IMAGE DYNAMIC RANGE ENHANCEMENT

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

BY

SERKAN ÖZYÜREK

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

SEPTEMBER 2011

Approval of the thesis:

IMAGE DYNAMIC RANGE ENHANCEMENT

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

Prof. Dr. Canan ÖZGEN   _____
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. İsmet ERKMEN   _____
Head of Department, **Electrical and Electronics Engineering**

Prof. Dr. Gözde B. AKAR   _____
Supervisor, **Electrical and Electronics Engineering Dept., METU**

Examining Committee Members:

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

Prof. Dr. Gözde B. AKAR   _____
Electrical and Electronics Engineering Dept., METU

Assoc. Prof. Dr. Çağatay CANDAN   _____
Electrical and Electronics Engineering Dept., METU

Assist. Prof. Dr. Ahmet Oğuz AKYÜZ   _____
Computer Engineering Dept., METU

Emre TURGAY, MSc.   _____
MGEO, ASELSAN

                                              Date: 16.09.2011

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

Name, Last Name   : Serkan Özyürek

Signature   :

ABSTRACT

IMAGE DYNAMIC RANGE ENHANCEMENT

Özyürek, Serkan

M.Sc., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. Gözde B. Akar

September 2011, 131 pages

In this thesis, image dynamic range enhancement methods are studied in order to solve the problem of representing high dynamic range scenes with low dynamic range images. For this purpose, two main image dynamic range enhancement methods, which are high dynamic range imaging and exposure fusion, are studied. More detailed analysis of exposure fusion algorithms are carried out because the whole enhancement process in the exposure fusion is performed in low dynamic range, and they do not need any prior information about input images. In order to evaluate the performances of exposure fusion algorithms, both objective and subjective quality metrics are used. Moreover, the correlation between the objective quality metrics and subjective ratings is studied in the experiments.

Keywords: Image Fusion, Exposure Fusion, Image Quality Metrics, High Dynamic Range Imaging

ÖZ

GÖRÜNTÜ DİNAMİK ARALIĞININ İYİLEŞTİRİLMESİ

Özyürek, Serkan

Yüksek Lisans, Elektrik-Elektronik Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Gözde B. Akar

Eylül 2011, 131 sayfa

Bu tezde, yüksek dinamik aralığa sahip sahnelerin düşük dinamik aralıkta görüntülenmesine çözüm olan görüntü dinamik aralık iyileştirme algoritmaları incelenmiştir. Bu amaç doğrultusunda, iki temel görüntü dinamik aralık iyileştirme metodu olan yüksek dinamik aralıkta görüntüleme ve pozlama füzyonu incelenmiştir. Pozlama füzyon algoritmaları daha detaylı analiz edilmiştir çünkü pozlama füzyonunda bütün iyileştirme işlemleri düşük dinamik aralıkta gerçekleştirilmektedir ve bu algoritmalar giriş resimleri ile ilgili hiçbir bilgiye ihtiyaç duymamaktadır. Pozlama füzyon algoritmalarının performanslarını değerlendirmek için nesnel ve öznel kalite ölçütleri kullanılmıştır. Ayrıca, deneylerde nesnel kalite ölçütleri ve öznel değerlendirmeler arasındaki korelasyon incelenmiştir.

Anahtar Kelimeler: Görüntü Füzyonu, Pozlama Füzyonu, Görüntü Kalite Ölçütleri, Yüksek Dinamik Aralıkta Görüntüleme

To My Parents ...

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

1D	One Dimensional
2D	Two Dimensional
AC	Alternating Current
CE	Cross Entropy
CRF	Camera Response Function
DC	Direct Current
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
E	Entropy
EF	Exposure Fusion
HDR	High Dynamic Range
HDRI	High Dynamic Range Imaging
IDCT	Inverse Discrete Cosine Transform
JPEG	Joint Photographic Expert Group
LL	Low Low Sub-bands of Discrete Wavelet Transform
LH	Low High Sub-bands of Discrete Wavelet Transform
HL	High Low Sub-bands of Discrete Wavelet Transform
HH	High High Sub-bands of Discrete Wavelet Transform
MD	Multiscale Decomposition
MI	Mutual Information
PDF	Probability Density Function
PT	Pyramid Transform
SD	Standard Deviation
SF	Spatial Frequency
SVD	Singular Value Decomposition Techniques
UIQI	Universal Image Quality Index

CHAPTER 1

INTRODUCTION

The main goal of digital cameras is to represent an acquired scene as realistic, and as similar as the scene observed by the human visual system. In parallel with the increase in digital camera technology, this goal is achieved to a great extent. The luminance, contrast and all the other parameters related with digital images are gradually getting closer to the parameters experienced by a human observer [29].

However, while the digital cameras have been improved in terms of performance and quality, photography is still having a problem with a wide range of radiance variations in the real world. The scenes in the real world comprise of harsh lightening conditions that cause shadows (underexposed regions) or highlights (overexposed regions) in digitally captured images. The reason is that the dynamic range of camera sensors is not high enough to capture the dynamic range of the scenes. Dynamic range of the scene is defined as the ratio of radiances between the brightest and the darkest points in the scene. For example, the radiance range of the real world reaches up to 1:500000 but today's camera sensors have pixel depth ranging from 8-bit to 14-bit which correspond 256 to 16384 digital values [1].

10^{-6} 10^{+6}

Real World

High Dynamic Range

10^{-6} 10^{+6}

Digital Picture
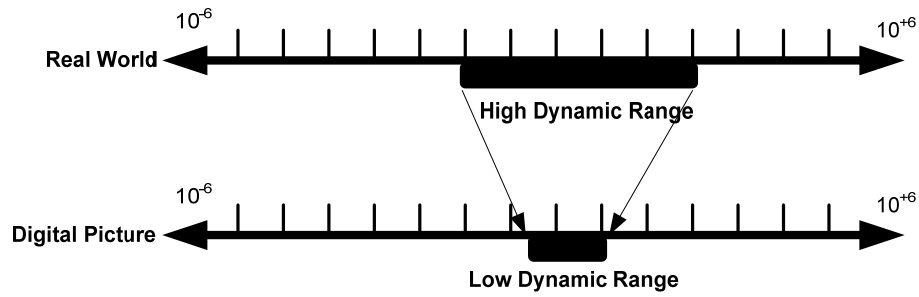
Low Dynamic Range

Figure 1.1 Mapping from High Dynamic Range to Low Dynamic Range

There are two different approaches to solve this problem. These are hardware and software approaches. The hardware approach tries to solve this problem in the image acquisition pipeline of the digital cameras. For this purpose, camera sensors, which have the capability of capturing wide dynamic range, have been developed [2]. In the software approach, the problem is solved by post processing techniques. Since the hardware solutions are expensive and need further developments in this field, they are not preferred by manufacturers or researchers. The software solutions are easier, less expensive than the hardware solutions and usually independent from hardware platforms. Therefore, the software solutions are more preferred than the hardware ones.

However, it is not possible to recover the details in saturated regions by using advanced software techniques directly. The well-known solution to capture a high dynamic range scene with a limited dynamic range device is splitting the full dynamic range into smaller strips. By means of this, the full dynamic range is represented via low dynamic range images. In this form, this sequence of images is not so useful because the details are distributed in several images. The interpretation and processing of these details are problematic. Therefore, these images should be merged into a single image that contains all the details. This overall process is called as image dynamic range enhancement, and methods used for this purpose are called as image dynamic range enhancement methods.

Image dynamic range enhancement methods are becoming popular day by day. The application areas have been expanded from amateur photography to scientific researches. At the same time, some image dynamic range enhancement programs are developed such as Enfuse, Photomatix, Adobe Photoshop add-ons and Bracket [3-5, 12].

1.1 Scope of the Thesis

In this thesis, image dynamic range enhancement methods are studied in order to solve the problem of representing high dynamic range scenes with low dynamic range images. For this purpose, two main image dynamic range enhancement methods, which are high dynamic range imaging and exposure fusion, are studied.

The main goals of this thesis are given as follows:

- Giving principles of high dynamic range imaging methods by implementing basic algorithms.
- Performing detailed analysis of exposure fusion methods.
- Comparison of objective quality metrics used in exposure fusion.
- Giving the correlation between the objective quality metrics and human subjective ratings.

Image sets, which are used in the performance comparison of the algorithms, are mostly obtained through the internet. All images have been already aligned, and registration algorithms are not included to the scope of this thesis. Moreover, this thesis only considers monochrome image based algorithms. If an algorithm is color based, it has been modified to work with the monochrome images.

The implementation of the algorithms and the supplementary programs used in this thesis are developed in MATLAB.

1.2   Outline of the Thesis

In Chapter 2, general background information about high dynamic range imaging is given. Moreover, the implemented recovering high dynamic range radiance map and tone mapping algorithms are explained.

In Chapter 3, the definition of exposure fusion is given, and the literature related with exposure fusion is explained. Then, the implemented exposure fusion algorithms are explained in details.

In Chapter 4, various objective quality metrics that are used to evaluate the results of image fusion algorithms are explained. This chapter forms a base line for comparing image fusion algorithms.

In Chapter 5, the evaluation of image dynamic range enhancement algorithms is given with the results of the objective and subjective tests. This chapter mainly focuses on the evaluation of the exposure fusion algorithms, but a brief comparison of the high dynamic range imaging algorithms is given.

In Chapter 6, conclusion is given, and possible future studies are discussed.

CHAPTER 2

HIGH DYNAMIC RANGE IMAGING

2.1   Introduction

High dynamic range imaging (HDRI) is one of the solutions to represent high dynamic range scenes with low dynamic range images. HDRI is a set of techniques that provide a greater dynamic range than traditional imaging techniques [6]. The output of HDRI methods is called as high dynamic range (HDR) images.   These images can represent full dynamic range of the scenes because they are directly proportional to the radiance values of captured scenes. Therefore, they are also called as HDR radiance maps. In this aspect, HDR radiance maps are different from the standard images because standard images consist of digital counts, which are the projection of the scene radiance on two-dimensional coordinate system.

HDRI techniques consist of two different methods. These are recovering HDR radiance map and tone mapping methods.

For recovering HDR radiance map, different methods have been proposed. The goal of these methods is to turn back to the true radiance values from the digital pixel values. By means of these methods, the information at saturated pixels is recovered, and the real luminance information is obtained. HDR radiance maps are used in various applications such as image based lightening and creating virtual images [6].

However, most of the displays have low dynamic ranges that are not capable of displaying HDR radiance maps. For this reason, a method is required to map HDR radiance maps into low dynamic range images. Tone mappings are the methods that provide this dynamic range compression. Most of these tone mapping methods try to simulate human visual system in order to preserve the important details in HDR radiance maps.

In the following sections, the detailed information about these two methods is given, and the implemented algorithms are explained.

2.2 Recovering High Dynamic Range Radiance Map

The HDR radiance map methods intend to find a radiance map which is a kind of lookup table. Then, they use the found radiance map to calculate scene radiance. However, turning back to the true radiance values by the help of radiance map is not an easy process. The reason of this difficulty is the nonlinearities which exist in the digital cameras. These nonlinearities are modeled by a transfer function (T), and it is named as camera response function (CRF) or radiometric response curve [8]. Several methods have been developed for the estimation of CRF in years [8-11].
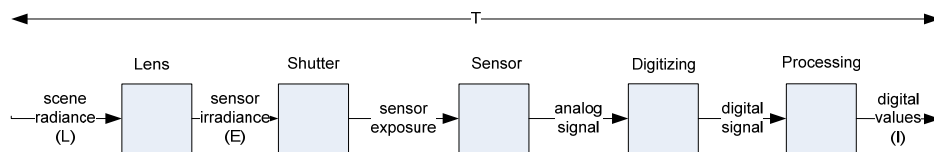
Figure 2.1 Transfer Function of a Typical Camera

These methods use a series of different exposed low dynamic range images in order to estimate CRF. CRF is estimated by assuming that each differently exposed image corresponds to the same scene radiance. In these methods, the

whole process for the estimation of CRF is awarded the name of radiometric calibration. After CRF is estimated, the sensor irradiance is calculated by using CRF and the exposure times of the input images. This operation can be explained mathematically as follows:

$$L(x,y) \approx E(x,y) * \Delta t_i = T^{-1}(I_i(x,y)) \ where \ i = 1, 2, ..., N \qquad \text{2-1}$$

$$I_i(x,y) = T(E(x,y) * \Delta t_i) \ where \ i = 1, 2, ..., N \qquad \text{2-2}$$

where $I_i(x,y)$ is the intensity value of the pixel at *(x,y)*, and Δt_i is the exposure time of ith image. The sensor irradiance is denoted by $E(x,y)$.

The calculated image represents an approximation of the sensor irradiance, and it is called as HDR image or HDR radiance map. The sensor irradiance has been calculated, but the scene radiance has to be found. Most of the time, it is assumed that scene radiance $L(x,y)$ is proportional to sensor irradiance $E(x,y)$. This assumption is valid because manufactured lenses include designs to compensate the nonlinearity between the scene radiance and sensor irradiance [10]. By the help of these designs, the relation between the radiance and the irradiance becomes almost constant.

In the next section, several algorithms for recovering HDR radiance map are explained. Since the first algorithm, HDR-ALG1, defines the most popular CRF estimation method in the literature, CRF estimation section of the other algorithm is not implemented. The steps to calculate HDR radiance map with a known CRF are implemented. In the literature, there are also others algorithms for recovering HDR radiance map [8, 9, 22].

2.2.1   HDR-ALG1 [10]

This algorithm is composed of two sections: 1) the estimation of CRF 2) the calculation of HDR radiance map with a known CRF [10].

For the estimation of CRF, equation 2-2 is modified by taking the natural logarithm of both sides in order for simplification. The modified equation is given as follows:

$$\ln\left(T^{-1}\big(I_i(x,y)\big)\right) = \ln\big(E(x,y)\big) + \ln(\Delta t_i)$$
$$g(I_i(x,y)) = \ln\big(E(x,y)\big) + \ln(\Delta t_i)$$

2-3

In this set of equations, I_i and Δt_i are knowns. Remaining terms are unknowns. In order to solve this overdetermined equation system, least square estimation method is used in the algorithm. Let Z_{min} is 0, and $Z_{max,}$ is 255 for the 8-bit pixel depth. P is the number of sample pixels, N is the number of the input images, and $g^{"}(z)$ is the second derivative of $g(z)$ where z is the pixel intensity. Then, the solution of equation 2-3 is given as follows:

$$O = \sum_{x,y \in S}^{P} \sum_{i=1}^{N} \big[g\big(I_i(x,y)\big) - \ln\big(E(x,y)\big) - \ln(\Delta t_i)\big]^2 + \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} g^{"}(z)^2$$
$$g^{"}(z) = g(z-1) - 2g(z) + g(z+1)$$

2-4

where S is the set of pixel samples. The elements of the set S are randomly chosen pixel intensities from the input images.

The above equation is composed of two terms. These are fitting and smoothness terms. There is also a parameter λ in the above equation. This parameter controls the weight of the smoothness term. After this step, the algorithm is finalized by adding a triangular shape weighting function to the above equation. This

8

weighting function provides much better approximation to the real shape of CRF. It guarantees that CRF has steep slopes near the extreme points and smooth transition at the middle points. Therefore, it gives more importance to the pixels through the middle of the curve than the pixels in the extreme points. Then, the solution is finalized as follows:

$$O = \sum_{x,y \in S}^{P} \sum_{i=1}^{N} \{w(I_i(x,y))[g(I_i(x,y)) - \ln(E(x,y)) - \ln(\Delta t_i)]\}^2$$
$$+ \lambda \sum_{z=Z_{min}+1}^{Z_{max}-1} [w(z)g''(z)]^2$$

2-5

The algorithm does not use all the pixels in the input images in equation 2-5. "Given measurements of P pixels in N images, we have to solve for P values of $\ln(E(x,y))$ and $(Z_{max} - Z_{min})$ samples of g" [10]. To guarantee a sufficiently overdetermined system, it is required to satisfy the following equation:

$$P(N-1) > (Z_{max} - Z_{min})$$

2-6

After the estimation of CRF, sensor irradiance is calculated. Any of the input images with its exposure time can be used to calculate HDR radiance map by equation 2-3. However, all the input images can be used in the calculation in order to reduce the calculation errors and noise in the output image. At this point, it is also suggested to use the same weighting function in equation 2-5. Since CRF is best approximated at the middle pixel values, giving higher weights to these pixels have an effect of decreasing error in the output image. The formula to calculate HDR radiance map is given as follows:

$$\ln L(x,y) \approx \ln E(x,y) = \frac{\sum_{i=1}^{N} w(I_i(x,y))[g(I_i(x,y)) - \ln(\Delta t_i)]}{\sum_{i=1}^{N} w(I_i(x,y))}$$

2-7

where *L(x,y)* is the scene radiance or HDR radiance map. As stated previously, the scene radiance and the scene irradiance can be assumed proportional.

2.2.2   HDR-ALG2 [11]

This algorithm stress that HDR-ALG1 [10] does not give importance to long exposed images and use all the input images equally [11]. Therefore, it suggests that giving higher weights to long exposed images reduces quantization error arising from the digitizing of the sensor irradiance. In addition to HDR-ALG1 [10], the noise in observation model is defined in this algorithm. By using this model, the relation between the sensor irradiance and the digital pixel values is expressed as follows:

$$T^{-1}\big(I_i(x,y)\big) = E(x,y) * \Delta t_i + N_i^C(x,y) + N_i^q(x,y) = C_i(x,y) \qquad \textbf{2-8}$$

where $N_i^C(x,y)$ is the additive noise in i^{th} exposed image, and $N_i^q(x,y)$ is the dequantization error in i^{th} exposed image.

To solve the above equation, the noise terms should be modeled. In the algorithm, they are modeled as zero mean Gaussian random variables. In order to model them as Gaussian variable, the variances of these random variables have to be calculated. This is not an easy task because there are many sources causing noise to the imaging system. Each of these sources should be characterized to insert their effects to the observation model. For this purpose, it is suggested to replace the variances with a weighting function which is inverse proportional to the variances of the noise terms. This weighting function which is defined below has high values where the noise power is less and has low values where the noise power is high.

$$w_i(z) = \exp\left(-4 * \frac{(z - 127.5)^2}{(127.5)^2}\right) \qquad \textbf{2-9}$$

This function gives more importance to the pixels which are at the middle of CRF and does not trust the pixels which are at the two ends of CRF.

Maximum likelihood technique is used to calculate HDR radiance map by the help of the joint probability density function of $I_i(x, y)$. Maximization process of this probability density function is equivalent to the minimization of the following equation:

$$O(E) = \sum_{x=1, y=1}^{X,Y} \sum_{i=1}^{N} w_i (C_i(x, y) - \Delta t_i * E(x, y))^2 \qquad \textbf{2-10}$$

where X and Y are the dimensions of the image in horizontal and vertical directions respectively. The solution of the equation is found by taking the gradient of the equation and equating the result of gradient to zero. Then, HDR radiance map is calculated as follows:

$$L(x, y) \approx E(x, y) = \frac{\sum_{i=1}^{N} w(C_i(x, y)) * \Delta t_i * C_i(x, y)}{\sum_{i=1}^{N} w(C_i(x, y)) * \Delta t_i^2} \qquad \textbf{2-11}$$

where $C_i(x, y)$ is calculated by inverse CRF. Resultant equation gives higher weights to long exposed images to decrease the quantization error.

2.3 Tone Mapping

The real world consists of high dynamic range scenes that cannot be captured by standard digital cameras. However, HDR radiance map (HDR images) can be calculated by the previously explained methods. There are many advantages of using HDR radiance map, but it also brings some difficulties. Displaying these images on the standard screens is impossible due to low dynamic range of the screens. Therefore, a compression technique is needed to map HDR radiance maps to low dynamic range images. These techniques are called as tone mapping

or tone reproduction operators [13]. The quality of these operators is measured by the contrast of output images and resulted artifacts.

Tone mapping operators are divided into two groups. These groups are called as global and local tone mapping operators. The global tone mapping operators are also known as spatially invariant operators [7, 14]. They use the same mapping function for all pixels regardless of surrounding pixel values. The advantages of the global tone mapping operators are their simplicity and fast processing time. However, using global information may cause the loss of details when the image dynamic range is high.

In order to solve the previously stated problem, local tone mapping operators are developed. The local tone mapping operators use the neighborhood information of pixels in order to calculate necessary mapping function [23-26]. The mapping function changes locally depending on the pixel location so it provides more contrast at the output images. The disadvantages of these operators are their high computational loads and the determination of the neighborhood size.

In the following subsections, the implemented tone mapping algorithms are explained. These algorithms are the former and basic tone mapping algorithms. They are implemented to give the principles of the tone mapping operators. However, there are also more complex and successful tone mapping algorithms in the literature [62-64].

2.3.1   TM-ALG1 [38]

The simplest tone mapping operator is linear scaling which you can use to map HDR radiance maps into low dynamic range images. In this method, the whole dynamic range is mapped linearly regardless of the distribution of the details in the dynamic range. Therefore, it does not preserve the details of HDR radiance map in the output image. Linear scaling function is given as follows:

$$L_d(x, y) = \frac{L(x, y) - L_{min}}{L_{max} - L_{min}} \qquad \text{2-12}$$

where $L(x, y)$ is the scene radiance, and $L_d(x, y)$ is the display luminance. L_{max} is the maximum scene radiance, and L_{min} is the minimum scene radiance.

2.3.2   TM-ALG2 [15]

Logarithmic mapping is another simple dynamic range compression operator. Although this operator is simple to achieve satisfactory results, this mapping operator is more successful than the linear mapping operator. In this case, a nonlinear mapping is used to compress the high dynamic range of HDR radiance map. The operator does not perform compression evenly for the whole dynamic range. The specific intervals in the dynamic range are compressed less than the remaining intervals. This approach prevents losing details in this interval but the remaining of the dynamic range is highly compressed. By doing so, the information in HDR radiance map is transferred with the minimum loss in the details. The logarithmic dynamic range compression is given as follows:

$$L_d(x, y) = \frac{log_{10}(L(x, y) + 1)}{log_{10}(L_{max} + 1)} \qquad \text{2-13}$$

2.3.3   TM-ALG3 [15]

This algorithm requires less tuning parameter than the former algorithms and is based on logarithmic compression. In this algorithm, the logarithmic function, which is proposed by Stockham [16], is used to model the relation between brightness and luminance. This function is given as follows:

$$L_d(x, y) = \frac{\log (L(x, y) + 1)}{\log (L_{max} + 1)} \qquad \text{2-14}$$

This function forms a basis for the proposed tone mapping operator.

In the algorithm, a preprocessing operation is applied before mapping high dynamic range scene into low dynamic range image. This operation calculates the brightness level of the input image in order to provide adaptation for various displays. For this purpose, the log-average of the input image is calculated. Then, the calculated value and a user defined parameter are used to scale the scene radiance. Hence, the brightness level of the output image can be adjusted by this scaling operation.

In addition, an adaptive compression is used to preserve the details in the HDR radiance map. This is achieved by adaptive changing the logarithmic base of the mapping function depending on pixel radiance. As a result of this, the details in relatively dark areas are preserved, and high radiance values are highly compressed. The smooth transition between the logarithmic bases is obtained by using a bias function. This function, which is proposed by Perlin and Hoffert [18], is given as follows:

$$bias_b(t) = t^{\frac{\log(b)}{\log(0.5)}} \qquad \text{2-15}$$

Then, the developed tone mapping operator is proposed. This new operator is given as follows:

$$L_d(x,y) = \frac{L_{dmax} * 0.01}{log_{10}(\tilde{L}_{max} + 1)} \frac{\log(\tilde{L}(x,y) + 1)}{\log\left(2 + \left(\left(\frac{\tilde{L}(x,y)}{\tilde{L}_{max}}\right)^{\frac{\log(b)}{\log(0.5)}}\right) * 8\right)} \qquad \text{2-16}$$

where L_{dmax} is the maximum luminance limit of the display, $\tilde{L}_{max}$ is the scaled maximum scene radiance, and $\tilde{L}$ is the scaled scene radiance. The proposed

operator is adaptive for any display medium by the help of L_{dmax}. Moreover, the optimum value of the bias parameter b is given as *0.85*.

2.3.4   TM-ALG4 [17]

This algorithm is composed of two different tone mapping algorithms which are independent operators. One of them is global tone mapping operator, and the other one is local tone mapping operator. These new operators take traditional photographic methods as bases and solve the problem of mapping HDR radiance maps into low dynamic range images by the help of these methods [17]. Zone system is one of the traditional photography methods, and it provides a systematical approach for photographers in order to define the relationship between the visualization of the scene and the final print [19].

