert J. Safranek. Perceptual criteria for image quality evaluation. In *in Handbook of Image and Video Processing*, pages 669–684. Academic Press, 2000.

[29] Kamisetty Ramamohan Rao and Patrick C Y Yip. *Discrete Cosine Transform: Algorithms, Advantages, Applications*. Academic Press Professional, Inc., San Diego, CA, USA, September 1990.

[30] Z. M. Parvez Sazzad. *No-Reference Image Quality Assessments for JPEG and JPEG2000 Coded Images*. PhD thesis, University of Toyama, Department of Systems Science and Engineering, March 2008.

[31] Z. M. Parvez Sazzad, Yuukou Horita, and Yoshikazu Kawayoke. No-reference image quality assessment for JPEG2000 based on spatial features. *Signal Processing: Image Communication*, 23(4):257–268, April 2008.

[32] Z. M. Parvez Sazzad, Shouta Yamanaka, Yoshikazu Kawayoke, and Yuukou Horita. Stereoscopic image quality prediction. In *International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 257–268, San Diego, CA, USA, July 2009.

[33] Pieter Seuntiens, Lydia Meestres, and Wijnand Ijsselsteijn. Perceived quality

of compressed stereoscopic images: Effects of symmetric and asymmetric JPEG coding and camera separation. *ACM Transactions on Applied Perception (TAP)*, 3(2):95–109, January 2006.

[34] Lew B. Stelmach, Wa James Tam, and Daniel V. Meegan. Stereo image quality: effects of spatio-temporal resolution. In *Proceedings of SPIE: Stereoscopic Displays and Virtual Reality Systems VI*, volume 3639, pages 4–11, May 2004.

[35] Video Quality Experts Group (VQEG). Final report from the VQEG on the validation of objective models of video quality assessment, FR-TV Phase II. `http://www.vqeg.org/`, Accessed on April 12, 2009.

[36] AN Volkovich. Construction of three-dimensional models using stereo images and taking into consideration its multisequencing. *Pattern Recognition and Image Analysis*, 18(4):716–719, 2008.

[37] Zhou Wan, Alan C. Bovik, and Ligang Lu. Why is image quality assessment so difficult? In *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Orland, CA, USA, 2002.

[38] Zhou Wang and Alan C. Bovik. *Modern Image Quality Assessment*. Morgan and Claypool Publishers, San Diego, CA, USA, March 2006.

[39] Zhou Wang, Hamid R. Sheikh, and Alan C. Bovik. No-reference perceptual quality assessment of JPEG compressed images. In *IEEE International Conference on Image Processing (ICIP)*, pages 477–480, Rochester, NY, USA, September 2002.

[40] Alan Wee-Chung Liew and Hong Yan. Blocking artifacts suppression in block-coded images using overcomplete wavelet representation. *IEEE Transactions on Circuits and Systems for Video Technology (TCSV)*, 14(4):450–461, April 2004.

[41] Albert M William and Darrell L Bailey. Stereoscopic visualization of scientific and medical content. In *The 33rd international conference on computer graphics and interactive techniques (SIGGRAPH) 2006 Educators program*, Boston, Massachusetts, USA, 2006.

[42] Matusik Wojciech and Pfister Hanspeter. 3d tv: a scalable system for real-time acquisition, transmission, and autostereoscopic display of dynamic scenes. *ACM Transactions on Graphics (TOG)*, 23, August 2004.

[43] Chee Sun Won. Improved block-based image segmentation. In *The International Conference on Image Processing (ICIP'99)*, volume 1, pages 329–332, 1999.

[44] Chee Sun Won. A block-based map segmentation for image compressions. *IEEE Transactions on Circuits and Systems for Video Technology*, 8(5):592–601, 2002.

[45] John Iselin Woodill, Ron Buck, Dave Jurasek, Gaile Gordon, and Terrance Brown. 3d vision: Developing an embedded stereo-vision system. *Computer*, 40:106–108, 2007.