The calculation of the global operator starts by the calculation of scene's key value. Key value is used to define the brightness of a scene. Bright scenes such as white clouds have high key values, and dark scenes such as shadow of an item have low key values. Log-average is used to calculate scene's key value, and the calculation of log-average is given as follows:

$$\bar{L} = \exp\left(\frac{1}{X * Y} \sum_{x=1,y=1}^{X,Y} \log(\delta + L(x,y))\right)$$
2-17

where $\bar{L}$ is the result of log-average, and δ denotes a small offset which is used to avoid infinite results. Then, a user defined parameter a is used to scale the scene radiance. The scaling operation is given as follows:

$$\tilde{L}(x,y) = \frac{a}{\bar{L}} L(x,y)$$
2-18

15

The log-average of the scene can be mapped to different values by changing the parameter *a*. Then, a tone mapping method is used to compress HDR radiance map and is given as follows:

$$L_d(x,y) = \frac{\tilde{L}(x,y)}{1 + \tilde{L}(x,y)} \qquad \text{2-19}$$

As can be seen from the above equation, high luminance values are mapped to one, and low luminance values are almost not changed. Then, the operator is extended to provide high contrast outputs. The extended operator allows high luminance values to be saturated. This process is the same as the burning operation defined in traditional photography. The extended operator is given as follows:

$$L_d(x,y) = \frac{\tilde{L}(x,y)(1 + \frac{\tilde{L}(x,y)}{L_{white}^2})}{1 + \tilde{L}(x,y)} \qquad \text{2-20}$$

where L_{white} denotes the minimum luminance value and is called as white threshold level. The radiance values which are greater than this threshold value are mapped to the maximum display intensity. Resultant global tone mapping operator gives good results for low dynamic range scenes. However, this operator is not good enough to preserve the details in high dynamic range scenes.

In order to solve this problem, a new local tone mapping operator is proposed. This new operator uses the dodging and burning techniques in the traditional photography. Therefore, it is suggested to define a key value for each pixel in the input image in order to preserve the details in high dynamic range scene. This operator is applied within a local neighborhood. The size of the neighborhood is determined by the help of a local contrast measure which is performed on different scales. The measure is defined as the subtraction of two circular symmetric Gaussian convolved images. The used circularly symmetric Gaussian filter is given as follows:

16

$$R_i(x, y, s) = \frac{1}{\pi(\alpha_i s)^2} \exp\left(-\frac{x^2 + y^2}{(\alpha_i s)^2}\right)$$

$$\tag{2-21}$$

$$V_i(x, y, s) = \tilde{L}(x, y) \otimes R_i(x, y, s)$$

$$\tag{2-22}$$

where V_i is the convolution result of Gaussian filter for scale s. Then, the contrast measure is defined as follows:

$$V(x, y, s) = \frac{V_1(x, y, s) - V_2(x, y, s)}{\dfrac{2^\phi a}{s^2} + V_1(x, y, s)}$$

$$\tag{2-23}$$

where ϕ is the sharpening parameter, and $V_1(x, y, s)$ has a smaller radius than $V_2(x, y, s)$. In the algorithm, the ratio between the radiuses of two center surrounded filters is used as *1.6*. Furthermore, in the equation, $2^\phi a/s^2$ prevents the response of the function becomes infinite when $V_1(x, y, s)$ becomes zero. As stated previously, equation 2-23 is used to measure the contrast of a local area. For that purpose, equation 2-23 is iterated for different scales by starting from the lowest scale. At each scale, the absolute value of equation 2-23 is compared to a predefined threshold value. The last scale s_m which is smaller than the threshold is selected as the local neighborhood size of the current pixel. Then, this process is performed for each pixel in the input image.

Since $V_1(x, y, s_m)$ defines the local average of the pixels, the tone mapping operator in equation 2-19 is converted into local tone mapping operator by replacing $\tilde{L}(x, y)$ with $V_1(x, y, s_m)$. By doing that, the global key value is replaced with a local key value. Then, the local tone mapping operator is given as follows:

$$L_d(x, y) = \frac{\tilde{L}(x, y)}{1 + V_1(x, y, s_m(x, y))}$$

$$\tag{2-24}$$

At a dark pixel surrounded by a relatively bright region, the relation is $V_1(x, y, s_m(x, y)) > \tilde{L}(x, y)$. Since the local average of this area is greater than the radiance value of the dark pixel, the operator decreases the radiance values of the pixel. At a bright pixel surrounded by a relatively dark region, the relation is $V_1(x, y, s_m(x, y)) < \tilde{L}(x, y)$. Therefore, the operator increases the radiance value of the pixel. As a result, the local contrast is enhanced in the output image. Furthermore, the operator does not guarantee to keep the radiance values in the allowable range. Therefore, the resultant pixel values may become greater than the range. In this case, these pixels become saturated. It has the same effect as equation 2-20.

2.3.5   TM-ALG5 [13]

The proposed approach does not require any manual tuning parameters, so it makes the operator adaptive. This approach solves the problem of manual tuning requirement for different input images and displays. The proposed operator is also computationally efficient because it uses the some of the pixels in HDR radiance map in order to calculate the required parameters of the operator.

The operator is composed of two stages. The first stage, which takes references from Tumblin-Rushmeier's and Reinhard's [20, 21] operators, is given as follows:

$$L_d(x, y) = m(\bar{L}) * L_{da} \left(\frac{L(x, y)}{\bar{L}} \right)^{\left(\frac{\gamma_w}{\gamma_d} \right)} \qquad \text{2-25}$$

where $\bar{L}$ is the log-average, and L_{da} is the display luminance. $\bar{L}$ is calculated by equation 2-17. The exponents γ_w and γ_d are used to control the contrast of the output image. The formula to calculate these exponents is given as follows:

$$\gamma(\bar{L}) = \begin{array}{l} 2.655, \quad if \ \bar{L} > 100 \ cd/m^2 \\ 1.855 + 0.4 \ log_{10}(\bar{L} + 2.3x10^{-5}) \ otherwise \end{array} \qquad \text{2-26}$$

18

where $\gamma(L_{da}) = \gamma_d$, and $\gamma(\bar{L}) = \gamma_w$. Mathematical expression to calculate $m(\bar{L})$ coefficient is proportional to the scene adaptation luminance and is given as follows:

$$m(\bar{L}) = \left(\sqrt{C_{max}}\right)^{(\gamma_{wd}-1)}$$

$$\gamma_{wd} = \left(\frac{\gamma_w}{1.855 + 0.4\log (L_{da})}\right)$$

2-27

where C_{max} is the display maximum contrast limit. As it has been mentioned previously, the above equations do not use all of the pixels of HDR radiance map. The small number of pixels is used to calculate necessary parameters, so it decreases computational load of the operator.

As stated previously, there is a second stage that compensates the oversaturated results of the first stage. In the second stage, the global tone mapping operator which is given in equation 2-20 is used for this purpose, and L_{white} in equation 2-20 is chosen such that %1 of the HDR radiance map pixels pass this value.

For tuning the parameters, $\bar{L}$ and $L_{white,}$ the operator includes preprocessing steps. When the HDR radiance map contains many dark regions, the operator may result in oversaturated regions due to equation 2-26. Dark pixels, which are below a specific threshold, are discarded during the calculation of $\bar{L}$ in order to solve this problem. The threshold is given as follows:

$$L_{trsh} = \min (\bar{L}/20, L_{white}/100)$$

2-28

where L_{trsh} is the defined threshold for the calculation of $\bar{L}$. Furthermore, a fine tuning method to determine white threshold level, L_{white}, is used in the operator. The algorithms states that initially used white threshold level gives satisfactory results for most of the scenes. However, it may fail when the scene contains so many bright regions. Therefore, the following equations are solved jointly to

calculate more proper white threshold level. These equations are defined as follows:

$$L_d = m(\bar{L}) * L_{da} * \left(\frac{L_{white}}{\bar{L}}\right)^{\left(\frac{\gamma_w}{\gamma_d}\right)}$$

$$L_{dwt} = \frac{L_d * \left(1 + \frac{L_d}{L_{fwhite}^2}\right)}{1 + L_d}$$

2-29

where L_{fwt} is the new white threshold level, and L_{dwt} is the display white threshold. L_{dwt} is generally chosen as *0.98* for canonical range. Then, the new white threshold level is used in the second stage of the proposed operator as L_{white}, and the mapping process is finalized. All the other post processing steps in the algorithm is not implemented in this thesis.

CHAPTER 3

EXPOSURE FUSION

3.1 Introduction

The term fusion is used to define the general procedure of extracting necessary information from different domains. The goal of image fusion is integrating information that belongs to different sources or has different characteristics into an enhanced single image. The fused image contains enhanced details that cannot be achieved by using only one of the input images.

Exposure fusion (EF), also called as multi exposure image fusion, is one of the fusion methods. This fusion method is used to blend differently exposed images into a single image which is well-exposed and represents the whole information of captured scene. The main goal of EF is creating high contrast images that represent the high dynamic range of captured scenes in low dynamic range. The whole process is performed in low dynamic range unlike HDRI. Since the fusion process is performed in low dynamic range, the fused image does not require any additional process to display it on the standard screens. In addition, EF methods do not need any prior information about exposure settings of input images and do not calculate CRF. These are the advantages of EF. The general EF scheme is given as follows:
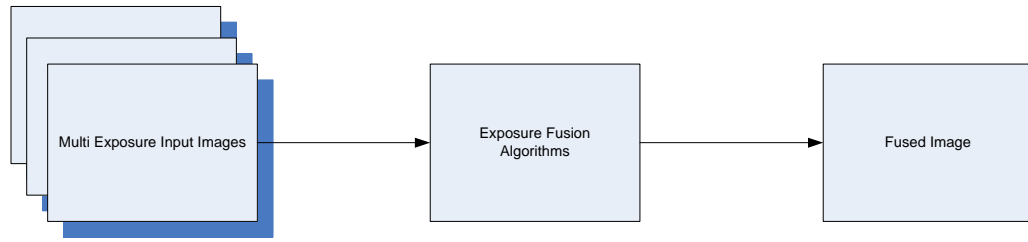
Figure 3.1 General Exposure Fusion Scheme

Image fusion is usually performed on different processing levels. These are pixel, feature and symbol levels. Each group defines different levels of information. In other words, they are different image information representations. Pixel level image fusion performs fusion process in pixel basis. Input images are combined to form a single fused image. Feature level image fusion performs fusion process on features or object labels which are already extracted from input images. Symbol level image fusion uses symbols to describe the information in input images and uses these symbols for image fusion [27].

In this thesis, pixel level EF methods are studied because they are the most popular and common EF methods. The following section describes pixel level methods used in EF and explains the implemented EF algorithms.

3.2   Pixel Level Image Fusion Methods

Pixel based image fusion methods combine multiple pre-registered input images from different sensors into a single fused image. The main goal of pixel based image fusion can be defined as; transferring salient information in the input images to the fused image by preserving the details of the input images and not causing any artifacts in the fused image [27].

Briefly, the success of the methods is defined as preserving the salient features of the input images in the fused image. However, in real life, it is usually not possible to preserve all important details. Therefore, the definition of image fusion

should be changed into this; pixel based image fusion tries to transfer the most important details of the input images to the fused image. At this point, another issue appears. It is stated that fusion defines the important details in the input images. The definition of the important details is a problem for image fusion methods [27].

Introducing artifacts and losing some important features of the input images are the main problems of pixel based methods. In image fusion, some unimportant details can be transferred to the fused image. This may be a form of transferring features that do not exist in the input images. Moreover, some features that are evaluated as important can be transferred to the fused image with some additive noise or unwanted effects. This leads to lose some important details of the input images. In addition, these features may cause misunderstanding and can be considered as important. As a conclusion, transferring details from the input images to the fused image is very critical.

In this thesis, the well-known pixel level EF algorithms are implemented. Other than the implemented algorithms, there are also several different EF algorithms in the literature [42-47]. However, the implemented algorithms include main methods which are used in EF. In the following sections, the algorithms are classified as space domain, transform domain and multiscale decomposition based methods. In the following sections, the implemented EF algorithms are explained.

3.2.1   Space Domain Based Methods

These methods perform fusion process in the space domain. Therefore, there is no need to use any transform or decomposition operators for image fusion. The advantages of these methods are that the computation load of the methods is low, and they can be implemented easily on embedded systems. Let there are N input images namely $I_i(x, y)$, $i \in 1, 2, ..., N$, and $I_f(x, y)$ is the fused image. If F is the

fusion operator, then the fusion process in the space domain can be written as follows:

$$I_f(x,y) = F\big(I_1(x,y), I_2(x,y), \dots, I_f(x,y) \big)$$

3-1

Moreover, these methods can be either pixel based or blocked based. In pixel based methods, the fusion process is performed in pixel basis. Therefore, their computational cost is a little bit high. In block based methods, the fusion process is performed on blocks rather than pixels. The goal of these methods is to decrease the computational load of the fusion process by working block based. In block based methods, firstly, the input images are divided into blocks. Then, the fusion rules are applied to each block. Due to the block based processing, resultant image includes blocking artifacts. To solve this problem, these methods include an additional processing step namely blending. This step performs the blending of the blocks to provide smooth transitions at the borders of the blocks. An image with a size of *MxN*, which is divided into non-overlapping blocks with a size of *KxL*, is shown as follows:
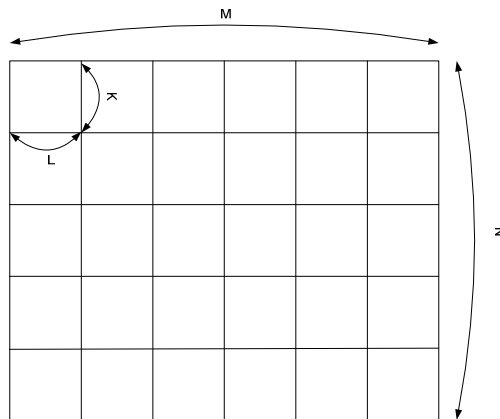
Figure 3.2 Illustration of Image Blocks in Block Based Methods

The following subsections explain some examples of the space domain based methods in EF.

3.2.1.1   EF-ALG1 & EF-ALG2 [28]

These methods, EF-ALG1 and EF-ALG2, are the simplest methods for image fusion. In EF-ALG1, the fusion operator calculates the average of the input images at each pixel coordinate. If $I_f(x, y)$ denotes the fused image, the fusion process is written as follows [28]:

$$I_f(x, y) = \frac{1}{N} * \sum_{i=1}^{N} I_i(x, y) \qquad \text{3-2}$$

where $I_i(x, y)$ represents the ith input image, N is the number of input images. Although it is very simple and basic, it gives good results when the input images have same characteristic. This method requires a very accurate image alignment due to pixel by pixel processing. If the input images are not aligned accurately, this method may cause problems such as the loss of contrast, blurring and false objects in the fused image. Other than computational advantage, this method provides good noise removal. The disadvantages of the method are the loss of contrast and the attenuation of salient features.

This method can also be alternated with a weighted average operation, EF-ALG2. This operation provides us to control the contribution of each input image to the fused image. If $k_1, k_2 \ldots k_N$ are the pixel independent weights, the fusion process is written as follows:

$$I_f(x, y) = \frac{k_1 * I_1(x, y) + \cdots + k_N * I_N(x, y)}{k_1 + \cdots + k_N} \qquad \text{3-3}$$

3.2.1.2   EF-ALG3 [29]

This algorithm is another example of the pixel based methods. The goal of the algorithm is to develop an algorithm that works on embedded processors and

25

mobile devices in real-time [29]. Therefore, the number of the input images to be fused is limited to two. Additionally, the algorithm uses three different blending functions for image fusion in order to eliminate expensive block or transform domain processes. The methodology can be explained symbolically as follows:
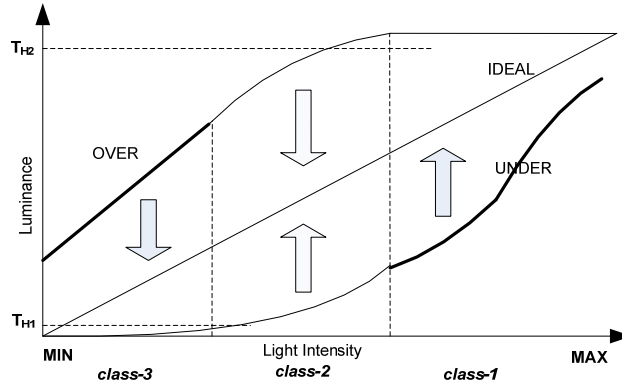
Figure 3.3 The Pixel Classes in EF-ALG3

The proposed method aims to form an ideal transfer function between the light intensity and the observed luminance. First of all, two thresholds TH_1 and TH_2 are defined in order to indicate the saturated regions in the input images. The algorithm uses TH_1 as %5 of the maximum intensity and uses TH_2 as %95 of the same value. Then, three different mapping functions are defined for the fusion process. Pixel intensity is used to choose which mapping function is going to be used for image fusion.

The first mapping function is used for the overexposed pixels. In the following equations, $I_{un}(x, y)$ represents the short exposed image, and $I_{ov}(x, y)$ represents the long exposed image. The difference between these two images is denoted by $I_{diff}(x, y)$. Then, the first mapping function is given as follows [29]:

$$I_f(x, y) = I_{un}(x, y) + \frac{I_{diff}(x, y)}{4}$$

3-4

26

This process is a simple brightness boost of the short exposed image.

The second mapping function is used for pixels which are not in the saturated regions. In this class, the weighted average of the input images is used for image fusion. The mapping function is given as follows:

$$I_f(x,y) = \frac{\left(I_{ov}(x,y) - \frac{I_{diff}(x,y)}{4}\right)(M - I_{ov}(x,y))}{M - I_{diff}(x,y)}$$

$$+ \frac{\left(I_{un}(x,y) + \frac{I_{diff}(x,y)}{4}\right)I_{un}(x,y)}{M - I_{diff}(x,y)}$$

<div align="right">3-5</div>

where M is the maximum pixel intensity value. For example, M is 255 when the pixel depth is 8-bit. By means of this mapping function, rapid intensity changes are prevented in the fused image.

The third mapping function which is defined below is used for the underexposed pixels and is very similar to the first mapping function.

$$I_f(x,y) = I_{ov}(x,y) - \frac{I_{diff}(x,y)}{4}$$

<div align="right">3-6</div>

In order to match the pixel intensity to the ideal pixel intensity, the brightness of the short exposed image is reduced by one quarter of the difference between two input images.

These fusion rules are applied to each pixel in the input images. Firstly, the pixel of the short exposed image is compared to TH_1. If it is smaller than this threshold value, the first mapping function is used in the fusion process. If this condition is not satisfied, the pixel of the long exposed image is compared to TH_2. If the pixel value is greater than TH_2, the third mapping function is used for image fusion. The

second mapping function is used for pixels that do not satisfy the above conditions. As it is seen from the fusion steps, the pixels of the long exposed image have higher priority in the fusion process. This priority provides brighter fused images.

3.2.1.3   EF-ALG4 [30]

This algorithm is one of the examples for the block based EF methods [30]. The proposed algorithm is different from the previously proposed algorithms. The former algorithms use pixel intensity averaging for image fusion. Therefore, the average of the input images is transferred to the fused image. This operation results low contrast and unsatisfactory fused images. However, this algorithm divides the input images into blocks and applies fusion rules to the blocks. Then, a blending function is performed to smoothly combine the blocks. Moreover, this algorithm uses a gradient ascent method in order to find an optimum block size. The methodology can be explained as follows [30]:

First of all, all input images are divided into blocks with a size of *dxd*. Then, the most detailed input image is selected in each block. The algorithm states that the block size should be greater than *16x16* to obtain accurate results. Moreover, the size of the blocks (*d*) is one of the input parameters and must be applied to the algorithm. The algorithm uses entropy to estimate the details of the images. For grayscale images, entropy equation is given as follows:

$$E_g = \sum_{z=0}^{255} -p(z) * \log_2(p(z))$$

3-7

where E_g is the entropy. In the equation, $p(z)$ is the probability of a pixel that has the intensity of z. After computing histogram of the image, $p(z)$ is calculated as n_z/n. The number of pixels that have intensity z is denoted by n_z, and n is the number of pixels in the images. In each block, the selection is made by comparing

28

the input images with respect to their entropies. Then, the input image with the highest entropy value is selected as the most detailed image for that block.

Let N is the number of input images, and each image is divided into an array of $n_r x n_c$. In an image, k and l are the row and the column indices of the blocks. $I_{kl}(x, y)$ represents the image that has the highest entropy in the block kl. Then, the blending process is given as follows:

$$O(x, y) = \sum_{k=1}^{n_r} \sum_{l=1}^{n_c} W_{kl}(x, y) * I_{kl}(x, y)$$

3-8

where *O(x,y)* is the output of the blending process, and $W_{kl}(x,y)$ is the blending function that is centered at kl^{th} block at location *(x,y)*. The blending function provides smoothing operation at the borders of the blocks. In the algorithm, Rational Gaussian (RaG) surface is used as the blending function. The blending function that is centered at kl^{th} block is given as follows:

$$W_{kl}(x, y) = \frac{G_{kl}(x, y)}{\sum_{m=1}^{n_r} \sum_{n=1}^{n_c} G_{mn}(x, y)}$$

3-9

As can be seen from the above equation, the blending function is normalized. Therefore, it has the values between zero and one. Then, *G_{kl}(x,y)* is defined as follows:

$$G_{kl}(x, y) = exp\left(-\frac{(x - x_{kl})^2 - (y - y_{kl})^2}{2 * \sigma^2}\right)$$

3-10

where *(x_{kl},y_{kl})* is the center coordinate of kl^{th} block, and σ is the sigma of Gaussian function. The sigma is the second parameter of the algorithm and should be manually tuned.

As it has been mentioned previously, the proposed algorithm requires two input parameters which are the block size (d) and the sigma (σ). It is proposed to use gradient ascent method to search for the optimum values of the parameters. The method starts from an initial d, σ and then performs previously defined fusion steps. At the end of the fusion, the entropy of the fused image is calculated, and gradient direction that maximizes the entropy is found. Then, these parameters are updated until each parameter does not change so much between iterations. By means of this method, the optimum values of the parameters are found for the given input images. However, in this thesis, gradient ascent method is not implemented, and the parameters are manually tuned to avoid long computation times.

3.2.1.4   EF-ALG5 [31]

This algorithm is another example of the block based image fusion methods [31]. The main goal of this algorithm is to decrease the high computational load of EF-ALG4 [30]. The former algorithm uses Rational Gaussian function for the blending operation, and this function is too complex to implement in real-time systems. Therefore, the blending function is changed with simple interpolation functions. The only required parameter is the block size (d) for this algorithm. The methodology can be explained as follows:

First of all, the input images are divided into the blocks, and the entropy of these blocks are calculated. These steps are the same as the ones in the previous algorithm. The difference between these two algorithms is the blending function used for image fusion. There are three blending functions in this algorithm, and they are interpolation functions. These functions are used for corner, edge and interior pixels.
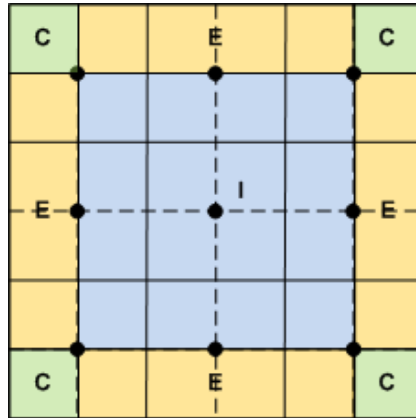
Figure 3.4 The Pixel Classes in EF-ALG5

In the figure, C, E and I represent the corner, edge and interior pixel classes in the image. The black dots are the centers of the blocks.

If $I_i(x,y), I_j(x,y), I_u(x,y)$ and $I_v(x,y)$ represent the input images where $i,j,u,v = 1,2,\ldots,N$, bilinear interpolation for interior pixels is given as follows:

$$p(x,y) = \frac{s}{r+s}\left(\frac{n}{m+n}I_i(x,y) + \frac{m}{m+n}I_j(x,y)\right) + \qquad \text{3-11}$$
$$\frac{r}{r+s}\left(\frac{n}{m+n}I_u(x,y) + \frac{m}{m+n}I_v(x,y)\right)$$
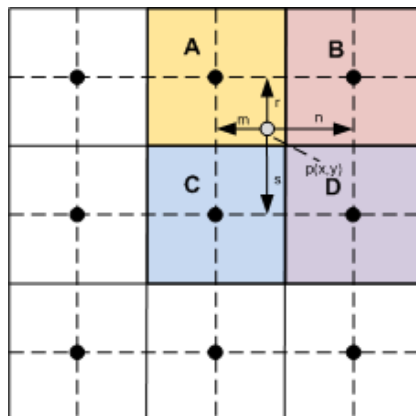
Figure 3.5 Bilinear Interpolation at p(x,y)

31

$I_i(x, y)$ is the input image that has the highest entropy in the block A. The others $I_j(x, y)$, $I_u(x, y)$ and $I_v(x, y)$ are the input images that are selected in the block B, C and D respectively. This fusion process is repeated for each interior pixel by considering its four neighbor blocks. Furthermore, linear interpolation is used for edge pixels and is given as follows:

$$p(x, y) = \left(\frac{n}{m + n} I_i(x, y) + \frac{m}{m + n} I_j(x, y) \right)$$

3-12

This function is a weighted average of two input images. For corner pixels, simple mapping function such as the pixel intensity of the closest block is used for image fusion.

3.2.2   Transform Domain Based Methods

Image fusion can also be performed in the transform domain. Methods which perform the fusion process in the transform domain are called as transform domain based methods. In these methods, the input images are transformed to a transform domain, and then the fusion rules are applied in this domain. To get the fused image, inverse transform is used to turn back to the spatial domain. If the transform operator is denoted by T, the fusion process in the transform domain can be represented as follows:

$$I_f(x, y) = T^{-1}\{ F\left(T\{I_1(x, y)\}, T\{I_2(x, y)\}, \ldots, T\{I_f(x, y)\} \right) \}$$

3-13

where T^{-1} represents inverse transform operator.

In the literature, there are several transform domain based methods. However, these methods mostly use Discrete Cosine Transform (DCT) for image fusion. Therefore, two DCT based image fusion algorithms are implemented in this thesis.

The following subsections explain the implemented DCT based EF algorithms.

3.2.2.1   EF-ALG6 [32]

This study introduces a new DCT based image fusion algorithm, the performance of which is similar to wavelet or pyramid based image fusion methods. However, the algorithm is more computationally efficient and simpler than multiscale decomposition based algorithms. Moreover, the proposed algorithm can be embedded into popular Joint Photographic Experts Group (JPEG) format. This operation can save time when the fused image is transmitted or stored in JPEG format.

The algorithm uses DCT coefficients to define a local contrast measure for image fusion. This measure is called as *contrast measure*. In the algorithm, 8x8 DCT is used for image fusion. One output of 8x8 DCT includes 64 different coefficients. Top-left coefficient is called as DC coefficient, and other remaining 63 coefficients are called as AC coefficients. The frequency content of the AC coefficients increases from top-left to bottom-right in a zigzag manner. The highest frequency component is represented by the bottom-right coefficient. The rest of the algorithm is explained below [32]:

In the algorithm, 15 different energy bands are defined in the DCT coefficients. n^{th} energy band contains coefficients that satisfy this equation $n = x + y$. Then, the contrast measure is defined as follows:

$$C(x, y) = \frac{d(x, y)}{\sum_{k=0}^{n-1} E_k}$$

3-14

$$E_k = \frac{\sum_{t+p=k} |d(p, t)|}{N}$$

3-15

$$N = \begin{cases} k + 1, & k < 8 \\ 14 - k + 1, & k \geq 8 \end{cases}$$

3-16

where *C(x,y)* is the contrast measure at location *(x,y)*, E_k is k^{th} energy band, and *d(x,y)* is the DCT coefficient.

After the contrast measure is calculated, the fusion process is performed. Assume that there are *N* input images namely, $I_1, I_2, ..., I_N$. The DCT transforms of the input images are $I_1^D, I_2^D, ..., I_N^D$. Moreover, $I_{i,t}^D$ represents the DCT output of tth block in i^{th} image.

In the algorithm, the fusion is performed in two different ways. The algorithm states that taking the average of the DC coefficients is enough to blend coarse details. Therefore, it uses averaging operation for fusing the DC coefficients. This process is given as follows:

$$I_{f,t}^D(0,0) = \frac{1}{N} \sum_{i=1}^{N} I_{i,t}^D(0,0) \qquad \text{3-17}$$

where $I_{f,t}^D(0,0)$ is the DC coefficient of tth block in the fused image.

The fusion of the AC coefficients is performed by considering the contrast measure of the AC coefficients. For the AC coefficients, the fusion process is given as follows:

$$I_{f,t}^D(m,n) = I_{q,t}^D(m,n), m \neq 0 \text{ and } n \neq 0 \qquad \text{3-18}$$

$$q = arg\max_i (C_{i,t}(m,n)) \qquad \text{3-19}$$

Finally, the fused image is calculated by IDCT transform. The explained steps form the proposed fusion algorithm namely *EF-ALG6-1*.

Moreover, there is another fusion method that is a variant of the first method. In the first method, averaging operation is only used for the DC coefficients but in

this algorithm, averaging operation is performed for both the DC and AC coefficients. This second method is given as follows:

$$I_{f,t}^D(m,n) = \frac{1}{N}\sum_{i=1}^{N} I_{i,t}{}^D(m,n)$$ **3-20**

where $I_{f,t}^D(m,n)$ is the tth block in the fused image. Then, IDCT is applied to get the fused image. This is the second fusion method namely *EF-ALG6-2*.

3.2.2.1.1   Proposed Improvement

A slightly modified version of EF-ALG6-1 has been developed in this thesis. The goal of this improvement is to decrease the computational cost of the original method by removing the contrast measure from the fusion process. In the original algorithm, the AC coefficients are chosen by comparing the contrast measures of the input images. However, the improved method directly uses the AC coefficients rather than the contrast coefficients. The AC coefficient with the maximum absolute value is selected as the AC coefficient of the fused image. This operation can be explained as follows:

$$I_{f,n}^D(m,n) = I_{t,n}^D(m,n), m \neq 0 \ and \ n \neq 0$$ **3-21**

$$t = arg\max_i (I_{i,n}(x,y))$$ **3-22**

Then, the same steps are performed as explained in the original algorithm. This algorithm is called as *EF-ALG6-3* in this thesis.

3.2.2.2   EF-ALG7 [33]

This algorithm is another DCT based fusion method for multi exposure and multi focus images [33]. The main goal of the method is to develop a low computational

image fusion algorithm that can work in real-time applications. It is stated that image fusion has been performed on coded images until that time. However, when the DCT method is used, the fusion can be performed on lossless data. JPEG format uses DCT coding as one of the blocks which is fed by a raw data. Therefore, the fusion process can be performed on raw data before it is compressed. Yet another advantage of the algorithm is to be applicable for both multi exposure and multi focus image fusions.

Furthermore, DCT is widely used in coding applications thanks to its low computational cost. Hence, the method uses 8x8 DCT which is similar within JPEG coding. The methodology can be explained as follows:

Assume that there are five different input images which are shown as I_{--}, I_{-}, I, I_{+}, I_{++}. *(--), (-) , (++)* and *(+)* are the underexposed and the overexposed images respectively. *(--)* and *(++)* show the highest underexposed and overexposed images. The center image represents an image which is captured by the automatic exposure setting. A DCT block with a size of *kxk* is shown as follows:

$$I^D = (dc, ac_1, ac_2, \dots, ac_{k^2-1}) = (dc, ac) \qquad \textbf{3-23}$$

where *dc* is the DC coefficient, and *ac* is the AC coefficient. Then, the DCT transform of the input images are $I^D_{--,j}, I^D_{-,j}, I^D_j, I^D_{+,j}, I^D_{++,j}$.

$$I^D_t = \left(dc_t, ac_{t,1}, ac_{t,2}, \dots, ac_{t,k^2-1}\right) = (dc_t, ac_t)$$
$$I^D_{--,t} = \left(dc_t{}^{--}, ac_{t,1}{}^{--}, ac_{t,2}{}^{--}, \dots, ac_{t,k^2-1}{}^{--}\right) = (dc_t{}^{--}, ac_t{}^{-})$$
$$I^D_{-,t} = \left(dc_t{}^{-}, ac_{t,1}{}^{-}, ac_{t,2}{}^{-}, \dots, ac_{t,k^2-1}{}^{-}\right) = (dc_t{}^{-}, ac_t{}^{-}) \qquad \textbf{3-24}$$
$$I^D_{++,t} = \left(dc_t{}^{++}, ac_{t,1}{}^{++}, ac_{t,2}{}^{++}, \dots, ac_{t,k^2-1}{}^{++}\right) = \left(dc_t{}^{++}, ac_t{}^{++}\right)$$
$$I^D_{+,t} = \left(dc_t{}^{+}, ac_{t,1}{}^{+}, ac_{t,2}{}^{+}, \dots, ac_{t,k^2-1}{}^{+}\right) = (dc_t{}^{+}, ac_t{}^{+})$$

The fusion process is composed of two different rules. These rules are used to fuse the DC coefficients and the AC coefficients separately. If $I_{f,t}^{D} = (dc_t^{*}, ac_t^{*})$ is the tth block of the fused image, the fusion rules are given as follows:

$$dc_t^{*} = average(dc_t^{--}, dc_t^{-}, dc_t, dc_t^{+}, dc_t^{++}) \qquad \textbf{3-25}$$

$$ac_t^{*} = max\ (|ac_t^{--}|_{L2}, |ac_t^{-}|_{L2}, |ac_t|_{L2,}, |ac_t^{+}|_{L2}, |ac_t^{++}|_{L2}) \qquad \textbf{3-26}$$

For the fusion of the AC coefficients, L2 norm that is known as Euclidean norm is used to define the most detailed input image block. The vector norm function is given as follows:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ . \\ x_N \end{bmatrix} \qquad \textbf{3-27}$$

$$|x|_{L2} = \sqrt{\sum_{i=1}^{N} |x_i|^2} \qquad \textbf{3-28}$$

The previously defined fusion rules are applied to each input image block, and then the fused image is calculated by inverse IDCT. The explained steps form the proposed fusion algorithm namely *EF-ALG7-1*.

3.2.2.2.1 Proposed Improvement

To solve the blocking problem, a modified version of this algorithm has been implemented in the thesis. As it has been explained previously, the original method takes the average of the DC coefficients and selects the AC coefficients with the largest L2 norm from the input images for image fusion. In addition to

selecting the AC coefficients with respect to their L2 norms, weighted average can be added into the fusion process. In the improved algorithm, the L2 norm of the AC coefficients is calculated, and the block with the maximum L2 norm is chosen from the input images. Then, it uses the weighted average of the AC coefficients instead of directly copying the chosen AC coefficients to the fused image. The mathematical explanation of the improved method is given as follows:

$$dc_t{}^* = average(dc_t{}^{--}, dc_t{}^-, dc_t, dc_t{}^+, dc_t{}^{++}) \qquad \textbf{3-29}$$

$$ac_t{}^* = average\ (ac_t{}^{--}, ac_t{}^-, ac_t, ac_t{}^+, ac_t{}^{++}, 2 * ac_t^x) \qquad \textbf{3-30}$$

where x denotes the AC coefficients of the input images with the maximum L2 norm. Then, the fused image is calculated by IDCT. This algorithm is called as *EF-ALG7-2* in this thesis.

3.2.3   Multiscale Decomposition Based Methods

Multiscale decomposition (MD) based methods are the third group of pixel level image fusion methods. MD methods are widely used in image fusion algorithms. These methods are of great importance due to the capability of extracting salient features for the purpose of image fusion. Even though it is not possible to obtain detailed fused images with most of the non-multiscale decomposition based methods, it can be achieved with MD based methods.

If $\psi^\uparrow$ and $\psi^\downarrow$ are decomposition and reconstruction operators respectively, the fusion process in multiscale decomposition can be represented as follows:

$$I_f(x, y) = \psi^\downarrow\{\ F\big(\ \psi^\uparrow\{I_1(x, y)\}, \psi^\uparrow\{I_2(x, y)\}, \dots, \psi^\uparrow\{I_f(x, y)\}\ \big)\ \} \qquad \textbf{3-31}$$

 The generic scheme for MD based image fusion is given as follows:

38
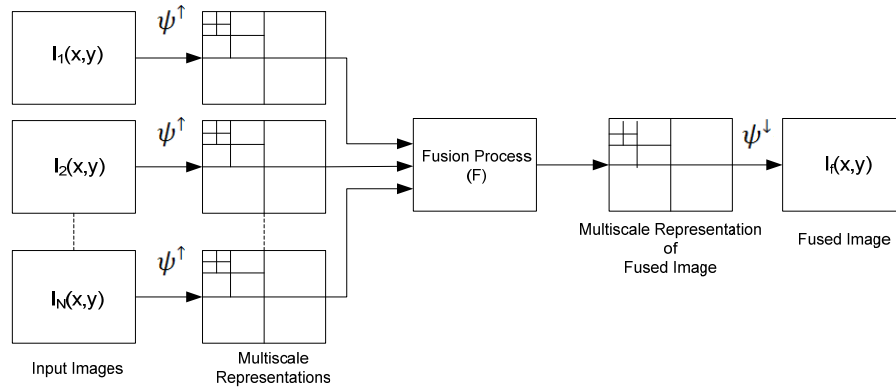
Figure 3.6 General Multiscale Decompostion Based Image Fusion Scheme

In the above figure, first of all, all the input images are decomposed into multiscale representations by a pyramid or wavelet transform. Then, the decomposed images are fused by using some fusion rules. Finally, the fused image is reconstructed by applying inverse multiscale transform.

Pyramid Transform (PT) is one of the MD methods. This method has been developed for machine vision and image compression applications. An image pyramid is a group of processed images which are arranged as the shape of a pyramid. In the pyramid, each image defines a pyramid level. The image resolution and the image size decrease while the pyramid level increases. The filters, which are used in PT, reveal different pyramid transforms. The general PT scheme is shown below:
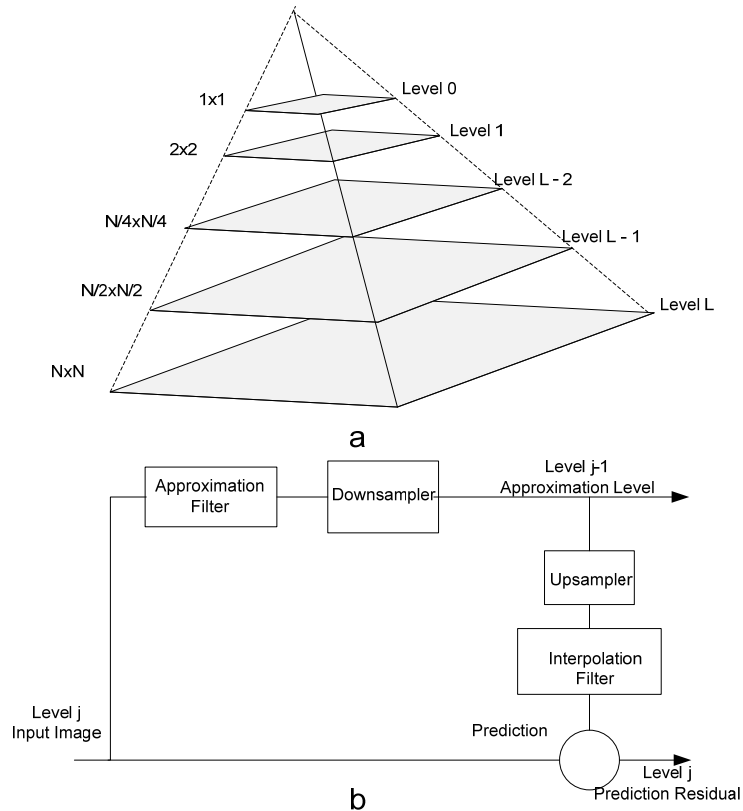
Figure 3.7 a) An Image Pyramid, b) Pyramid Decomposition Scheme

As can be seen in the figure, one level of PT is composed of upsampling, downsampling, approximation and interpolation filters. PT begins by taking the original image as an input. The first step is calculating the approximation of the original image. This is achieved by using an approximation filter and a downsampler. The approximation filtering can be a simple averaging filter or Gaussian filter. If the Gaussian filter is used for the filtering operation, the resulted pyramid is called as *Gaussian pyramid*. The approximation filter is used in order to avoid the aliasing problem of the downsampler. The approximation filter is applied prior to the downsampler in order to limit the signal before the sampling operation. If the filter is removed, aliasing may occur at the higher levels of the pyramid. After the downsampler, the image size is reduced to half in two directions, and the output is called as approximation level. By iterating the

40

same steps for each approximation level, Gaussian pyramid is constructed [34, 35].

If $W(m,n)$ represents a Gaussian filter, the Gaussian pyramid of an image can be mathematically explained as follows:

$$G_l(m,n) = |W(m,n) \otimes G_{l-1}(m,n)|_{2\downarrow} \qquad \textbf{3-32}$$

where $G_l(m,n)$ is the Gaussian pyramid at l^{th} level, and $G_L(m,n) = I(m,n)$. In the equation, $|\quad|_{2\downarrow}$ is the downsampling by 2, and $\otimes$ is the convolution operator.

At each pyramid level, prediction residual error is also computed by subtracting the approximated image from the output of the previous level. Before the subtraction operation is performed, the approximated image should be upsampled to match its size to the size of the previous level's output. Therefore, the image is upsampled by a factor of two and then is filtered by the interpolation filter. If the image is not filtered by the interpolation filter, the blocking effects may be visible at upper levels. The output of this step is used to generate *Laplacian pyramid* [34, 35].

If $R(m,n)$ represents an interpolation filter, the Laplacian pyramid of an image can be mathematically explained as follows:

$$L_l(m,n) = R(m,n) \otimes |G_{l-1}(m,n)|_{2\uparrow} \qquad \textbf{3-33}$$

where $L_l(m,n)$ is the Laplacian pyramid at l^{th} level, and $|\quad|_{2\uparrow}$ is the upsampling by 2.

Figure 3.8 a) Gaussian Pyramid of an Image b) Laplacian Pyramid of an Image

Gradient pyramid is another important pyramid transform. It is widely used in image fusion algorithms and gives good results. Gradient pyramid uses four basis functions. The advantage of using four basis functions is that it can represent the details of the images more than the other pyramid transforms. The disadvantage of this method is its high computational cost. To generate the Gradient pyramid of an image, one must calculate the Gaussian pyramid of the image as a first step. Then, the Gradient pyramid of the image is calculated as follows [37]:

$$D_{l,k}(m,n) = d_k(m,n) \otimes [\, G_l(m,n) + w(\dot{m},n) * G_l(m,n)\,] \qquad \textbf{3-34}$$

where $D_{l,k}(m,n)$ is the Gradient pyramid of the image at l^{th} level in k^{th} direction and

$$\dot{w}(m,n) = \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$d_1(m,n) = [1-1], \quad d_2(m,n) = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\frac{1}{\sqrt{2}} \qquad \textbf{3-35}$$

$$d_3(m,n) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, \quad d_4(m,n) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\frac{1}{\sqrt{2}}$$

There are several pyramid transforms in the literature other than the explained ones. *Ratio pyramid* is one of them. This pyramid transform is obtained by taking the ratio of two consecutive Gaussian pyramid levels at each pyramid level. Another pyramid transform is *Filter-Subtract-decimate* pyramid, which is the computationally efficient version of the Gaussian pyramid. *Contrast* and *Morphological* pyramids are some of the other pyramid transforms in the literature [36].

Another multiscale decomposition method is <u>*Discrete Wavelet Transform*</u> (DWT). In image fusion, DWT is used to extract the salient information in images. DWT decomposes images into sub-images which represent different frequency bands. As it has been explained before, PT also decomposes images into sub-images. However, the output of DWT gives information about spatial orientations such as edge orientations. Moreover, the total number of pixels in DWT is equal to the number of pixels in the original image. Therefore, it is not overcomplete unlike PT. The general DWT scheme is given below:
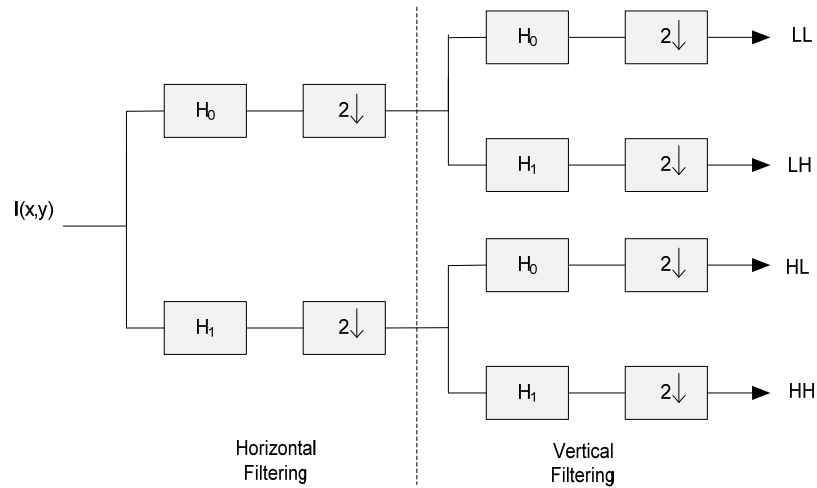
Figure 3.9 Wavelet Decomposition of an Input Image I(x,y)

As can be seen in the figure, same filters H_0 and H_1 are used for the horizontal and vertical filtering operations. The sub-bands are named as *LL* (low-low), *LH* (low-high), *HL* (high-low) and *HH* (high-high). Each sub-band corresponds to different spatial frequency. *LL* sub-band is called as approximation sub-band, and other three sub-bands are called as detail sub-bands. The approximation sub-band contains the coarse details of the original image. The other sub-bands contain the details about vertical, horizontal and diagonal spatial orientations. The second DWT level is generated by continuing the decomposition process on the approximation sub-band of the first level. DWT of an image is obtained by iterating the same operations for each level. An example of DWT is given in the below figure. As can be seen in the figure, each detail sub-band represents different edge orientations [35].

a
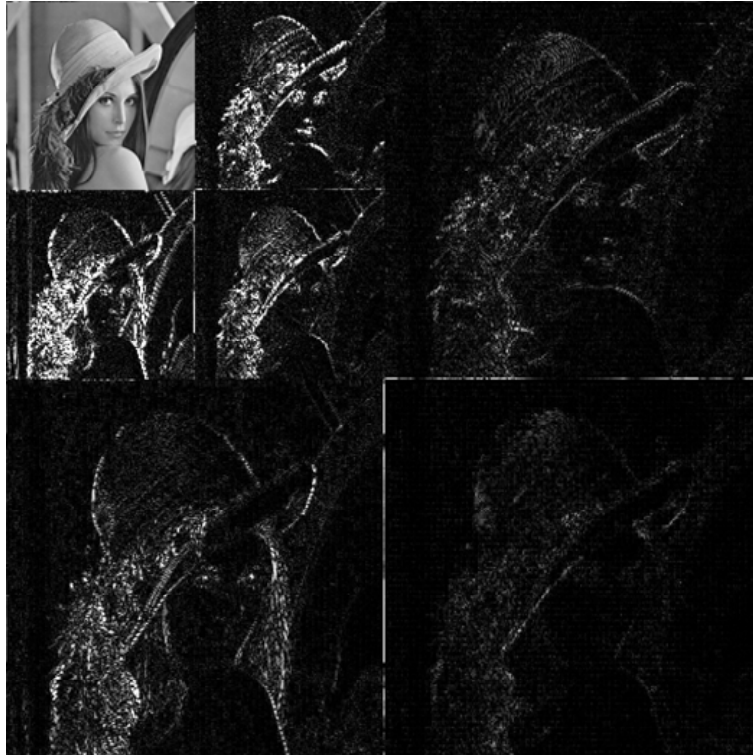
b

Figure 3.10 a) Two Level 2D Wavelet Transform of the Image b) The Names of Wavelet Sub-bands

In the following subsections, the examples of the MD based EF algorithms are explained.

3.2.3.1   EF-ALG8 [37]

This algorithm introduces a new fusion method which uses Gradient pyramid for image fusion [37]. The proposed algorithm creates a general fusion scheme that can be applicable to many applications including EF. The algorithm proposes two new measures for image fusion. These measures are called as match and salience and are calculated within a neighborhood. This approach provides greater shift invariance and noise immunity in the fused image. The methodology of the algorithm is explained as follows [37]:

The algorithm accepts multiple input images, but the number of the input images to be fused is limited to two for the ease of the explanation. First of all, the Gaussian pyramids of the input images are calculated, and then the Gradient pyramids of the input images are calculated from the calculated Gaussian pyramids.

The second step is performing the fusion process by using the measures. As it has been indicated previously, the algorithm uses two measures in the fusion process. These are *match* and *salience* measures.

Salience measure shows that a pattern in an input image contains any important information. If the salience measure is high, it means that this pattern represents important details in the image. Low salience measure indicates unwanted or corrupted data. This measure is defined as the local energy or the local variance of the pixels. Let I_1 and I_2 are the input images whose Gradient pyramids are $D_{1,k,l}$ and $D_{2,k,l}$. Then, the calculation of the salience measure is given as follows:

$$S_{I_i,l,k}(m,n) = \sum_{u,v \, \epsilon \, \mathcal{N}} D_{I_i,l,k}(m+u, n+v)^2 \qquad \text{3-36}$$

where $S_{I_i,l,k}(m,n)$ is the salience measure, and $D_{I_i,l,k}(m,n)$ is the Gradient pyramid of ith input image at l^{th} level in k^{th} orientation. Additionally, $\mathcal{N}$ is the neighborhood of the sample point at *(m,n)*.

Match measure is used to select combination mode that is going to be used for image fusion. These modes are called as *selection* and *averaging*. At pixel locations where the similarity of the input images is low, the selection mode is used for the fusion. At these locations, input image with the largest salience measure is selected. The pixel value is directly copied to the fused image. This operation keeps dominant details while rejecting unwanted or less salient features. At pixel locations where the similarity is high, the averaging mode is used for image fusion. At these locations, the average of the input images is used to increase stability and reduce noise in the fused image. The match measure is calculated as follows:

$$M_{l,k}^{12}(m,n) = \frac{2 * \sum_{u,v \in \mathcal{N}} D_{I_1,l,k}(m+u,n+v) * D_{I_2,l,k}(m+u,n+v)}{S_{I_1,l,k}(m,n) + S_{I_2,l,k}(m,n)} \qquad \textbf{3-37}$$

where $M_{l,k}^{12}(m,n)$ is the match measure. $M_{l,k}^{12}(m,n)$ takes value 1 when the input images have identical patterns and takes value -1 when the images have identical patterns with different signs.

The fusion process with the match and salience measures is defined mathematically as follows:

$$D_{I_f,l,k}(m,n) = w_1(m,n) * D_{I_1,l,k}(m,n) + w_2(m,n) * D_{I_2,l,k}(m,n) \qquad \textbf{3-38}$$

If $M_{l,k}^{12}(m,n) < \alpha$, the selection mode is used for image fusion. Then, the weights are;

$$\begin{aligned} w_{min}(m,n) &= 0 \\ w_{max}(m,n) &= 1 \end{aligned} \qquad \textbf{3-39}$$

If $M_{l,k}^{12}(m,n) > \alpha$, the averaging mode is used for the fusion. Then, the weights are;

$$w_{min}(m,n) = \frac{1}{2} - \frac{1}{2}\left(\frac{1 - M_{l,k}^{12}(m,n)}{1 - \alpha}\right)$$
$$w_{max}(m,n) = 1 - w_{min}(m,n)$$

3-40

Then, the weights of the input images are assigned as follows:

$$w_1(m,n) = \begin{cases} w_{max}(m,n) & if\ S_{I_1,l,k}(m,n) > S_{I_2,l,k}(m,n) \\ w_{min}(m,n) & otherwise \end{cases}$$

3-41

$$w_2(m,n) = \begin{cases} w_{min}(m,n) & if\ S_{I_1,l,k}(m,n) > S_{I_2,l,k}(m,n) \\ w_{max}(m,n) & otherwise \end{cases}$$

where $D_{I_f,l,k}(m,n)$ is the Gradient pyramid of the fused image, and α is the user defined threshold. If the image similarity is low, the selection mode is used for the fusion. In this mode, image weight with the highest salience measure is set to one, and other weight is set to zero. However, if the similarity is high, the averaging mode is used for image fusion. In this mode, the weighted average of the input images is calculated. The image with the higher salience measure takes higher weight in this calculation. After the fusion rules are applied, the fused image is calculated from $D_{I_f,l,k}(m,n)$ by inverse PT.

3.2.3.2   EF-ALG9 [38]

This algorithm is another example of MD based image fusion methods [38]. It is computational efficient and easy to implement thanks to the Laplacian pyramid transform. The steps of the algorithm are explained as follows:

The first step of the algorithm is calculating the Laplacian pyramids of the input images. The algorithm does not use any measures in the fusion process and

accepts multiple input images. Let there are N input images namely $I_1, I_2,..., I_N$, and $L_{I_1,l}, L_{I_2,l}, ..., L_{I_N,l}$ are the Laplacian pyramids of the input images at l^{th} level.

The second step is performing fusion process at each pyramid level. The fusion process is composed of two different rules. For the 0th level and the remaining levels, different fusion rules are applied. For the 0th level, averaging function is used for the fusion and is given as follows:

$$L_{I_f,0}(m,n) = \frac{1}{N}\sum_{i=1}^{N} L_{I_i,0}(m,n) \qquad \textbf{3-42}$$

where $L_{f,0}$ is the 0th level of the fused image.

For the remaining levels, simple absolute maximum function is used to extract details in the input images. The fusion operation is given as follows:

$$L_{I_f,l}(m,n) = maximum(|L_{I_i,l}(m,n)|) \ for \ l \neq 0 \qquad \textbf{3-43}$$

where $L_{I_f,l}(m,n)$ is l^{th} level of the fused image. After the fusion rules are applied, the fused image is calculated from $L_{I_f,l}(m,n)$ by using inverse PT.

3.2.3.3   EF-ALG10 [39]

This algorithm is an extension to existing pyramid based fusion methods [39]. The extension provides fusion method to perform well even though the input images have significantly large brightness variations. Furthermore, the contribution of each input image to the fused image is controlled by a mask. The mask is calculated automatically from the pixel intensities. This operation eliminates the need of input parameters. The methodology of the algorithm can be explained as follows:

The algorithm uses a spline function to fuse two input images by the help of a mask. In the algorithm, fusion process is extended by taking this spline fusion as a basis. The proposed spline function is given as follows [39]:

$$I_S(x, y) = spline2\left(I_i(x, y), I_j(x, y), R(x, y)\right)$$

{

$$L_{S,l}(m, n) = G_{R,l}(m, n)L_{I_i,l}(m, n) + \left(1 - G_{R,l}(m, n)\right)L_{I_j,l}(m, n) \qquad \textbf{3-44}$$

$$I_S(x, y) = Reconstruction\ of\ Laplacian\ Pyramid\{L_{I_S,l}(m, n)\}$$

}

where $L_{I_S,l}(m, n)$, $L_{I_i,l}(m, n)$ and $L_{I_j,l}(m, n)$ are the Laplacian pyramids of the spline function output, I_i and I_j respectively. $G_{R,l}(m, n)$ is the Gaussian pyramid of the mask R at level l^{th}.

The algorithm assumes that the input images are put in order by boosting brightness. This constraint is important because wrong ordering causes unsatisfactory results. Firstly, the input images are fed to the algorithm in required order, and then input image with the smallest number of saturated pixels is found. At the last step of the algorithm, the found image is used to finish the fusion process. This image is denoted by an index C. Then, the remaining steps of the algorithm are given as follows:

The below equations are iterated for $i = 2, 3, ...,N$ and $\hat{I}_1 = I_1$. u and o show predefined exponents, which are one and two respectively. At each iteration, the mask R is normalized before applying it to the *spline2* function. Then, the operation is defined as follows:

$$R(m, n) = \hat{I}_{i-1}^u(x, y) + I_i^o(x, y)$$

$$\hat{I}_i(x, y) = spline2\left(\hat{I}_{i-1}(x, y), I_i(x, y), R(x, y)\right) \qquad \textbf{3-45}$$

50

The below equations are repeated for $i = N - 1, N-2, ..., 1$. Then, the operation is defined as follows:

$$R(x, y) = \hat{I}^u_{i-1}(x, y) + \hat{I}^o_i(x, y)$$
$$\hat{I}_i(x, y) = spline2\big(\hat{I}_{i-1}(x, y), I_i(x, y), (1 - R(x, y))\big)$$

<div align="right">3-46</div>

Finally, the fusion is completed by the following step:

$$R(x, y) = \hat{I}^u_1(x, y) + \hat{I}^o_C(x, y)$$
$$I_f(x, y) = spline2\big(\hat{I}_1(x, y), \hat{I}_C(x, y), R(x, y)\big)$$

<div align="right">3-47</div>

3.2.3.4   EF-ALG11 [40]

This algorithm uses Laplacian and Gaussian pyramids with three quality measures [40]. It gives extension to former methods in terms of giving a specific solution for EF. Their quality measures are directly related to EF. The methodology can be explained symbolically as follows:

As it has been mentioned previously, the method uses quality measures in image fusion. The measures are defined as follows [40]:

- The first measure is *contrast*, and it is used to define the edge and salient details in the input images. The measure is calculated by Laplacian filter and is denoted by C.

- The second measure is *saturation,* which is denoted by S. The measure helps the fused images to be more vivid and is defined as the standard deviation of the input images.

- The third measure is *well-exposedness,* which is denoted by *E*. This measure shows that how well a pixel is exposed and is calculated by Gaussian filter.

Then, for each input image, a weight map is calculated by using these quality measures. The calculation of the weight maps is given as follows:

$$W_i(x,y) = \left(C_i(x,y)\right)^{w_C} * \left(S_i(x,y)\right)^{w_S} * \left(E_i(x,y)\right)^{w_E}$$
$$\widehat{W}_i(x,y) = \frac{W_i(x,y)}{\sum_{i=1}^{N} W_i(x,y)}$$

<div align="right">3-48</div>

where $\widehat{W}_i(x,y)$ is the normalized weight map of ith image. The exponents, which are denoted by w_C, w_S and w_E, are used to control the contribution of each quality metric to the weight map.

The next step is calculating the pyramid transforms of the input images. The proposed algorithm performs blending operation in PT. It is indicated that when the blending operation is performed in the spatial domain, the weight map may contain sharp transitions due to different intensity values from the input images. This problem has been already solved by applying a Gaussian filter to the weight map prior to the blending operation. However, this filtering causes artifacts in the fused image. Therefore, the blending operation is performed in PT. The fusion operation is defined as follows:

$$L_{I_f,l}(m,n) = \sum_{i=1}^{N} G_{\widehat{W}_i,l}(m,n) L_{I_i,l}(m,n)$$

<div align="right">3-49</div>

where $L_{I_i,l}(m,n)$ is the Laplacian pyramid of input image I_i at level l^{th}, and $G_{\widehat{W}_i,l}(m,n)$ is the Gaussian pyramid of the normalized weight map $\widehat{W}$ at level l^{th}. MD based blending operation provides consistent results because MD based

methods blend features rather than pixel intensities. Therefore, the resultant image does not have any blending artifacts or sharp transitions. Finally, the fused image is calculated by inverse PT.

3.2.3.5   EF-ALG12 [41]

This algorithm is a modified version of EF-ALG11 [41]. The modification is the replacement of PT with DWT. Additionally, the measures, which are proposed by the former algorithm, are used directly without any modifications. Since DWT cannot be applied on color images without a preprocessing operation, the color domain is changed from RGB to L*a*b. Then, DWT is applied to the L component of L*a*b. For monochrome images, DWT can be directly applied to the image. The steps of the algorithm are explained as follows:

First of all, the quality measures, which are contrast, saturation and well-exposedness [40], are used to form the weight maps. The next step is calculating the DWT of the weight maps for L levels. The calculated detail and approximation sub-bands are given as follows:

$$LH_{W_{i,l}}(m,n), HL_{W_{i,l}}(m,n), HH_{W_{i,l}}(m,n), LL_{W_{i}L}(m,n) \qquad \textbf{3-50}$$

where $LH_{W_{i,l}}(m,n), HL_{W_{i,l}}(m,n),$ and $LL_{W_{i,l}}(m,n)$ are the Horizontal, Vertical and Diagonal sub-band of ith weight map at l^{th} level. $LL_{W_{i,L}}(m,n)$ is the approximation sub-band of ith weight map at L^{th} level. Then, the DWT of the input images is calculated. Resultant sub-bands are given as follows:

$$LH_{I_{i,l}}(m,n), HL_{I_{i,l}}(m,n), HH_{I_{i,l}}(m,n), LL_{I_{i,L}}(m,n) \qquad \textbf{3-51}$$

where $LH_{I_{i,l}}(m,n), HL_{I_{i,l}}(m,n)$ and $HH_{I_{i,l}}(m,n)$ are the Horizontal, Vertical and Diagonal detail sub-band of ith image at l^{th} level. $LL_{I_{i,L}}(m,n)$ is the approximation sub-band of ith image at L^{th} level.

53

After the calculation of the sub-bands, the fusion process is performed. The process is composed of two different rules. For the approximation and detail sub-bands, different fusion rules are applied. Before the fusion process, the approximation sub-bands of the weight maps are preprocessed to satisfy the following constraint:

$$sum(m,n) = \sum_{i=1}^{N} LL_{W_i,L}(m,n)$$

$$sum(m,n) = 1$$

3-52

where $sum(m,n)$ is the sum of all approximation sub-bands of the weight maps. This constraint forces the weight maps to be consistent. After that step, the approximation sub-bands of the weight maps are normalized as follows:

$$LL_{\widehat{W}_i,L}(m,n) = \frac{LL_{W_i,L}(m,n)}{2^L}$$

3-53

where $LL_{\widehat{W}_i,L}(m,n)$ is the normalized approximation sub-band of ith weight map. Then, for the approximation sub-bands, the fusion process is given as follows:

$$LL_{I_f,L}(m,n) = \sum_{i=1}^{N} LL_{I_i,L}(m,n) * LL_{\widehat{W}_i,L}(m,n)$$

3-54

where $LL_{I_f,L}(m,n)$ is the approximation sub-band of the fused image.

The next step is the fusion of the detail sub-bands. The detail sub-bands of the fused image are calculated by summing each horizontal, vertical and diagonal sub-band of the input images. The summation of the detail sub-bands transfers the details in the input images to the fused image. In the paper, this operation is given as follows:

$$LH_{I_f,l}(m,n) = \sum_{i=1}^{N} \left(\frac{LH_{I_i,l}(m,n)}{threshold} \right)$$

$$HL_{I_f,l}(m,n) = \sum_{i=1}^{N} \left(\frac{HL_{I_i,l}(m,n)}{threshold} \right) \qquad \text{3-55}$$

$$HH_{I_f,l}(m,n) = \sum_{i=1}^{N} \left(\frac{HH_{I_i,l}(m,n)}{threshold} \right)$$

where $LH_{I_f,l}(m,n), HL_{I_f,l}(m,n)$ and $HH_{I_f,l}(m,n)$ are the Horizontal, Vertical and Diagonal detail sub-band of the fused image at l^{th} level. There is a threshold parameter in the above equations. This threshold parameter controls the contribution of each detail sub-band to the fused image. The edge intensity in the fused image can be adjusted by changing the threshold value. The last step of the algorithm is calculating the fused image by inverse DWT.

CHAPTER 4

QUALITY METRICS

4.1   Introduction

In years, many image fusion methods have been developed so far. These methods have some advantages and disadvantages over other image fusion methods. However, an important problem arises when it is asked to compare or evaluate the performance of the methods. Image quality assessment techniques are proposed to solve this problem. Image quality assessment is an important but a difficult issue in image fusion.

The subjective evaluation of image fusion's results is one of the solutions in order to evaluate the performance of the methods. In this approach, image quality assessment is carried out by humans [48 - 52]. Subjective tests are developed and prepared for this purpose. However, the repeatability of these tests is very low, and their time cost is very high. Therefore, researchers have tried to find quality assessment techniques which are reliable and do not require any human interactions.

Then, objective quality metrics are proposed to solve this problem. Firstly, reference-based objective quality metrics are developed. These metrics compare the results of image fusion methods by using an ideal reference image. Some of the widely used metrics are root mean square error (RMSE), mean squarer error (MSE), peak signal to noise ratio (PSNR) and correlation (CORR). However, this

approach is not realistic, because there is no such a reference image in image fusion. Due to lack of an ideal reference image in image fusion, researchers have tried to find a novel benchmarking method that does not depend on any reference images. Several metrics have been developed in the last decades in order to solve this problem. Then, non-reference objective quality metrics are developed. Some of them are standard deviation (STD), entropy (E), cross entropy (CE), mutual information (MI), universal quality index (UIQI), C.S. Xydeas and V. Petrovic and spatial frequency (SF) [51-56]. These metrics measure the similarity between input and fused images or amount of transferred salient information to fused image. In another sense, the metrics try to measure the performance of image fusion methods without any reference images.

4.2 Objective Quality Metrics

In the following sections, the well-known non-reference objective quality metrics are explained.

4.2.1 Standard Deviation

Standard Deviation (SD) is one of the objective quality metrics. SD only considers the histogram of the fused image and evaluates the fused image with respect to the width of its histogram. In other words, it only evaluates the details in the fused image. In image fusion, a larger SD means better image fusion. The formulation of SD is given as follows [51]:

$$SD = \sqrt{\sum_{z=0}^{L} (z - \bar{z})^2 h_{I_f}(z)}$$

$$\bar{z} = \sum_{z=0}^{L} z\, h_{I_f}(z)$$

4-1

57

where $h_{I_f}(z)$ is the normalized histogram of the fused image $I_f(x,y)$, and L is the number of bins in the histogram.

4.2.2   Entropy

In communication, describing information content of a signal is very important, and several researches have been carried out to develop a method for this purpose. In 1948, Claude E. Shannon introduced a new concept to solve this problem. In this concept, entropy (E) is defined as the measure of the overall information content of the given signal or data [52]. Entropy uses the probability density function of the signal in order to measure its randomness. High entropy denotes high randomness, and it shows that the information content of the signal is rich. In image fusion, entropy is used to measure the salient features in the fused image. The larger entropy means that the fused image contains more information and implies better image fusion. The formulation of E is given as follows [51]:

$$E = -\sum_{z=0}^{L} h_{I_f}(z)log_2\, h_{I_f}(z) \qquad \text{4-2}$$

4.2.3   Cross Entropy

Cross Entropy (CE) is another objective quality metric to evaluate the outputs of image fusion methods. In image fusion, CE is used to measure the dissimilarity between the input images and the fused image. The dissimilarity measure is used to describe how much information is transferred from the input images to the fused image. The smaller the cross entropy, better image fusion results are obtained. The calculation of CE is given as follows [54];

$$CE\left(I_1, I_2, ..I_N; I_f\right) = \frac{CE\left(I_1; I_f\right) + CE\left(I_2; I_f\right) + \cdots + CE\left(I_N; I_f\right)}{2} \qquad \text{4-3}$$

$$CE\ (I_i; I_f) = \sum_{z=0}^{L} h_{I_i}(z) log_2 \frac{h_{I_i}(z)}{h_{I_f}(z)} \qquad \text{4-4}$$

where $h_{I_i}(z)$ is the normalized histogram of the image $I_i(x,y)$.

4.2.4   Mutual Information

Mutual Information (MI) is an important concept in information theory. MI denotes the dependence between two random variables [57]. In image fusion, this concept is adapted to measure the amount of information transferred from the input images to the fused image. The higher MI means better image fusion. The formulation of MI is given as follows [57]:

$$MI\ = MI_{I_1 I_f} + MI_{I_2 I_f} + \cdots + MI_{I_N I_f} \qquad \text{4-5}$$

$$MI_{I_i I_f} = \sum_{x=1}^{X} \sum_{y=1}^{Y} h_{I_i I_f}(x,y) log_2 \left(\frac{h_{I_i I_f}(x,y)}{h_{I_i}(x,y) h_{I_f}(x,y)} \right) \qquad \text{4-6}$$

where $h_{I_i I_f}(z)$ is the joint histogram of the fused image $I_f(x,y)$ and $I_i(x,y)$.

4.2.5   Universal Image Quality Index

Universal Image Quality Index (UIQI) is another well-known objective quality metric. It was proposed by Zhou Wang and Alan C. Bovik in 2002 [53]. The proposed metric does not require any reference images. For N input images, theoretically explanation of UIQI is given as follows:

$$Q = \frac{Q^{I_1 I_f} + Q^{I_2 I_f} + \cdots + Q^{I_N I_f}}{N} \qquad \text{4-7}$$

Let's assume that there are two images that are the input and fused images.

$$Q = \frac{4\sigma_{I_i I_f} \bar{I}_i \bar{I}_f}{\left(\sigma_{I_i}^2 + \sigma_{I_f}^2\right)\left[(\bar{I}_i)^2 + (\bar{I}_f)^2\right]}$$

4-8

where

$$\bar{I}_i = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} I_i(x,y), \; \bar{I}_f = \frac{1}{XY}\sum_{x=1}^{X}\sum_{y=1}^{Y} I_f(x,y)$$

4-9

$$\sigma_{I_i}^2 = \frac{1}{XY-1}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_i(x,y) - \bar{I}_i)^2, \sigma_{I_f}^2 = \frac{1}{XY-1}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_f(x,y) - \bar{I}_f)^2$$

4-10

$$\sigma_{I_i I_f} = \frac{1}{XY-1}\sum_{x=1}^{X}\sum_{y=1}^{Y}(I_i(x,y) - \bar{I}_i)\,(I_f(x,y) - \bar{I}_f)$$

4-11

The dynamic range of Q is [-1,1]. Q takes the value of one, only if two images are exactly same. The representation of Q in equation 4-7 is composed of three different terms. The alternative representation is given as follows:

$$Q = \frac{\sigma_{I_i I_f}}{\sigma_{I_i}\sigma_{I_f}} * \frac{2\bar{I}_i\bar{I}_f}{(\bar{I}_i)^2 + (\bar{I}_f)^2} * \frac{2\sigma_{I_i}\sigma_{I_f}}{\sigma_{I_i}^2\sigma_{I_f}^2}$$

4-12

These terms represent the loss of correlation, luminance distortion and contrast distortion respectively.

Most of the time, images are non-stationary, so the quality metric should not be calculated for the whole image. Therefore, the metric is calculated within a

neighborhood. Then, the local metrics are combined to form the overall quality metric. The equation is given as follows:

$$Q^{I_i I_f} = \frac{1}{|W|} \sum_{w \in W} Q(I_i, I_f | w)$$

$$\text{4-13}$$

4.2.6   C.S. Xydeas & V. Petrovic Quality Measure

In 2000, C.S. Xydeas and V. Petrovic developed another objective quality metric. This metric measures the amount of edge detail transferred from the input images to the fused image. Since human visual system is more sensitive to edge information, the visual details in images are represented by edge information.

The edge information of the images is obtained by using Sobel edge operator. At each pixel location, edge strength $g(x,y)$ and edge orientation $\alpha(x,y)$ are calculated as follows:

$$g_{I_i}(x,y) = \sqrt{s_{I_i}^x(x,y)^2 + s_{I_i}^y(x,y)^2}$$

$$\text{4-14}$$

$$\alpha_{I_i}(x,y) = tan^{-1}\left(\frac{s_{I_i}^y(x,y)}{s_{I_i}^x(x,y)}\right)$$

$$\text{4-15}$$

where $s_{I_i}^x(x,y)$, $s_{I_i}^y(x,y)$ are the outputs of image $I_i(x,y)$ convolved with horizontal and vertical Sobel kernels. Then, the relative strength $G^{I_i I_f}(x,y)$ and relative orientation $A^{I_i I_f}(x,y)$ values of an input image $I_i(x,y)$ with respect to the fused image $I_f(x,y)$ are calculated as follows:

$$G^{I_iI_f}(x,y) = \begin{cases} \dfrac{g_{I_f}(x,y)}{g_{I_i}(x,y)} & \text{if } g_{I_i}(x,y) > g_{I_f}(x,y) \\ \dfrac{g_{I_i}(x,y)}{g_{I_f}(x,y)} & otherwise \end{cases} \qquad \textbf{4-16}$$

$$A^{I_1I_f}(x,y) = \frac{\left| \left| \alpha_{I_i}(x,y) - \alpha_{I_f}(x,y) \right| - \dfrac{\pi}{2} \right|}{\pi/2} \qquad \textbf{4-17}$$

By using the above equations, edge strength and edge orientation preservation values are calculated as follows:

$$Q_g^{I_iI_f}(x,y) = \frac{\Gamma_g}{1 + e^{\kappa_g(G^{I_iI_f}(x,y) - \sigma_g)}} \qquad \textbf{4-18}$$

$$Q_\alpha^{I_iI_f}(x,y) = \frac{\Gamma_\alpha}{1 + e^{\kappa_\alpha(G^{I_iI_f}(x,y) - \sigma_\alpha)}} \qquad \textbf{4-19}$$

$Q_g^{I_iI_f}(x,y)$ and $Q_\alpha^{I_iI_f}(x,y)$ represent the perceptual loss of information in the fused image. "The constants Γ_g, κ_g, σ_g and Γ_α, κ_α, σ_α determine the exact shape of the sigmoid functions used to form the edge strength and orientation preservation values." [54, 55] Then, the overall edge preservation values are calculated by using the above equations as follows:

$$Q^{I_iI_f}(x,y) = Q_g^{I_iI_f}(x,y)Q_\alpha^{I_iI_f}(x,y) \qquad \textbf{4-20}$$

The dynamic range of $Q^{I_iI_f}(x,y)$ is [0-1]. If $Q^{I_iI_f}(x,y)$ is zero, it means that none of the edge information is transferred from the input image $I_i(x,y)$ to the fused image $I_f(x,y)$. In other words, edge information totally is lost during image fusion. When $Q^{I_if}(x,y)$ equals to one, all information exist in the input image is

transferred to the fused image without any losses. At the final step, the normalized image quality metric for N input images is defined as follows:

$$Q^{(I_1 \dots I_N | I_f)} = \frac{\sum_{x=1}^{X} \sum_{y=1}^{Y} Q^{I_1 I_f}(x,y) w^{I_1}(x,y) + \cdots + Q^{I_N I_f}(x,y) w^{I_N}(x,y)}{\sum_{x=1}^{X} \sum_{y=1}^{Y} (w^{I_N}(x,y) + \cdots + w^{I_N}(x,y))} \qquad \textbf{4-21}$$

The quality metric contains additional weight functions namely $w^{I_i}(x,y)$. These weights stress the perceptual importance of the edge elements in the input images.

They are defined as follows:

$$w^{I_i}(x,x) = \left(g_{I_i}(x,y)\right)^{L} \qquad \textbf{4-22}$$

where L is a constant. Most of the time, this constant is set to one. Moreover, the dynamic range of $Q^{(I_1 \dots I_N | I_f)}$ is [0-1]. The value of one means the input and fused images are totally similar with respect to the edge information.

4.2.7 Spatial Frequency

Spatial Frequency (SF) is another objective quality metric which is convenient for human visual system. SF denotes overall activity level in an image using spatial frequency [56]. The formulation of SF is given as follows;

$$SF = \sqrt{(RF)^2 + (CF)^2} \qquad \textbf{4-23}$$

$$RF = \sqrt{\frac{1}{X * Y} \sum_{x=1}^{X} \sum_{y=2}^{Y} [I_f(x,y) - I_f(x,y-1)]^2} \qquad \textbf{4-24}$$

$$CF = \sqrt{\frac{1}{X*Y} \sum_{x=2}^{X} \sum_{y=1}^{Y} [I_f(x,y) - I_f(x-1,y)]^2}$$ **4-25**

where *RF* and *CF* is the row and column frequencies of the image.

4.2.8 Saturated Pixel Percentage

Saturated pixel percentage is the ratio of the number of saturated pixels to the number of pixels in the fused image. This metric has not been used in the literature as a performance measure. However, image dynamic range enhancement methods try to reduce saturated regions and increase the contrast of the fused image. Therefore, the percentage of undersaturated and oversaturated pixels in the fused image can be used to evaluate the performance of image fusion methods. This metric does not show how much information is transferred from the input images to the fused image or the similarity between the input images and the fused image. This metric only shows the percentage of the saturated pixels in the fused image.

CHAPTER 5

EXPERIMENTAL RESULTS

5.1   Introduction

Until now, the implemented algorithms have been explained by giving the advantages and disadvantages of the algorithms with respect to the authors of the algorithms. However, most of these studies do not include a detailed comparison between proposed algorithms and former algorithms. Therefore, in this chapter, the performance evaluation of the implemented algorithms is examined, and the algorithms are compared with each other through two different tests. In this thesis, algorithms related to EF and HDRI are implemented and explained. However, this thesis focuses on EF algorithms and does not provide much information about HDRI algorithms. Therefore, the results related to EF algorithms are given in this chapter. This chapter is composed of two different sections: 1) the objective evaluation of the algorithms 2) the subjective evaluation of the algorithms.

In the first section, the algorithms are compared with each other by using the objective quality metrics, which are explained in Chapter 4. The goals of this section are as follows: 1) giving the detailed comparison of the EF algorithms 2) determining the most appropriate objective quality metric for the evaluation of EF algorithms. This section consists of two subsections. The first subsection includes the comparison of the EF algorithms. The second subsection includes a brief comparison of the HDRI algorithms through Xydeas & Petrovic quality metric.

This section only gives general information about the performance of the algorithms and shows necessity of the successful HDRI algorithms.

In the second section, the algorithms are compared through a subjective test. The goals of this section are as follows: 1) evaluating the EF algorithms by human subjective scores 2) showing the correlation between the objective quality metrics and the human subjective ratings 3) finding an objective quality metric that is the most correlated with human visual system. For those purposes, a subjective test is performed, and the results related to the subjective test are given in this section.

The algorithms and objective quality metrics are all implemented in MATLAB, and the results given in this chapter are obtained in MATLAB.

In the experiments, the implemented algorithms are named as follows:

Table 5.1 Acronym of the Algorithms

Acronyms	Descriptions of the Algorithms
EF-ALG1	Simple Average [28]
EF-ALG2	Weighted Average [28]
EF-ALG3	"A Real Time Algorithm for Exposure Fusion of Digital Images", *Tomislav Kartalov, Aleksandar Petrov, Zoran Ivanovski, Ljupcho Panovski* [29]
EF-ALG4	"Fusion of Multi-Exposure Images", *A. Ardeshir Goshtasby* [30]
EF-ALG5	"Real-Time Exposure Fusion on a Mobile Computer", *Asheer Kasar Bachoo* [31]
EF-ALG6-1	"Contrast Based Image Fusion Technique in the DCT Domain", *Jinshan Tang* (**The First Original Algorithm**) [32]
EF-ALG6-2	"Contrast Based Image Fusion Technique in the DCT Domain", *Jinshan Tang* (**The Second Original Algorithm**) [32]
EF-ALG6-3	"Contrast Based Image Fusion Technique in the DCT Domain", *Jinshan Tang* (**Improved Algorithm**) [32]

Table 5.1 (continued)

Acronyms	Description of Algorithms
EF-ALG7-1	"Multi-Exposure & Multi-Focus Image Fusion in Transform Domain", *L.Zafar, E.A.Edirisinghe, H.E.Bez* (***Original Algorithm***) [33]
EF-ALG7-2	"Multi-Exposure & Multi-Focus Image Fusion in Transform Domain*", L.Zafar, E.A.Edirisinghe, H.E.Bez* (***Improved Algorithm***) [33]
EF-ALG8	"Enhanced Image Capture Through Fusion", *Peter J. Burt, Raymond J. Kolczynski* [37]
EF-ALG9	"Fusion of Differently Exposed Images", *Ron Rubinstein* [38]
EF-ALG10	"High Dynamic Range Imaging Through Multi-Resolusion Spline Fusion", *Hung-Son Le, Adi Anani, Haibo Li* [39]
EF-ALG11	"Exposure Fusion", *Tom Mertens, Jan Kautz, Frank Van Reeth* [40]
EF-ALG12	"Wavelet Based Exposure Fusion", *Madiha Hussain Malik, S. Asif M. Gilani, Anwaar-ul-Haq* [41]
HDR-ALG1	"Recovering High Dynamic Range Radiance Maps from Photographs", *Paul E. Debevec, Jitendra Malik* [10]
HDR-ALG2	"Dynamic Range Improvement Through Multiple Exposures", *Mark A. Robertson, Sean Borman, Robert L. Stevenson* [11]
TM-ALG1	Linear Mapping [38]
TM-ALG2	Logarithmic Mapping [15]
TM-ALG3	"Adaptive Logarithmic Mapping for Displaying High Contrast Scenes", *F. Drago, K. Myszkowski, T. Annen, N. Chiba* [15]
TM-ALG4-1	"Photographic Tone Reproduction for Digital Images", *Erik Reinhard, Michael Stark, Peter Shirley, James Ferwerda* (***Global Operator***) [17]
TM-ALG4-2	"Photographic Tone Reproduction for Digital Images", *Erik Reinhard, Michael Stark, Peter Shirley, James Ferwerda* (***Local Operator***) [17]
TM-ALG5	"An Effective Tone Mapping Operator for High Dynamic Range Images", *B. Kh. Barladian, A. G. Voloboi, V. A. Galaktionov, E. A. Kopylov* [13]

As stated previously, this thesis only considers monochrome image based algorithms. Therefore, if an algorithm is color based, it has been modified to work with the monochrome images. Moreover, some of the implemented algorithms require several input parameters. The performed modifications in the algorithms and the input parameters used in this chapter are given as follows:

As stated previously, *EF-ALG2* uses weighted average of the input images for image fusion. In this algorithm, the weights are chosen as 0.11, 0.79, 0.11 so the images that are underexposed and overexposed has been multiplied with 0.11.

As stated previously, *EF-ALG4* and *EF-ALG5* require the block size as an input parameter. In the experiments, the block size has been chosen as %15 of the minimum dimension of the input image. Moreover, the sigma parameter of the blending function in EF-ALG4 has been chosen as two times of the chosen block size.

In DCT based algorithms, which are *EF-ALG6-X* and *EF-ALG7-X*, 8x8 DCT has been used for image fusion.

In *EF-ALG8*, the fusion process is explained for two input images, and there is no information about the fusion process for multiple input images. Therefore, the algorithm is modified to cover the case for three input images. The fusion process is performed in two steps. In the first step, the first two input images are fused as explained in the algorithm. Then, the fusion result of this step is fused with the third input image. As a result, three input images are fused by EF-ALG8. Furthermore, the input parameter α in this algorithm has been selected as -0.5 to obtain satisfactory results.

In *EF-ALG11*, the exponents, which control the contribution of each quality measure to the weight map, have been set to 1 in the experiments.

In *EF-ALG12*, the threshold value, which controls the edge intensity in the fused image, has been set to 2.2 in the experiments.

Moreover, pyramid and wavelet transforms have been generated depend on the sizes of the input images. The pyramid and wavelet representations have been generated for the possible maximum level for the given input images.

In *HDR-ALGX*, smoothness term has been chosen as 40.

In *TM-ALG3* and *TM-ALG4-X*, the key values have been set to 1 and 0.48 respectively. Moreover, the parameters ϕ and the threshold have been set to 8 and 0.05 respectively.

All other parameters of the algorithms have been set to the values defined in the algorithms.

5.2   Objective Evaluation of the Algorithms

In this section, the implemented algorithms are compared with each other by the help of the objective quality metrics. Until now, the algorithms have been explained individually without giving any results. By performing an objective evaluation test, we aim to compare the algorithms and see their weakness and powerfulness with respect to the other algorithms. Moreover, the result of this evaluation also gives an idea about the performance of the objective quality metrics in the evaluation of the EF algorithms. Then, the most appropriate objective quality metric to evaluate EF algorithms is determined by considering the results of this test.

In this test, ten different image sets are used. The images are selected to represent all scenarios that can be come across in the real world. The each image set consists of three differently exposed images. These images represent

underexposed, normal and overexposed images. The number of the input images is limited to three because this number is enough to represent most of the high dynamic range scenes. The number of images greater than three may increase the performance of the algorithms. However, time, memory and processor requirements increase significantly depending on the number of input images. Three is the optimum number for this study and gives satisfactory results. The image sets, mostly obtained through the internet, can be found in Appendix A. The resolutions and exposure settings of the images are given at the bottom of the image sets. The only exposure setting of the images changes in each image set. All other settings of the images are same in each image set.

In the following subsections, the comparisons of the algorithms through the objective quality metrics are given.

5.2.1 Experimental Results for the Exposure Fusion Algorithms

In this subsection, the results are given as follows:

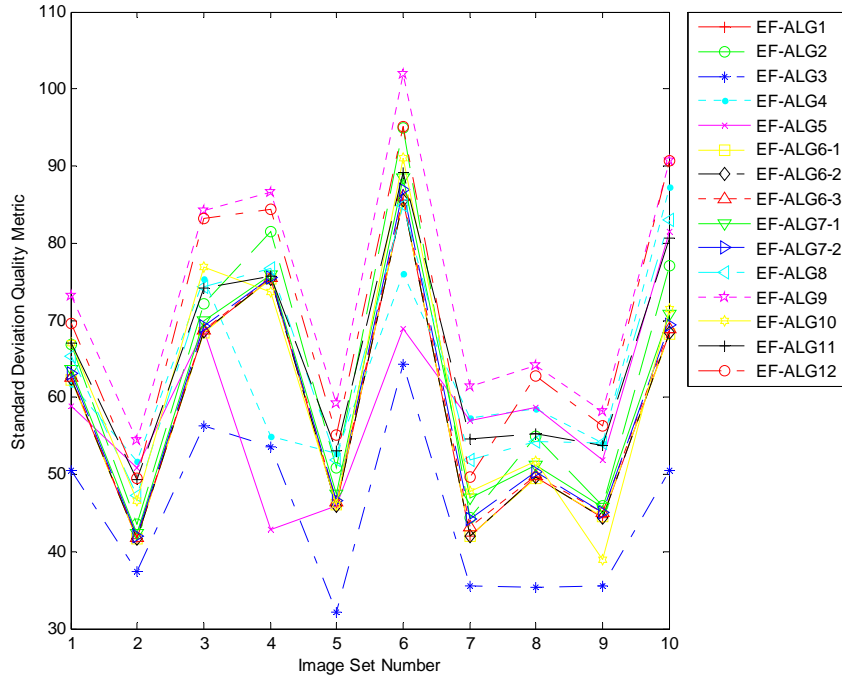## 5.2.1.1   Standard Deviation Quality Metric

Figure 5.1 Comparison of the EF Algs. by Using SD

In standard deviation quality metric, it is observed that EF-ALG3 is rated lower than the other algorithms in most of the image sets. EF-ALG3 tries to fuse the input images by the help of three different mapping functions. Since these functions are not good at preserving the contrast of the input images, the fused image has low contrast and limited histogram. That is why EF-ALG3 performs the worst in this quality metric.

The above figure shows that EF-ALG9 performs the best in this quality metric. Most of the time, it gives high contrast fused images. However, most of the saturated pixels in the input images are not recovered. EF-ALG9 uses absolute maximum function for the fusion of Laplacian pyramid's levels. This nonlinear process does not guarantee that the fused image is in the acceptable range. Therefore, some pixels in the fused image may have intensity values that are out

of the acceptable range. Consequently, the fused image has many saturated pixels and high peaks in the histogram.

As a result, standard deviation is not reliable for comparing different EF algorithms because it gives high points to the fused images that have many saturated pixels. This approach causes inaccurate results. However, it can give information about the global contrast of the fused image.

5.2.1.2   Entropy Quality Metric
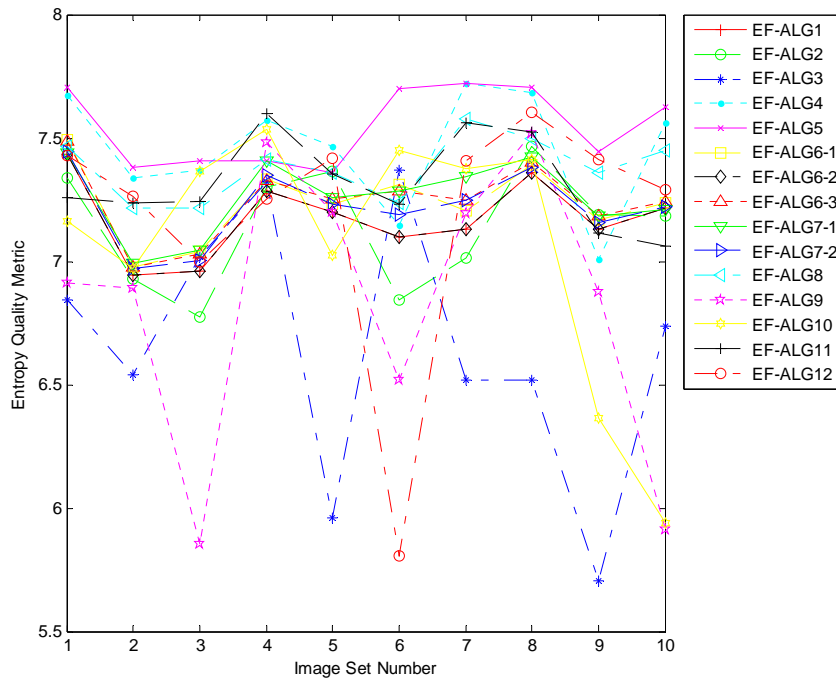
Figure 5.2 Comparison of the EF Algs. by Using Entropy

In the entropy quality metric, EF-ALG5 and EF-ALG4 have the highest scores in most of the image sets. The reason is that these algorithms give good contrast images while recovering the details in saturated regions. Therefore, the output of these algorithms has less saturated pixels and wide histogram. Moreover, these algorithms use the entropy to define the details in the input images. That is why they are rated higher than the others.

In this quality metric, EF-ALG3 and EF-ALG9 perform worse than the others. EF-ALG3 usually gives poor contrast images because of its fusion rules. Furthermore, the output of EF-ALG9 has good contrast, but it has many saturated pixels due to the previously stated reason. The saturated pixels create peaks in the histogram, and these peaks cause randomness to decrease. Consequently, EF-ALG3 and EF-ALG-9 are rated lower than the others.

As a summary, entropy is a measure of randomness and can be used to measure the details in the fused image. It provides information about the contrast of the fused image and the number of saturated pixels. Since these parameters are important for EF algorithms, EF algorithms can be evaluated by means of entropy objective quality metric.

5.2.1.3   Cross Entropy Quality Metric

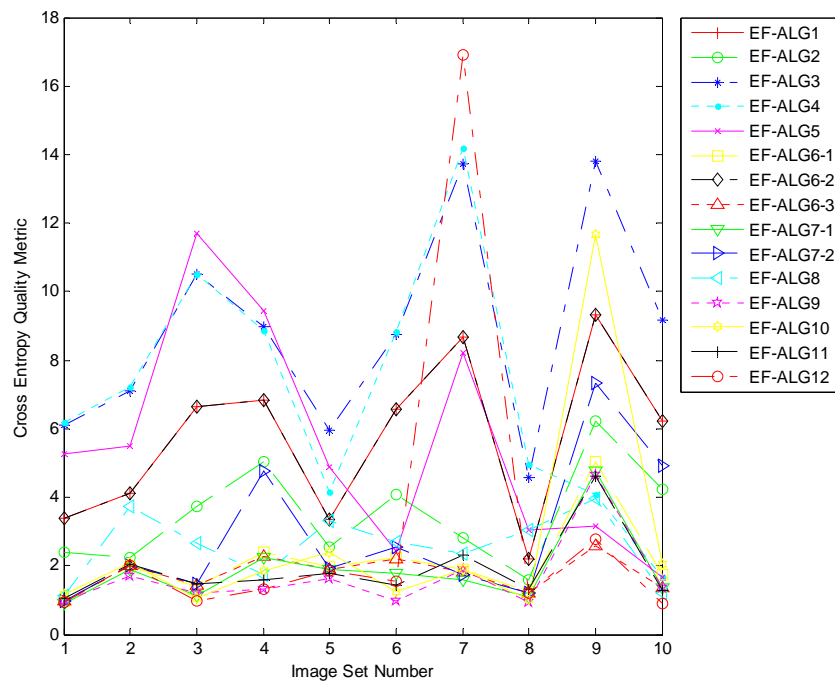**Figure 5.3 Comparison of the EF Algs. by Using Cross Entropy**

In cross entropy quality metric, EF-ALG3 is by far the worst among the algorithms because EF-ALG3 gives low contrast fused images. Since EF-ALG3 does not preserve the details in the input images, the dissimilarity between the input images and the fused image increases. Moreover, EF-ALG4 and EF-ALG5 have the second and third highest scores in most of the image sets. These algorithms use block based methods for image fusion. Since working on block based has a significant impact on histogram shape, the probability density function (PDF) of the fused image differentiates from the PDF of the input images so much. Furthermore, these algorithms do not use one of the input images, or they use one of the input images more that the others when the dynamic range of the scene is high. Therefore, the resultant image loses the similarity with one or two input images more than the other methods, so the average cross entropy decreases.

All MD based algorithms and some of the DCT based algorithms perform better than the others in this quality metric. Their performances are very similar, and it is difficult to distinguish which one is better. However, when the exact scores are examined, it is observed that EF-ALG9 performs slightly better than the others. Since EF-ALG9 gives saturated fused images, the similarity between the saturated input images and the fused image increases. Therefore, EF-ALG9 is rated the lowest scores in this quality metric.

As a summary, the results show that the quality metric cannot distinguish the successful algorithms from the others, and the similarity is not a reliable feature to evaluate the transferred information from the input images to the fused image in EF.

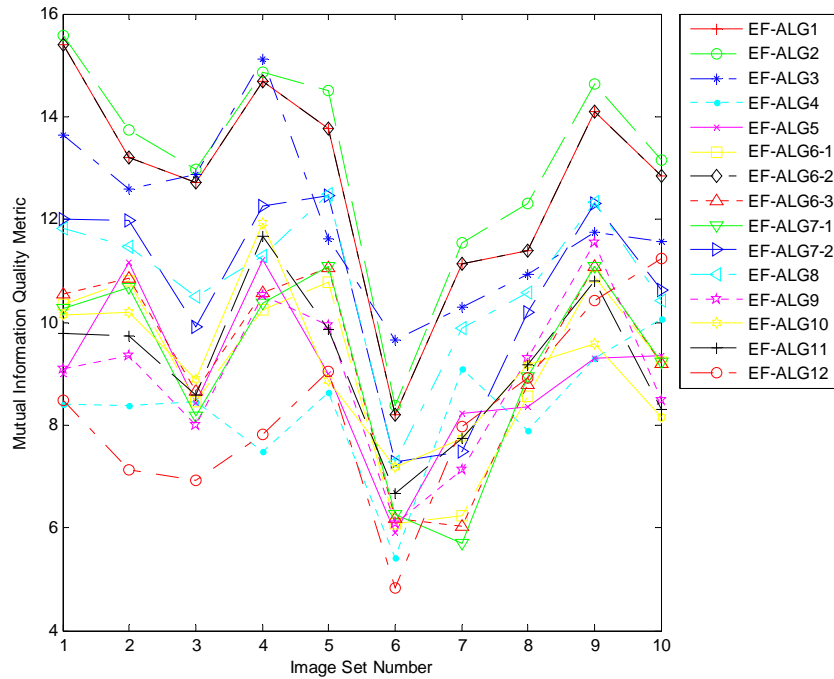### 5.2.1.4   Mutual Information Quality Metric

Figure 5.4 Comparison of the EF Algs. by Using Mutual Information

In mutual information quality metric, EF-ALG1, EF-ALG2 and EF-ALG6-2 perform better than the others. Since these algorithms use average operation for image fusion, the dependence between the input images and the fused image increases. That is why they are rated higher than the others. Furthermore, EF-ALG3 has the next highest scores due to its simple fusion rules.

EF-ALG9 and EF-ALG12 are rated lower than the others in this quality metric. The output of these algorithms has many saturated pixels because these algorithms do not guarantee that the fused image has pixels which are in the acceptable range. Therefore, EF-ALG9 and EF-ALG12 lose the dependence between the input images and the fused image. That is why they perform the worst in this quality metric. Additionally, EF-ALG4 and EF-ALG5 are rated lower than the most of the algorithms in this quality metric.

As a summary, mutual information evaluates the dependence between the input images and the fused image. However, it is observed that the simple algorithms are rated higher than the other successful algorithms in this quality metric. Therefore, the dependence is not a reliable feature to evaluate the performance of EF algorithms.

5.2.1.5   Universal Image Quality Index Quality Metric
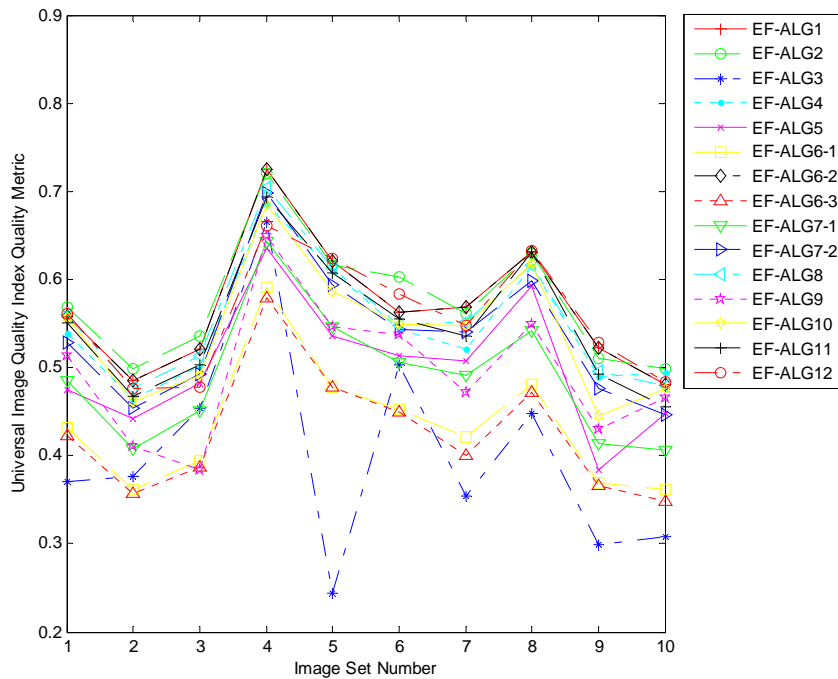
Figure 5.5 Comparison of the EF Algs. by Using UIQI

In UIQI quality metric, EF-ALG1 and EF-ALG2 have the highest scores. The MD based algorithms, which are EF-ALG12, EF-ALG11 and EF-ALG8, have the next highest scores. This is an expected result for MD based algorithms because they are good at transferring salient features from the input images to the fused image. Therefore, they can preserve the correlation between the input images and the fused image. However, this is not an expected result for EF-ALG1 and EF-ALG2. Since they use average operation for image fusion, the correlation and the mean

luminance similarity between the input images and the fused image are high. Therefore, UIQI gives high scores to EF-ALG1 and EF-ALG2.

In this quality metric, EF-ALG6-1, EF-ALG6-3 and EF-ALG3 perform the worst. Since these algorithms usually give fused images with blocking artifacts, the contrast and the mean luminance closeness are low. Hence, they are rated lower than the others.

As a conclusion, UIQI cannot distinguish simple averaging algorithms from the other successful algorithms. Therefore, it is not appropriate to evaluate EF algorithms.

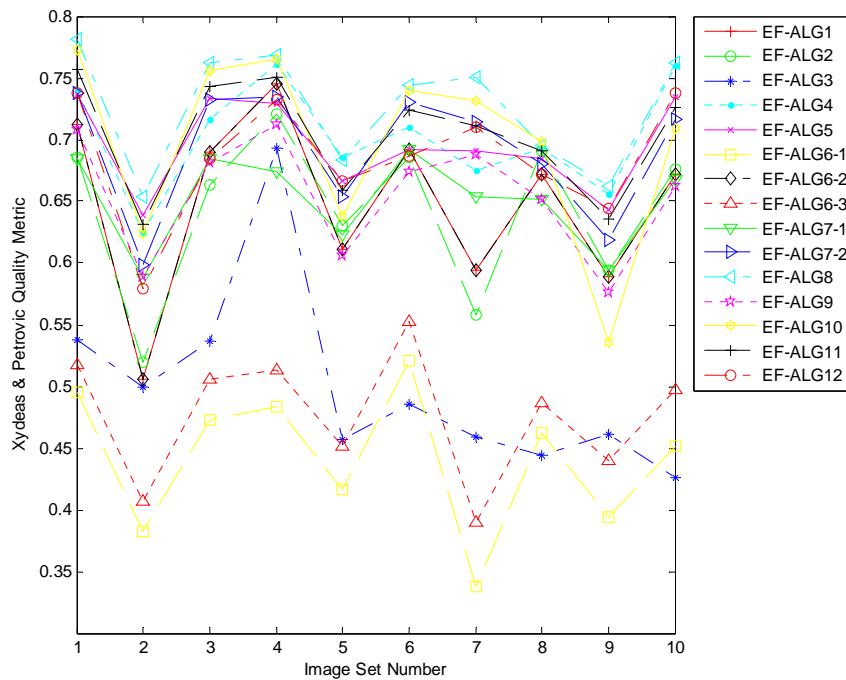## 5.2.1.6   Xydeas & Petrovic Quality Metric

Figure 5.6 Comparison of the EF Algs. by Using Xydeas & Petrovic

In Xydeas & Petrovic quality metric, MD based algorithms perform better than the others. EF-ALG8 has the highest scores. This result is reasonable because this algorithm uses Gradient pyramid for image fusion. EF-ALG10 has the second highest scores in most of the image sets. This is an expected result because EF-ALG10 uses Laplacian pyramid, which is good at transferring the edge details to the fused image. EF-ALG11 has the third highest scores which can also be explained with the previous reason.

In the above figure, EF-ALG6-1 and EF-ALG6-3 have the lowest scores in most of the image sets. This is an expected result because the output of these algorithms has noticeable blocking artifacts. These artifacts cause the edge similarity between the input images and the fused image to decrease. Therefore, these algorithms are rated the lowest scores. Then, EF-ALG-3 has the next lowest scores. This result is reasonable because most of the time, the output of EF-ALG-3 has low contrast and poor edge details. That is why EF-ALG-3 is rated the lowest.

As a conclusion, the results show that the edge detail is a reliable quality metric to measure the transferred salient information from the input images to the fused image. Therefore, Xydeas & Petrovic can be used to evaluate the performance of EF algorithms.

5.2.1.7 Spatial Frequency Quality Metric
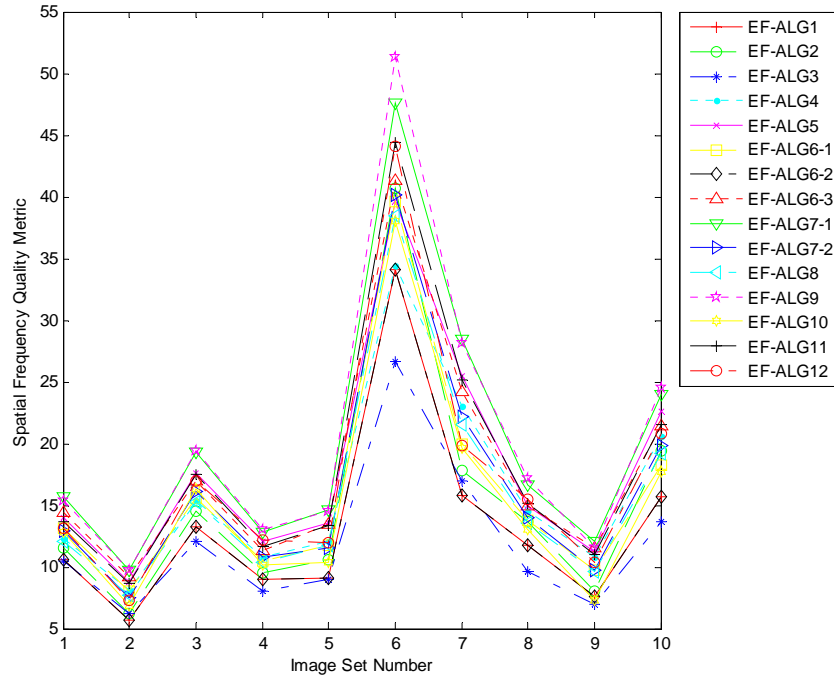
Figure 5.7 Comparison of the EF Algs. by Using Spatial Frequency

In the above figure, EF-ALG9 has the highest scores in most of the image sets. Most of the time, the output of EF-ALG9 has high contrast and many saturated pixels. These features increase the spatial frequency in an image. That is why EF-ALG-9 is rated the highest. EF-ALG7-1, EF-ALG6-1 and EF-ALG6-3 have the next highest scores. The common point of these algorithms is that they suffer from blocking artifact problem. As it has been mentioned before, blocking artifacts cause unwanted and nonexistent edges in the fused image. That is why they have high points in this quality metric.

In this quality metric, EF-ALG3 and EF-ALG1 perform the worst in the most of the image sets. It is an expected result because the output of these algorithms has poor contrast and poor edge details.

This metric considers the spatial frequency in an image and shows us how much details image has. However, it cannot distinguish blocking artifacts from the salient features. These artifacts are considered as details, and the algorithms, which cause these artifacts, are rated the highest scores. Moreover, algorithms that recover the details in saturated regions are rated lower than the algorithms that suffer from the blocking artifacts. As a result, spatial frequency is not appropriate for evaluating EF algorithms.
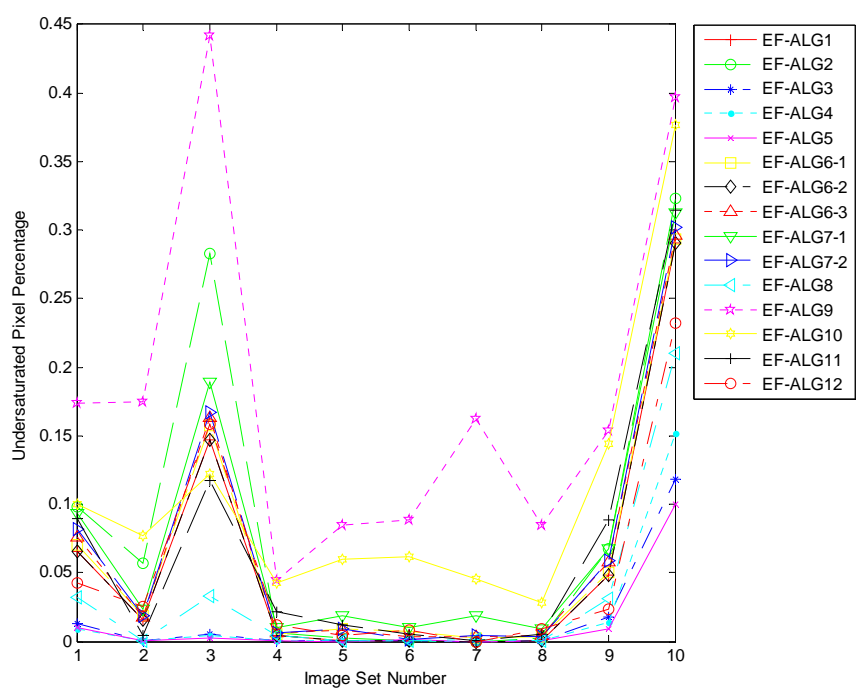
5.2.1.8   Saturated Pixel Percentage

Figure 5.8 Comparison of the EF Algs. by Using Undersaturated P. P.
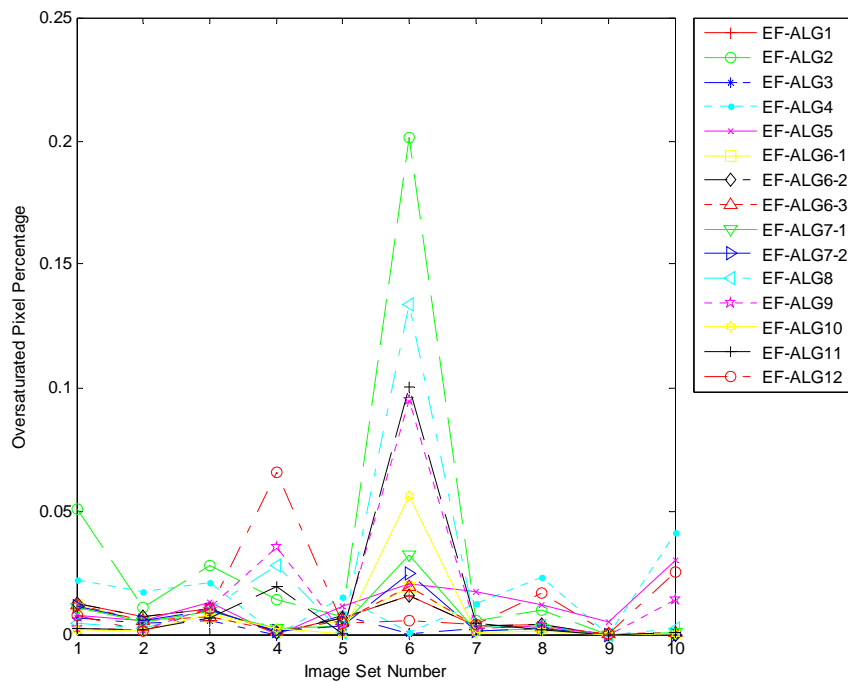
Figure 5.9 Comparison of the EF Algs. by Using Oversaturated P. P.

Saturated pixel percentage is not a well-known quality metric and is implemented to evaluate the performance of EF algorithms in this thesis. In this part, undersaturated and oversaturated pixel percentage quality metrics are examined. In undersaturated pixel percentage, EF-ALG9 performs the worst. This result is reasonable because the output of EF-ALG9 usually has many saturated pixels. In oversaturated pixel percentage, EF-ALG2 performs the worst because EF-ALG2 uses the weighted average which usually gives bright fused images. However, remaining algorithms have similar scores, and it is hard to compare the performances of the algorithms.

As a conclusion, the results show that saturated pixel percentage is not appropriate to evaluate the performance of EF algorithms. This is a simple quality metric that should be supported with the other quality metrics. It only gives information about

the number of saturated pixels in the fused image. It does not consider the contrast or the edge details of the fused image.

5.2.2 Experimental Results for the High Dynamic Range Imaging Algorithms

In the literature, there are comprehensive and detailed tests to evaluate the performance of the tone mapping operators [65-67]. Since this thesis focuses on EF algorithms and aims to understand general principles of the HDRI algorithms, a brief comparison of the HDRI algorithms by using objective quality metric is given. This section shows necessity of the successful HDRI algorithms to enhance the dynamic range of the images. Therefore, new HDRI algorithms have been proposed in the literature.

In this subsection, the implemented HDRI algorithms are compared with each other through Xydeas & Petrovic quality metric. This quality metric is used for the comparison of HDRI algorithms because it gives the most accurate results in the comparison of the EF algorithms. There are many combinations for the HDRI algorithms. In this thesis, two algorithms for recovering HDR radiance map and six tone mapping algorithms are implemented. These algorithms are combined with each other, which makes a total 12 combination, to make a comparison within the algorithms. The examination of these combinations is very problematic. Therefore, firstly, one of the HDR radiance map algorithms is chosen, and HDR radiance map of the each image set is calculated by using the chosen algorithm. Then, the tone mapping algorithms are compared with each other by using the calculated HDR radiance maps. This step is repeated for two HDR radiance map algorithms. After this step, the HDR radiance map algorithms are compared with each other by using one of the tone mapping algorithms. Since there are six different tone mapping algorithms, this process is repeated for six times.

In the following sections, the results related to the evaluation of the HDR radiance map and tone mapping algorithms are given.

5.2.2.1   Evaluation of the Tone Mapping Algorithms
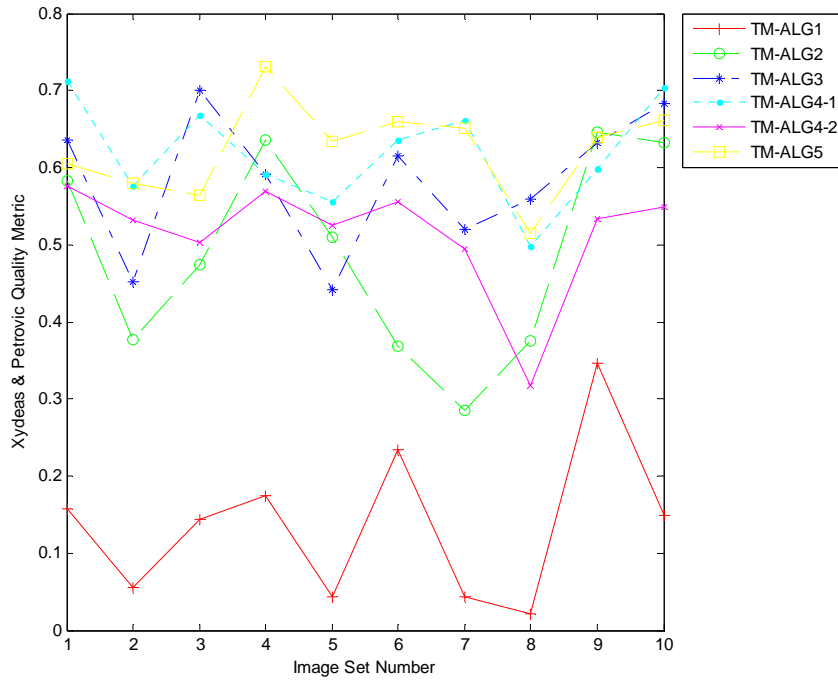
Figure 5.10 Comparison of the TM-ALGs by Using the HDR-ALG1

First of all, HDR-ALG1 is selected to calculate the HDR radiance maps of the image sets. Then, the tone mapping algorithms are compared with each other. As seen in the figure, TM-ALG1 is rated lower than the others in all image sets. This is an expected result because TM-ALG1 is a simple tone mapping algorithm, which uses linear scaling for the dynamic range compression. TM-ALG2 has the second lowest scores. However, TM-ALG2 performs better than the others in some of the image sets. The common characteristic of these image sets is that they are relatively dark scenes. Since TM-ALG2 has a characteristic curve that compresses high radiance values more than low radiance values, it performs better in the dark scenes. That is why TM-ALG2 performs well in these image sets.

83

TM-ALG5 performs the best in most of the image sets. However, it is rated lower than TM-ALG3 and TM-ALG4-1 in some image sets. The common characteristic of these image sets is that they include bright objects that lead dynamic range to increase. It shows that the performance of TM-ALG5 deteriorates in the high dynamic range scenes.

Figure 5.11 Comparison of the TM-ALGs by Using the HDR-ALG2
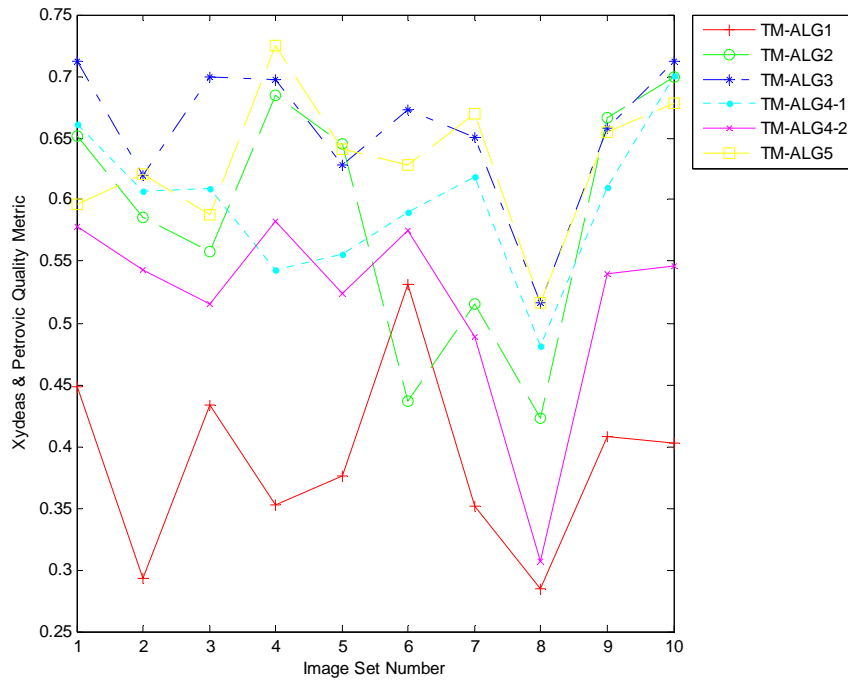
At the next step, HDR-ALG2 is chosen to calculate required HDR radiance maps, and then the tone mapping algorithms are compared with each other. The results are very similar to the previously given results. TM-ALG1 has again the lowest scores. This is expected due to the above stated reason. TM-ALG5 and TM-ALG3 have similar scores, so the difference is not remarkable.

5.2.2.2 Evaluation of the HDR Radiance Map Algorithms

Until now, the performance evaluation of the tone mapping algorithms has been discussed. In this part, the implemented HDR radiance map algorithms are compared with each other. Since objective quality metrics accept low dynamic range images as input, the outputs of HDR radiance map algorithms is mapped to low dynamic range images by the help of the tone mapping algorithms. Then, the comparison of the HDR radiance map algorithms is repeated for each tone mapping algorithms. However, there are six different tone mapping algorithms. To avoid vast figure number, the results are not plotted as figures but a table which summarizes the comparison result of the HDR radiance map algorithms is given as follows:

Table 5.2 Comparison of HDR Algs. by Using Xydeas & Petrovic QM

Name of the Tone Mapping Algorithms	First Rank	Second Rank
TM-ALG1	HDR-ALG2	HDR-ALG1
TM-ALG2	HDR-ALG2	HDR-ALG1
TM-ALG3	HDR-ALG2	HDR-ALG1
TM-ALG4-1	HDR-ALG1	HDR-ALG2
TM-ALG4-2	HDR-ALG2	HDR-ALG1
TM-ALG5	HDR-ALG1	HDR-ALG2

The results, which are given in the previous two parts, show that the edge information is a reliable quality measure and can be used to evaluate the performance of HDRI algorithms. As a conclusion, Xydeas & Petrovic metric is the most appropriate objective quality metric to evaluate the performance of image dynamic range enhancement algorithms.

5.2.3 Visual Results of Objective Test

The goal of this subsection is to give some visual results that have been obtained in the objective test. As it has been mentioned before, the test includes ten image sets, and each image set consists of three differently exposed images. It makes thirty input images total. In addition, there are fifteen EF algorithms and several HDRI algorithms in the test. Therefore, the number of output images is more than one hundred just for this test. It is not possible to show all output images, so some of the images are given in this subsection. In the test, it is observed that Xydeas & Petrovic is the most appropriate quality metric to evaluate EF algorithms. Therefore, the outputs of the best and worst performed EF algorithms with respect to Xydeas & Petrovic quality metric are given for each image set. In the following figures, the images on the left belong the algorithms that have the highest scores in each image set, and the right ones belong the worst algorithms. The figures are given as follows:

a)   b)

Figure 5.12 Image Set-1 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a) b)

Figure 5.13 Image Set-2 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a) b)

Figure 5.14 Image Set-3 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a) b)

Figure 5.15 Image Set-4 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a)   b)

Figure 5.16 Image Set-5 in Objective Test a) EF-ALG4 b) EF-ALG6-1

a)   b)
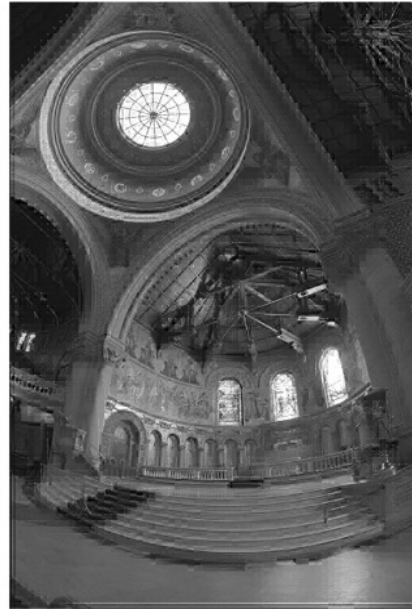
Figure 5.17 Image Set-6 in Objective Test a) EF-ALG8 b) EF-ALG3

a)   b)

Figure 5.18 Image Set-7 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a)   b)

Figure 5.19 Image Set-8 in Objective Test a) EF-ALG10 b) EF-ALG3

a)   b)

Figure 5.20 Image Set-9 in Objective Test a) EF-ALG8 b) EF-ALG6-1

a)   b)

Figure 5.21 Image Set-10 in Objective Test a) EF-ALG8 b) EF-ALG3

5.3   Subjective Evaluation of the Algorithms

Subjective assessment has been used to compare the performances of image fusion algorithms for years. In parallel, some objective quality metrics have been developed so far, but these metrics are required to be improved. Therefore, the subjective evaluation of image fusion algorithms is still an important issue in the literature [58-60]. Subjective assessments mainly consist of two subcategories. These subcategories are task related (active) and descriptive (passive) tests [59]. In the task related tests, participants are asked to evaluate different scenarios by

considering a defined task. Then, the performance of algorithms is evaluated by considering the performance of the participants in completing the task, average time to define and locate a specific target in the images etc [61]. In the descriptive tests, participants directly score or rank fused images based on their impressions.

In this thesis, a descriptive test is performed to evaluate different image fusion algorithms and to see the correlation between objective quality metrics and human subjective ratings.

In the following sections, the details of the subjective test are explained.

5.3.1 Selection of Image Sets and Algorithms

In this thesis, fifteen EF algorithms and eight HDRI algorithms are implemented. Including all algorithms into the test makes the test longer and inefficient. Therefore, one of the image dynamic range enhancement methods has to be chosen. Since this thesis gives more importance to EF algorithms, EF is chosen. Seven EF algorithms are included to the subjective test in order to avoid long test duration.

In the subjective test, the number of image sets is limited to five, and one additional image set is used to explain the goal of the test to the participants. Each image set consists of three different exposed images which are underexposed, normal and overexposed. The selected images represent different illumination conditions in the real life. Furthermore, image sizes are kept alike to avoid any distortion may result in rescaling. The image sets, used in the subjective test, can be found in Appendix A. The resolutions of the images are given at the bottom of the image sets.

As stated previously, seven EF algorithms are used in the test. Therefore, the chosen algorithms should represent most of the implemented algorithms. EF-

91

ALG4 and EF-ALG7-2 are chosen from the blocked and DCT based algorithms respectively. Since EF-ALG-7-1 suffers from blocking artifacts, the improved algorithm, which is EF-ALG7-2, is included to the test. Other five algorithms are selected from the MD based algorithms. Since there are five MD based algorithms, all algorithms are included to the test.

5.3.2   Subjective Test Procedure

Test environment is a very crucial and must be controllable. Display used in the test, the participant's distance to display, lighting conditions and interaction with the participant are very important points. In this test, the required rules in a subjective test have been followed as much as possible. Since this assessment is a simple implementation of real subjective tests, some rules have been violated.

As stated previously, the subjective test consists of five image sets and seven EF algorithms. It makes thirty five fused images total. In additional to this number, some of the fused images have been repeated to measure the consistency of the participants. For each image set, two of the fused images have been repeated. Therefore, the number of repeated images becomes ten for five image sets. However, there are seven different EF algorithms. To match the algorithm number, which is seven, and the repeated image number, which is ten, three of the algorithms' outputs have been repeated twice. The remaining algorithms have been repeated only once. When the total number of repeated images and the fused images are summed, it makes forty five images. In the test, the number of the images has been kept minimum as much as possible to avoid long test duration.

The representation of the fused images is another crucial point in subjective tests. In the literature, there are different representation methods that have been used in the past subjective tests. In this test, the fused images and the corresponding input images have been shown together. Example test image that has been used to explain the goal of the subjective test to the participants is given as below:

Figure 5.22 Example Image Used in Subjective Test

The above image is composed of four sub-images. Three sub-images at the top of the image are the input images, and the sub-image at the bottom is the fused image.

The input images have been added to the output image to make scoring easy for the participants. By means of this, the participants can see how much salient information is transferred from the input images to the fused image. Other algorithms in the same image set can be compared with each other by using the input images. They constitute a reference point for the comparisons.

In the test, these test images have been shown image set by image set. The order of the output images has been organized randomly in each image set. However, the image sets have not been ordered randomly to provide easy and accurate scoring.

In the test, the images have been shown on a standard notebook under standard white office light. The model of the notebook is "HP EliteBook 8540w Mobile Workstation". It has 15.6'' matt screen. Matt screen has been chosen to avoid reflection problems on the screen. Furthermore, standard "Microsoft Office

Picture Manager" has been used to show the images. To evaluate the algorithms, one to ten scoring scale has been used in the test. The worst and the best performances are denoted by one and ten respectively. There has been no time limit in the test.

The goal of the test has been explained by using same explanations and examples before each subjective test. First of all, the definition of dynamic range has been made and then the necessity of the image dynamic range enhancement has been explained. The participants have been asked to score forty five images by considering two performance measures. These measures are the capability of transferring details from the input images to the fused image and the contrast of fused images. The first requirement allows measuring the details in bright and dark areas. As it has been mentioned before, the output of some algorithms has limited contrast, but the details in saturated regions are recovered. It is an unwanted situation for EF. Ideal fused image should have high contrast and enhanced dynamic range. Therefore, all participants have been asked to score the fused images with respect to these two performance measures. The participant's instructions used in this test can be found given in Appendix B.

5.3.3 Participants

Thirty participants (25 males, 5 females) have attended this subjective test. The average age of the participants is 25.3, and standard deviation is 2.73. All participants have normal or corrected-to-normal vision. The age distribution of the participants is given as follows:
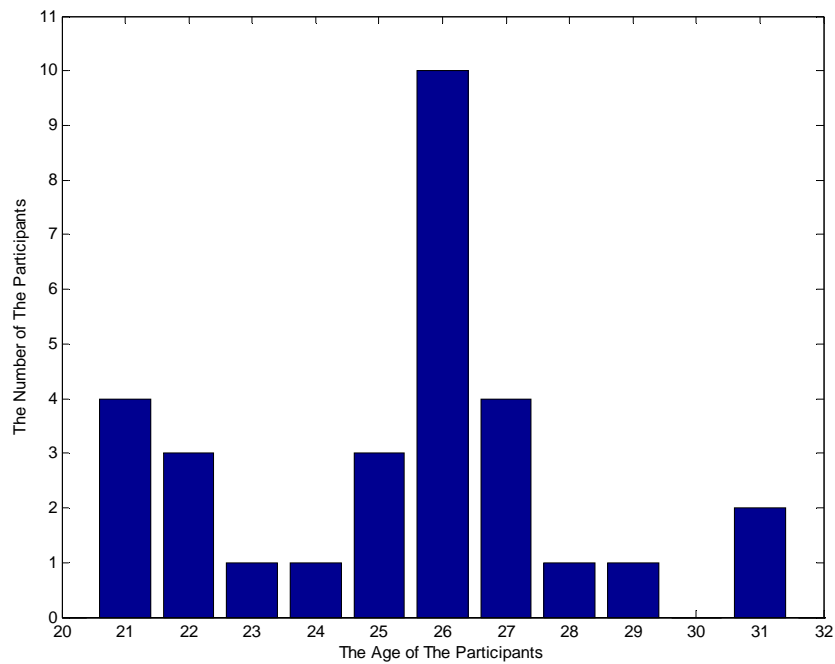
Figure 5.23 Age Distribution of Participants

5.3.4 Correlation between Objective and Subjective Scores

The main goal of this subsection is to evaluate the correlation between the objective quality metrics and the subjective scores.

There have been thirty participants in the test, and the participants have been asked to score images using the numbers ranging from one to ten. There is no restriction about scoring, so the score of each participant for a specific image become a little bit different. Hence, there should be a filter to remove scores that are so different from the mean of the scores. This filter can be realized by using a simple standard deviation filter. For each test image, the mean and the standard deviation of all participants' scores are calculated, and scores that are far away from one standard deviation are discarded. By means of this, some inaccurate scores are eliminated.

In the next step, the objective and subjective scores are scaled to bring all scores into the same range. In this study, all scores are scaled into the range of zero to one by linear scaling. However, the chosen scale method is very crucial, and it may change the result of the experiment. Therefore, different scale methods have been tried. First of all, the subjective scores are scaled. In the test, the participants have scored the images with respect to the other images in the same image set. Firstly, they have checked all test images in the current image set and then have scored the images relatively. The participants have not considered different image sets together during scoring. It means that the subjective scores are image set based. Therefore, the most appropriate scaling method is image set based scaling. Each score is scaled by only considering scores in the same image set. Therefore, all subjective scores are mapped locally in each image set. Then, the mean of the participants' scores for each test image is calculated. The calculated mean values are used as the subjective scores of the test images. For the objective scores, image set based scaling method is again used to keep the correlation with the subjective scores.

As stated previously, some of the images are repeated in order to measure the consistency of the participants in the test. However, it is observed that the scores are mostly consistent, and the standard deviation filter discards most of the inaccurate scores. Therefore, there is no need any extra processes to check the consistency.

The following subsections give the results related to the correlation between the objective and subjective scores. In the following figures, only image sets are indicated with different colors. Furthermore, the correlation coefficient of the scores is calculated for each quality metric and is given in the table as follows:

Table 5.3 Correlation Coefficients of Objective - Subjective Scores

Quality Metric	Correlation Coefficient
Standard Deviation	-0.5170
Entropy	0.5064
Cross Entropy	-0.2090
Mutual Information	0.3748
Universal Image Quality Index	0.3426
Xydeas & Petrovic	0.6792
Spatial Frequency	-0.3111
Undersaturated Pixel Percentage	-0.3988
Oversaturated Pixel Percentage	-0.3317

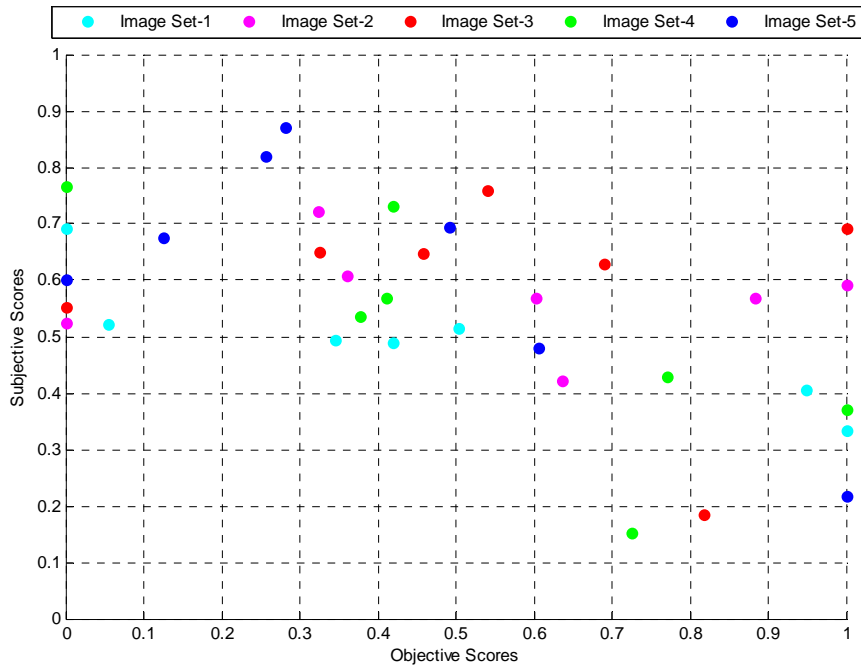**5.3.4.1   Standard Deviation Quality Metric**

Figure 5.24 Standard Deviation vs Subjective Scores

As can be seen in the figure, there is a negative correlation between the objective and subjective scores. In the objective evaluation section, it has been already stated that this quality metric does not give accurate results. Since it uses standard deviation as a quality metric, algorithms that result in so many saturated pixels are evaluated higher than the others.

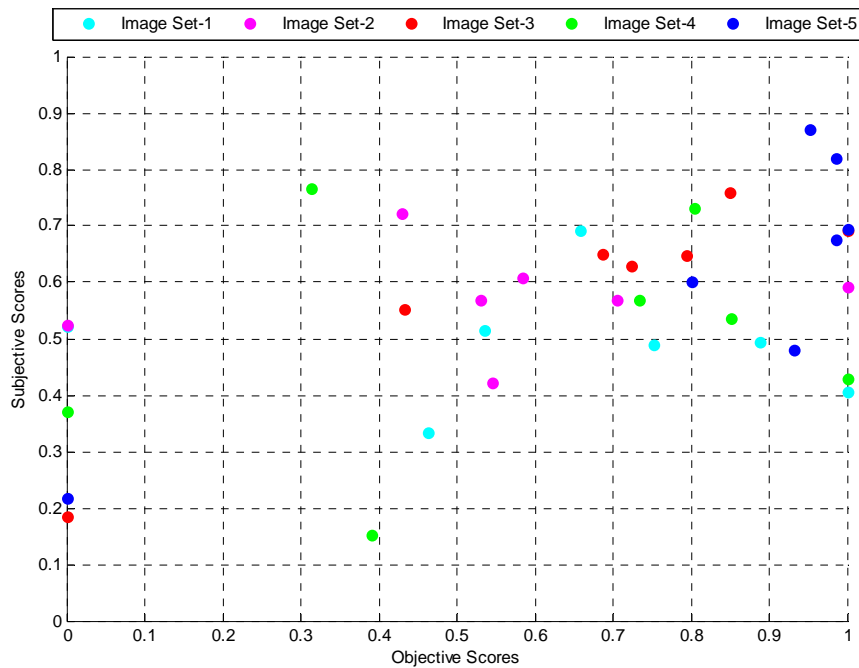### 5.3.4.2   Entropy Quality Metric

Figure 5.25 Entropy vs Subjective Scores

In this quality metric, there is a positive correlation between the scores, but the correlation is not very strong. In the objective evaluation section, it has been observed that the performance of this quality metric is better than most of the quality metrics. It has the capability of measuring the contrast of the fused images and detecting saturated pixels. This capability provides accurate evaluation of EF algorithms.
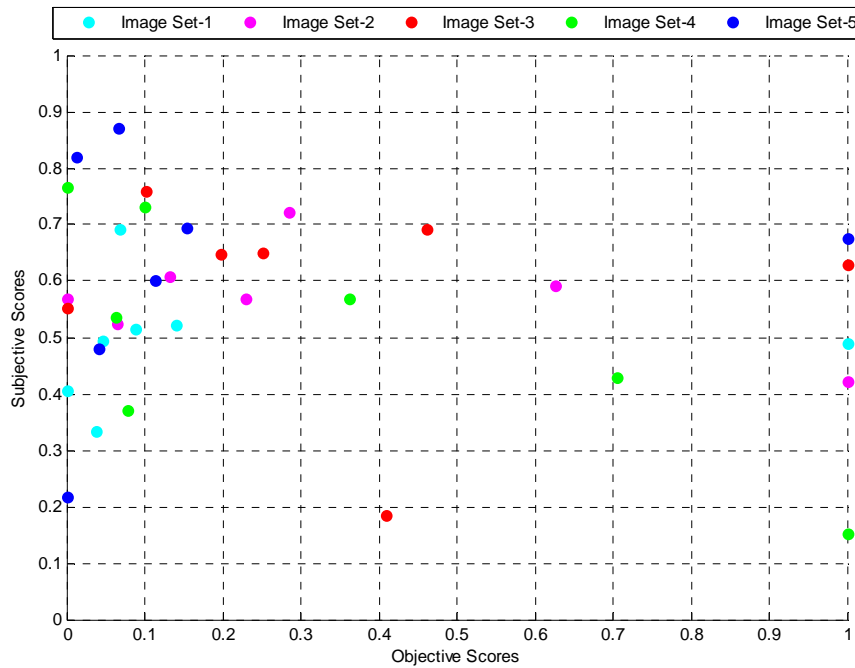
5.3.4.3   Cross Entropy Quality Metric

Figure 5.26 Cross Entropy vs Subjective Scores

In the above figure, there is a weak negative correlation between the scores. There is an exception for this quality metric because lower cross entropy means better image fusion. Therefore, the negative correlation is a desired situation in this quality metric. However, the correlation coefficient is very small and can be thought that the scores are uncorrelated with each other. This result is the same as what we have already obtained in the objective evaluation section. In that section, it has been observed that algorithms that give saturated outputs are rated higher than the others.

5.3.4.4   Mutual Information Quality Metric
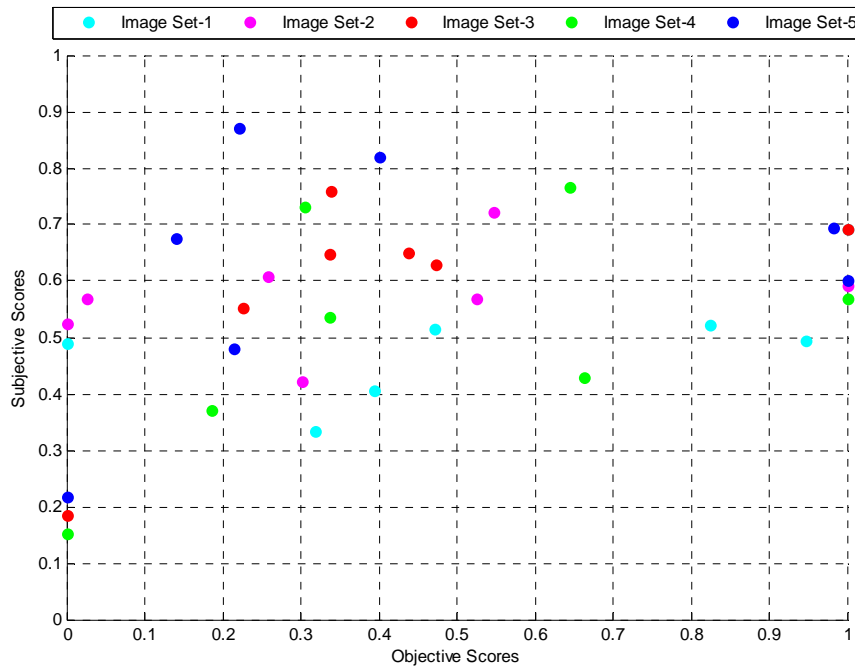
Figure 5.27 Mutual Information vs Subjective Scores

In the above figure, there is a weak correlation, but it can be accepted as uncorrelated. Mutual information uses dependence feature to measure the transferred information from the input images to the fused image. In the objective evaluation section, it has been observed that mutual information cannot distinguish the best and worst algorithms. This result is supported by the subjective scores.

5.3.4.5   Universal Image Quality Index Quality Metric
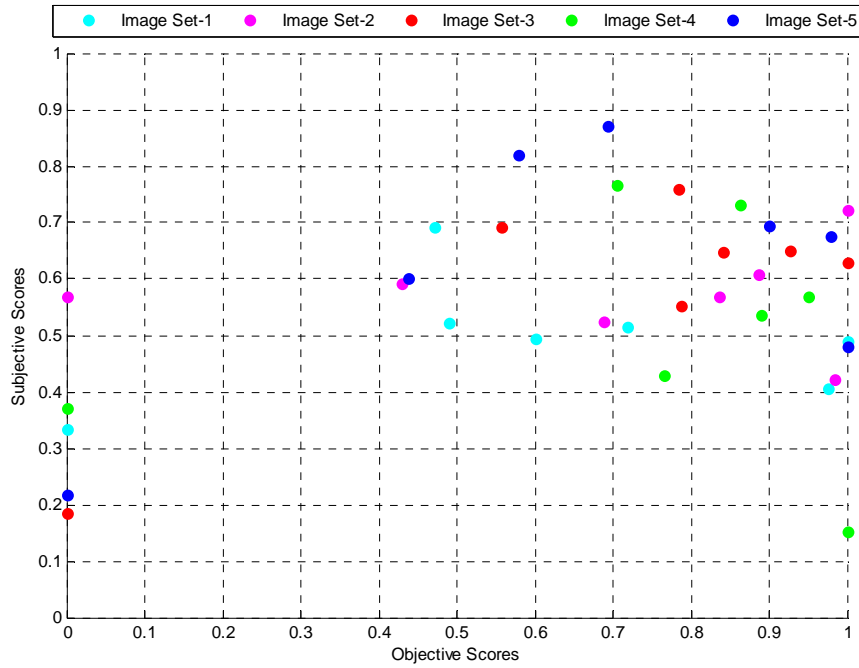

Figure 5.28 Universal Image Quality Index vs Subjective Scores

There is no correlation between the objective and subjective scores in this quality metric. This is an expected result because this quality metric is not good at measuring the transferred important information from the input images to the fused image as stated previously.

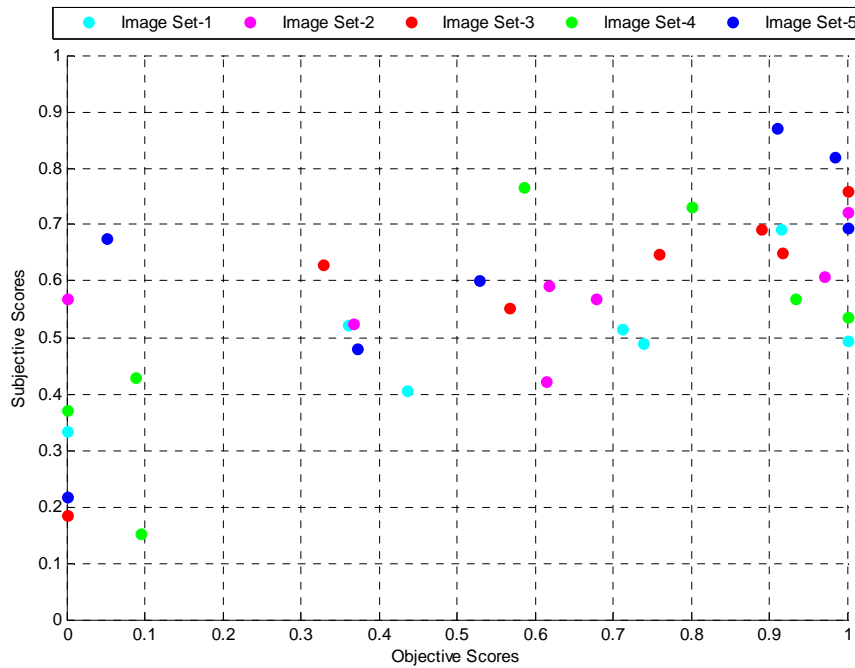### 5.3.4.6   Xydeas & Petrovic Quality Metric

Figure 5.29 Xydeas & Petrovic vs Subjective Scores

As can be seen in the figure, Xydeas & Petrovic quality metric has a significant positive correlation with the subjective scores. This quality metric has the highest correlation with the subjective scores among the quality metrics. It uses the edge strength and the edge orientation in order to measure the transferred important information from the input images to the fused image. These features are very accurate in measuring the blocking artifacts in the fused images and the contrast of the fused images. This conclusion is fully supported by the subjective scores. The conclusion has been made in the objective evaluation section is supported by the subjective scores. As a conclusion, Xydeas & Petrovic quality metric is the most appropriate quality metric to evaluate the performance of EF algorithms.

5.3.4.7   Spatial Frequency Quality Metric
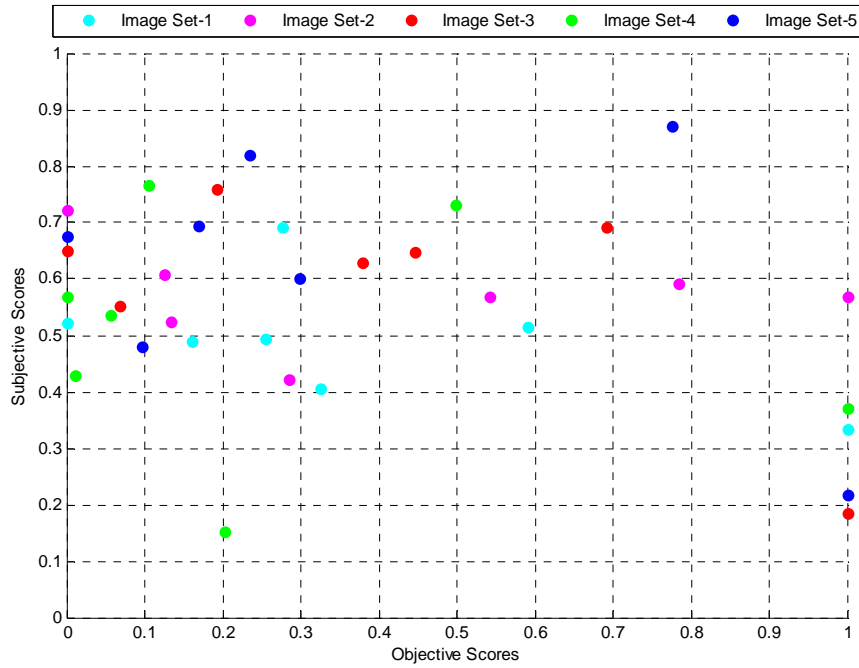
Figure 5.30 Spatial Frequency vs Subjective Scores

In the above figure, the scores are almost uncorrelated with each other. It shows that this quality metric is not good at evaluating the performance of EF algorithms. This conclusion has already been made in the objective evaluation section.

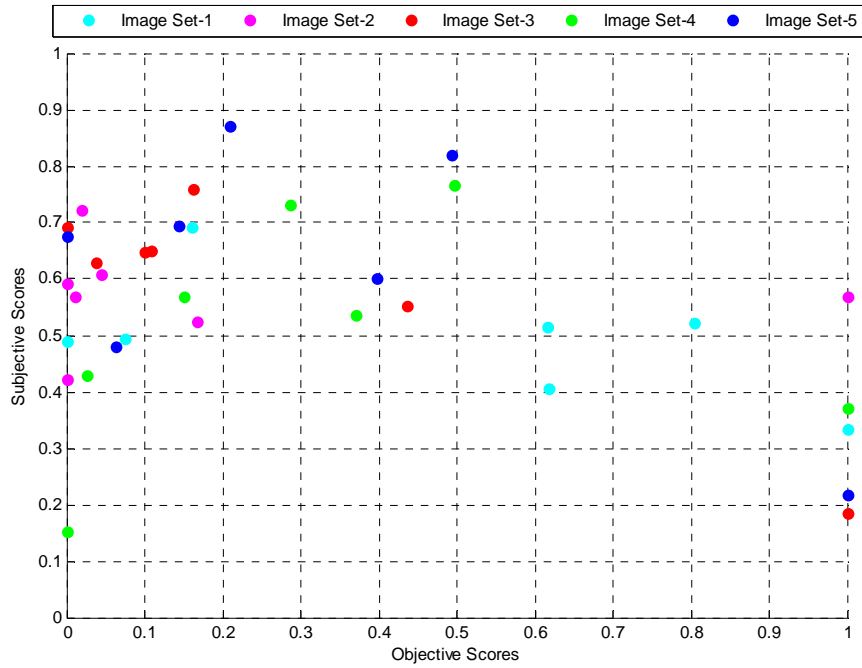## 5.3.4.8   Saturated Pixel Percentage

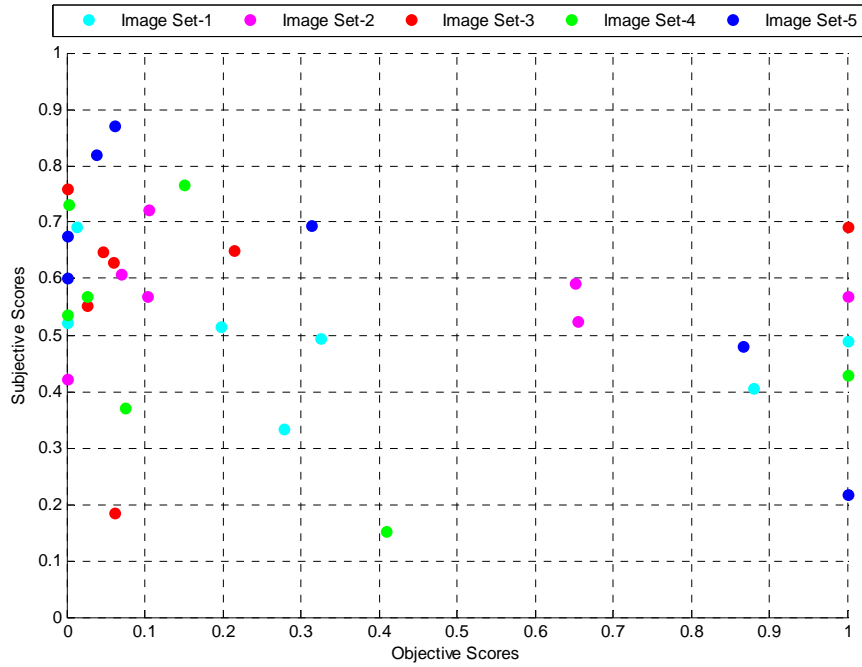**Figure 5.31 Undersaturated P.P. vs Subjective Scores**

Figure 5.32 Oversaturated P.P. vs Subjective Scores

In the above figures, there are negative correlations between the scores.

5.3.5 Evaluation of the Algorithms

Until now, the results related to the correlation between the scores have been discussed. In this subsection, the individual performance of the algorithms is examined by using the objective and subjective scores in the subjective test. Therefore, the subjective scores and the objective scores belong to Xydeas & Petrovic quality metric are plotted separately to evaluate the performance of the algorithms.

Since this subsection focuses on the performance of the algorithms not the correlation between the scores, the objective and subjective scores are used directly without performing any scaling. For the subjective scores, the mean scores, which have been calculated previously, are directly used. Then, the mean scores belong to each algorithm are plotted for each image set. The image sets are denoted by different colors in the figures, and the mean of the scores in each image set is indicated with a different marker. The same visualization is used in the figure of the objective scores. By doing that, the deviation and mean of the scores are made visible for each algorithm. The figures are given below:
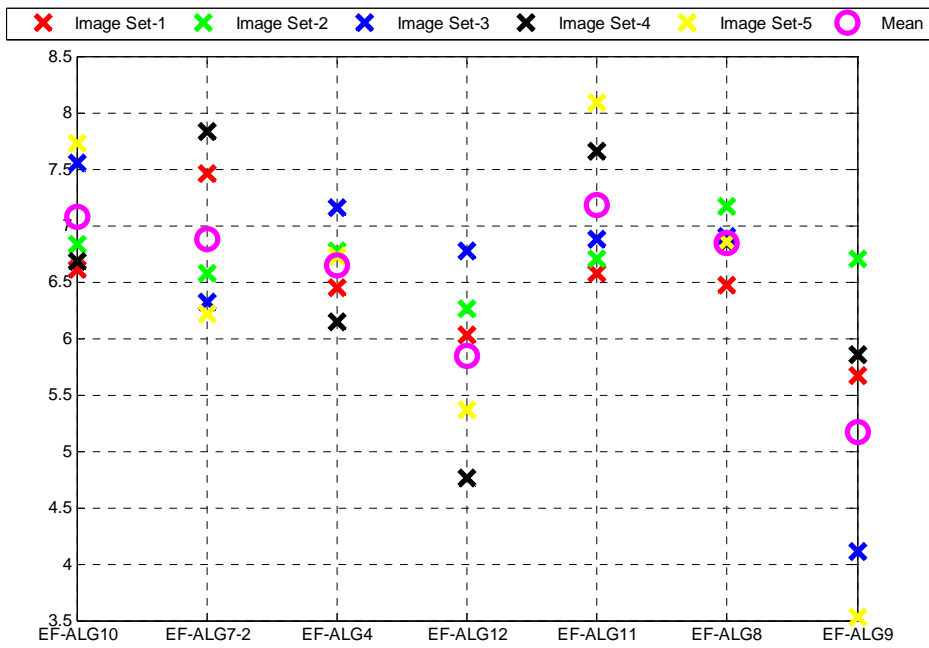
Figure 5.33 Subjective Scores of the EF Algorithms
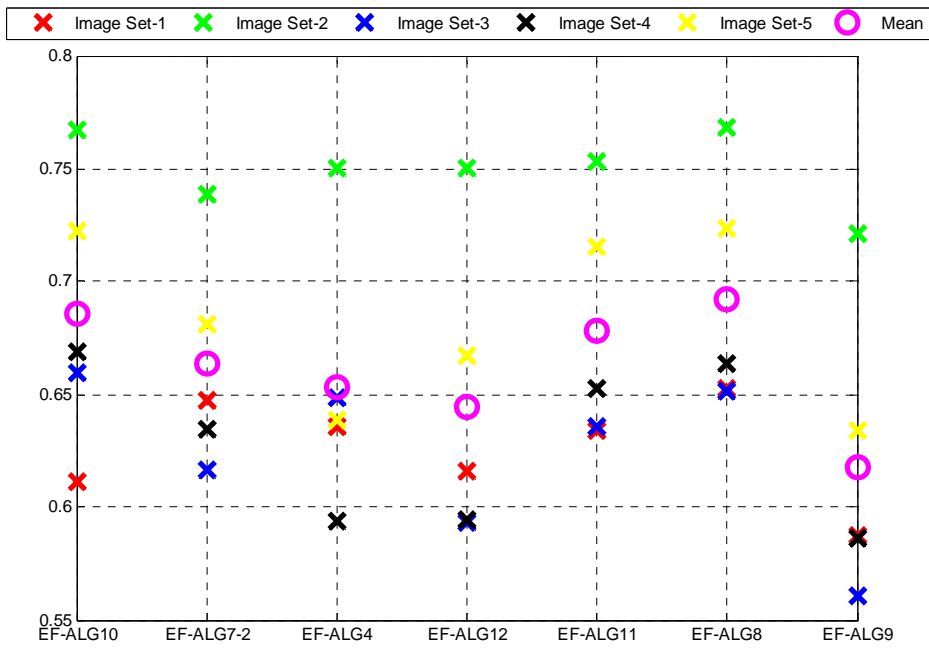
Figure 5.34 Objective Scores of the EF Algorithms

106

As can be seen in the figures, EF-ALG7-2, EF-ALG8, EF-ALG10 and EF-ALG11 are rated relatively higher than the others in both tests. Most of these algorithms are MD based image fusion algorithms except EF-ALG7-2, which is DCT based. This is an expected result because MD based algorithms are good at extracting the salient features in the input images. Additionally, algorithms that have relatively low scores in both tests are EF-ALG12 and EF-ALG9. This result is reasonable because these algorithms give saturated fused images due to the previously stated reasons.

5.3.6 Visual Results of Subjective Test

The goal of this subsection is to give some visual results that have been obtained in the subjective test. In this subsection, the outputs of the algorithms which have performed the best and the worst in the subjective test are presented. For each image set, the image that has the highest score is shown on the left side, and the image with the lowest score is shown on the right side.

a)   b)

Figure 5.35 Image Set-1 in Subjective Test a) EF-ALG7-2 b) EF-ALG9

Figure 5.36 Image Set-2 in Subjective Test a) EF-ALG8 b) EF-ALG12

Figure 5.37 Image Set-3 in Subjective Test a) EF-ALG10 b) EF-ALG9

Figure 5.38 Image Set-4 in Subjective Test a) EF-ALG7-2 b) EF-ALG12

a)   b)

Figure 5.39 Image Set-5 in Subjective Test a) EF-ALG11 b) EF-ALG9

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

6.1 Summary and Conclusions

In this thesis, image dynamic range enhancement methods are studied by using monochrome and static images. The main methods used in dynamic range enhancement are introduced briefly, and the detailed information about EF is presented. The well-known algorithms are implemented, and the performance of the algorithms is compared through the well-known objective quality metrics. Additionally, the correlation between the objective quality metrics and human visual system is studied in order to find the most appropriate objective quality metric.

Firstly, several HDRI algorithms are implemented to understand their approach in image dynamic range enhancement. It is observed that HDRI requires two processing steps to represent high dynamic range scenes with low dynamic range images. The first step includes HDR radiance map algorithms to calculate the radiance values of captured scene. The goal of these algorithms is to turn back to sensor radiance values from digital pixels values in order to recover the high dynamic range of captured scene. Therefore, most of the time, the exposure settings of the input images should be known. The outputs of these algorithms are high dynamic range images. However, these images cannot be displayed on standard screens. Hence, tone mapping algorithms are required. Several tone

mapping algorithms are implemented in this thesis. Then, the studies on this subject are not evaluated profoundly, and just a brief introduction has been made. Then, the second image dynamic range enhancement method, which is EF, is studied in details. The implemented EF algorithms consist of important image fusion methods such as DCT, pyramid transform and wavelet transform. Furthermore, literature searches show that EF is a new concept in image dynamic range enhancement. Unlike HDRI algorithms, all enhancement processes are performed in low dynamic range. Therefore, only one processing step is enough to enhance the dynamic range of captured scenes. It is more suitable for dynamic range enhancement in low dynamic range because it does not create high dynamic range images as an intermediate step. Due to the reasons mentioned above, EF has been becoming more popular than HDRI in image dynamic range enhancement.

The performances of the EF algorithms are compared with each other by the objective and the subjective tests. The results of the objective test show that the most appropriate quality metric to evaluate the performance of EF algorithms is Xydeas & Petrovic. Since Xydeas & Petrovic uses transferred edge information to the fused images as a performance measure, it can detect the blocking artifacts and the saturated pixels in the fused images. Additionally, it is observed that MD based algorithms perform better than the space and transform domain based algorithms. Then, the results of the subjective test are used to evaluate the performances of the EF algorithms and the correlation between the objective and subjective scores. In the subjective test, MD based algorithms are again rated higher than the other algorithms. This result is the same as the result in the objective test. Additionally, it is observed that Xydeas & Petrovic is the most correlated quality metric with the subjective scores. The cross correlation between the Xydeas & Petrovic and subjective scores is 0.6792 in the subjective test. Other quality metrics are not appropriate to evaluate image dynamic range enhancement algorithms. This shows that Xydeas & Petrovic quality metric is a reliable quality metric and can be used to compare image dynamic range algorithms without performing any subjective tests. In both tests, it is observed that EF-ALG8, EF-

ALG10 and EF-ALG11 are rated higher than the other algorithms in Xydeas & Petrovic quality metric. This result indicates that these algorithms are the most successful ones among the EF algorithms.

6.2 Future Work

Due to the time and computational cost of the subjective test, the performance of all the algorithms cannot be evaluated. To give whole comparison of the EF algorithms, more advanced and comprehensive subjective test can be performed. The results of this test will give the detailed information about the performance of the EF algorithms.

The most appropriate EF algorithm for real-time operation can be determined by comparing the computational cost of the EF algorithms. Then, the most appropriate one can be implemented on an embedded hardware. Therefore, the algorithms can be evaluated with respect to their real time processing performances.

The detailed study in HDRI algorithms can be performed. The objective and subjective tests can be repeated for the HDRI algorithms. Additionally, the comparison of HDRI algorithms with EF algorithms can be performed to show the advantages and disadvantages of each method.

The scope of the thesis can be extended to cover image dynamic range enhancement in color images. The same studies can be performed to see the performance of dynamic range enhancement algorithms in color images.

REFERENCES

1. **Seung -Jun Youm,Won-Ho Cho, Ki-Sang Hong**. *"High Dynamic Range Video through Fusion of Exposure-Controlled Frames"*, Conference on Machine Vision Applications, Tsukuba Science City, Japan, May 16-18, 2005.

2. **Dpreview**,http://www.dpreview.com/news/0301/03012202fujisuperccdsr.asp, last visited on 07/07/2011.

3. **Enblend/Enfuse**, http://enblend.sourceforge.net, last visited on 07/07/2011.

4. **Photomatix**, http://www.hdrsoft.com, last visited on 07/07/2011.

5. **Photoshop**, http://www.photoshop.com, last visited on 07/07/2011.

6. **Erik Reinhard, Greg Ward, Sumanta Pattanaik, Paul Debevec.** *"High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting"*, Morgan Kaufmann Publishers. December 2005.

7. **Patrick Ledda, Luis Paulo Santos, Alan Chalmers**. *"A Local Model of Eye Adaptation for High Dynamic Range Images"*, AFRIGRAPH '04 Proceedings of the 3rd international conference on Computer graphics, virtual reality, visualization and interaction in Africa, 2004.

8. **Tomoo Mitsunaga, Shree K. Nayar**. *"Radiometric self calibration"*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition Volume 1, 1999.

9. **Steve Mann, Rosalind W. Picard**. *"On Being 'Undigital' With Digital Cameras: Extending Dynamic Range by Combining Differently Exposed Pictures"*, IS&T's 48th Annual Conference; Society for Imaging Science and Technology, 1995.

10. **Paul E. Debevec, Jitendra Malik**. *"Recovering High Dynamic Range Radiance Maps from Photographs"*, SIGGRAPH '97 Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 1997.

11. **Mark A. Robertson, Sean Borman, Robert L. Stevenson**. *"Estimation-theoretic approach to dynamic range enhancement using multiple exposures"*, J. Electron. Imaging 12, 219, 2003.

12. **Bracket**. http://www.ceng.metu.edu.tr/~akyuz/bracket/bracket.html, last visited on 19/09/2011.

13. **B. Kh. Barladian, A. G. Voloboi, V. A. Galaktionov, E. A. Kopylov**. *"An Effective Tone Mapping Operator for High Dynamic Range Images"*, Programming and Computing Software Volume: 30, Issue: 5, 2004.

14. **Tsun-Hsien Wang, Wei-Su Wong, Fang-Chu Chen, Fang-Chu Chen**. *"Design and Implementation of a Real-Time Global Tone Mapping Processor for High Dynamic Range Video"*, International Conference on Image Processing - ICIP , pp. 209-212, 2007.

15. **F Drago, K Myszkowski, T Annen, N Chiba**. *"Adaptive Logarithmic Mapping For Displaying High Contrast Scenes",* The European Association for Computer Graphics 24th Annual Conference: EUROGRAPHICS, 2003.

16. **Thomas G. Stockham**. *"Image processing in the context of a visual model"*, IEEE, Volume: 60, Issue: 7, Pages: 828-842, 1972.

17. **Erik Reinhard, Michael Stark, Peter Shirley, James Ferwerda**. *"Photographic tone reproduction for digital images"*, 29th annual conference on Computer graphics and interactive techniques, 2002.

18. **K.Perlin and E.M. Hoffert**. *"Hypertexture",* Computer Graphics (Proceedings of ACMSIGGRAPH 89), ACM, 23, 253-262, 1989.

19. **Ansel Adams, Robert Hardy Baker**. *"The Negative (Ansel Adams Photography, Book 2)"*, 1995.

20. **John Erwin Tumblin**, **Jessica K. Hodgins**, **Brian K. Guenter**. *"Two Methods for Display of High Contrast Images"*, ACM Transactions on Graphics Volume 18 Issue 1, January 1999.

21. **John Erwin Tumblin**. *"Three methods of detail-preserving contrast reduction for displayed images"*, Doctoral Dissertation, 1999.

22. **Michael D. Grossberg**, **Shree K. Nayar**. *"Determining the Camera Response from Images: What Is Knowable?"*, Pattern Analysis and Machine Intelligence, IEEE Transactions, 2003.

23. **Michael Ashikhmin**. *"A Tone Mapping Algorithm for High Contrast Images"*, 13th Eurographics workshop on Rendering, 2002.

24. **Jack Tumblin, Greg Turk**. *"LCIS: A Boundary Hierarchy for Detail-Preserving Contrast Reduction"*, 26th annual conference on Computer graphics and interactive techniques, 1999.

25. **J. DiCarlo, B. Wandell**. *"Rendering High Dynamic Range Images"*, Proc. of the SPIE: Image Sensors 3965, 2000, pp. 392–401, 2000.

26. **Raanan Fattal, Dani Lischinski, Michael Werman**. *"Gradient Domain High Dynamic Range Compression"*, 29th annual conference on Computer graphics and interactive techniques, 2002.

27. **Vladimir S.Petrovic**. *"Multisensor Pixel-level Image Fusion"*, Manchester School of Engineering, February 2001.

28. **H. B. Mitchell**. *"Image Fusion: Theories, Techniques and Applications"*, 1st Edition, 200 p, 2010.

29. **Tomislav Kartalov, Ljupcho Panovski**. *"A Real Time Algorithm for Exposure Fusion of Digital Images"*, MELECON 2010 15th IEEE Mediterranean Electrotechnical Conference, 2010.

30. **A. Ardeshir Goshtasby**. ***"Fusion of multi-exposure images"***, Journal Image and Vision Computing Volume 23 Issue 6, June, 2005.

31. **Asheer Kasar Bachoo**. *"Real-time exposure fusion on a mobile computer"*, in The Twentieth Annual Symposium of the Pattern Recognition Association of South Africa (PRASA), 2009.

32. **Jinshan Tang**. *"A contrast based image fusion technique in the DCT domain"*, Digital Signal Processing Volume 14, Issue 3, Pages 218-226, May 2004.

33. **I. Zafar, E.A. Edirisinghe, H.E. Bez**. *"Multi-exposure & multi-focus image fusion in transform domain"*, Visual Information Engineering IET International Conference, Page 606, 26-28 Sept. 2006.

34. **Rafael C. Gonzalez, Richard E. Woods**. *"Digital Image Processing"*, 2nd Edition, Chapter 7 Wavelets and Multiresolution Processing, 2002.

35. **Alan C. Bovik**. *"The Essential Guide to Image Processing"*, 1st Edition, Chapter 6 Multiscale Image Decompositions and Wavelets, May 2009.

36. **Firooz Sadjadi**. *"Comparative Image Fusion Analysais"*, IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 03, 2005.

37. **Peter J. Burt, Raymond J. Kolczynski**. *"Enhanced Image Capture Through Fusion"*, Computer Vision Fourth International Conference, May 1993.

38. **Ron Rubinstein**. *"Fusion of Differently Exposed Images"*, Final Project Report, October 2004.

39. **Hung-Son Le, Adi Anani, Haibo Li.** *"High dynamic range imaging through multi-resolusion spline fusion"*, Signal Processing and Its Applications, 9th International Symposium, 2007.

40. **Tom Mertens, Jan Kautz, Frank Van Reeth**. *"Exposure Fusion",* 15th Pacific Conference on Computer Graphics and Applications, 2007.

41. **Madiha Hussain Malik, S. Asif M. Gilani, Anwaar-ul-Haq**. *"Wavelet Based Exposure Fusion",* World Congress on Engineering 2008 Vol I, 2008.

42. **Jinhua Wang, Songhe Feng, Qingsong Bao**. *"Pyramidal dual-tree directional filter bank based exposure fusion for two complementary images"*, Signal Processing (ICSP), 2010 IEEE 10th International Conference, 2010.

43. **Jinhua Wang, De Xu, Congyan Lang, Bing Li**. *"Exposure Fusion Based on Shift-Invariant Discrete Wavelet Transform"*, Journal of Information Science and Engineering, 2011.

44. **Wei Zhang, Wai-Kuen Cham**. *"Gradient-directed Composition of Multi-exposure Images"*, Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference, 2010.

45. **Annamária R. Várkonyi-Kóczy, András Rövid, Szilveszter Balogh, Takeshi Hashimoto, Yoshifumi Shimodaira**. *"High Dynamic Range Image*

Based on Multiple Exposure Time Synthetization", Technology Volume: 4, Issue: 1, Pages: 5-15, 2007.

46. **Gang Yu, Xiaojun Wu, Zhan Song**. *"A Novel Multiple Exposure Merging Method for High Dynamic Range Image Generation"*,2010 2nd International Conference on Signal Processing Systems (ICSPS), 2010.

47. **Jinhua Wang, De Xu, Bing Li**. *"Exposure fusion based on steerable pyramid for displaying high dynamic range scenes"*, Optical Engineering Volume: 48, Issue: 11, Pages: 117003, 2009.

48. **Toet, J. K. IJspeert**. *"Perceptual evaluation of different image fusion schemes",* Proceedings of Signal Processing, Sensor Fusion, and Target Recognition X, Vol. 4380, 2001.

49. **A. Toet, N. Schoumans, J. Ijspeert**. *"Perceptual evaluation of different nighttime imaging modalities",* Proceedings of Fusion 2000, Paris, 2000.

50. **A. Toet, E. Franken.** *"Perceptual evaluation of different image fusion schemes",* The International Society for Optical Engineering, Volume: 24, Pages: 25-37, 2003.

51. **Jitendra R. Raol**. *"Multi-Sensor Data Fusion with MATLAB"*, Chapter 10 Pixel and Feature Level Image Fusion Concepts and Algorithms, December 2009.

52. **Wikipedia.**http://en.wikipedia.org/wiki/Entropy_%28information_theory%29, last visited on 15/08/2011.

53. **Zhou Wang, Alan C. Bovik**. *"A Universal Image Quality Index"*, IIIE Signal Processing Letters. Y, Vol. XX, March 2002.

54. **V. Petrovic, C. S. Xydeas.** *"On the effects of sensor noise in pixel-level image fusion performance"*, Third International Conference on Information Fusion, July 2000.

55. **V. Petrovic, C. S. Xydeas.** *"Sensor noise effects on signal-level image fusion performance"*, Information Fusion, September 2003.

56. **S. T. Li, B. Yang**. *"Multifocus image fusion using region segmentation and spatial frequency"*, Image Vis. Comput., 2008.

57. **Guihong Qu, Dali Zhang, Pingfan Yan**. *"Information measure for performance of image fusion"*, Electronics Letters, Volume: 38, Issue: 7, Pages: 313-315, 2002.

58. **Timothy D. Dixon, Eduardo Fernández Canga, Stavri G. Nikolov, Tom Troscianko, Jan M. Noyes, C. Nishan Canagarajah, Dave R. Bull**. *"Selection of image fusion quality measures: objective, subjective, and metric assessment"*, J. Opt. Soc. Am. A 24, B125-B135, 2007.

59. **Vladimir Petrović.** *"Subjective tests for image fusion evaluation and objective metric validation"*, Information Fusion Volume 8, Issue 2, April 2007, Pages 208-216.

60. **E.F. Canga, S.G. Nikolov, C.N. Canagarajah, D.R. Bull, T.D. Dixon, J.M. Noyes, T. Troscianko**. *"Characterisation of Image Fusion Quality Metrics for Surveillance Applications over Bandlimited Channels"*, Information Fusion, 2005 8th International Conference on., 08/2005.

61. **Toet A., IJspeert J.K., Waxman A.M., Aguilar M.** *"Fusion of visible and thermal imagery improves situational awareness"*, Displays, Volume 18, Number 2, pp. 85-95(11), 30 December 1997.

62. **Raanan Fattal, Dani Lischinski, Michael Werman.** *"Gradient domain high dynamic range compression"*, SIGGRAPH '02 Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002.

63. **Frédo Durand, Julie Dorsey.** *"Fast bilateral filtering for the display of high-dynamic-range images"*. SIGGRAPH '02 Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002

64. **Qi Shan, Jiaya Jia, Brown, M.S.**. *"Globally Optimized Linear Windowed Tone Mapping"*, Visualization and Computer Graphics, IEEE Transactions, 2010.

65. **Mikko Kuhna, Mikko Nuutinen, Pirkko Oittinen.** *"Method for evaluating tone mapping operators for natural high dynamic range images"*, IS&T/SPIE Electronic Imaging, 7876-23, 2011.

66. **Martin Cadík, Michael Wimmer, Laszlo Neumann, Alessandro Artusi.** *"Evaluation of HDR tone mapping methods using essential perceptual*

attributes", COMPUTERS & GRAPHICS-UK, 32 (3). pp. 330-349. ISSN 0097-8493, 2008.

67. **Ahmet Oğuz Akyüz, Erik Reinhard.** *"Perceptual evaluation of tone-reproduction operators using the Cornsweet--Craik--O'Brien illusion"*, ACM Transactions on Applied Perception (TAP) TAP Homepagearchive Volume 4 Issue 4, January 2008.

APPENDIX A

IMAGE SETS

A.1   Image Sets in Objective Evaluation

Figure A.1 Image Set 1 in Objective Evaluation[1024x681][1/30, 1/125, 1/500]

Figure A.2 Image Set 2 in Objective Evaluation[1024x683][1, 1/4, 1/15]

Figure A.3 Image Set 3 in Objective Evaluation[1024x768][3.2, 0.8, 1/4]

Figure A.4 Image Set 4 in Objective Evaluation[870x552][1/13, 1/30, 1/80]

Figure A.5 Image Set 5 in Objective Evaluation[640x480][1/13, 1/30, 1/80]

Figure A.6 Image Set 6 in Objective Evaluation[816x612][1/100, 1/400, 1/1000]
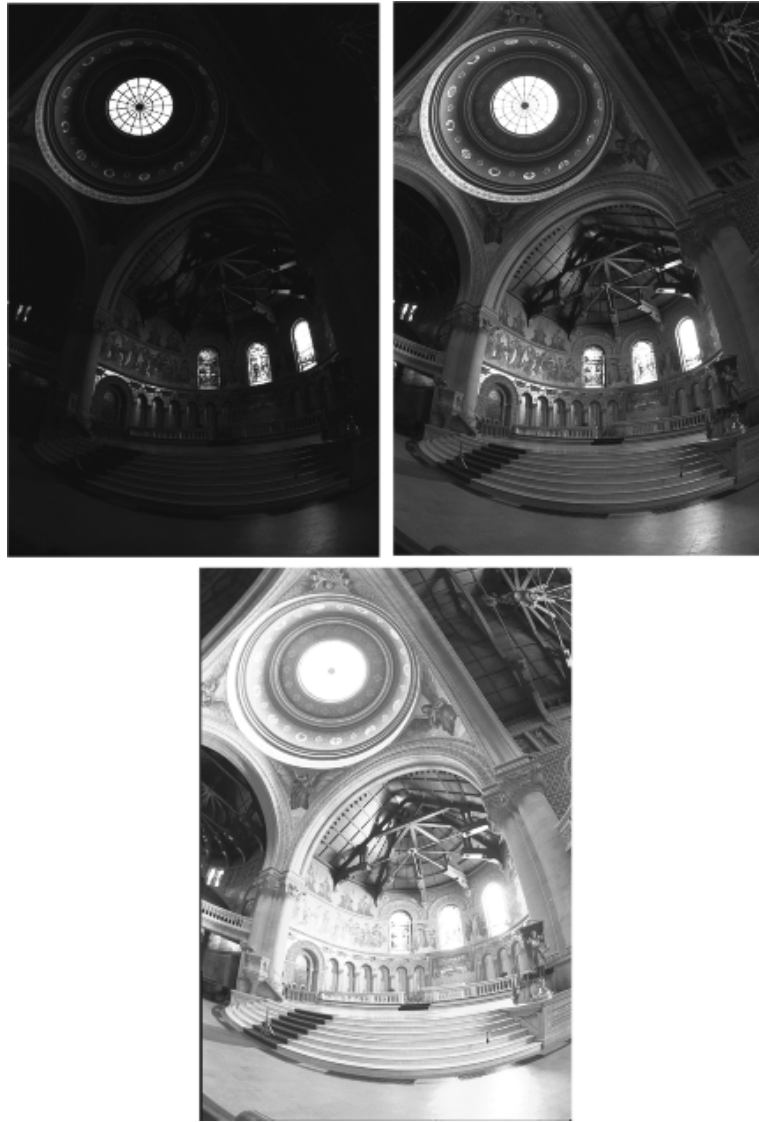
Figure A.7 Image Set 7 in Objective Evaluation[512x768][4, 1/2, 1/16]

Figure A.8 Image Set 8 in Objective Evaluation[1024x768][6, 1/2, 1/40]

Figure A.9 Image Set 9 in Objective Evaluation[819x614][0.6, 1/6, 1/25]

Figure A.10 Image Set 10 in Objective Evaluation[730x548][1/2, 1/10, 1/80]

A.2   Image Sets in Subjective Evaluation

Figure A.11 Image Set 1 in Subjective Evaluation[1024x768]

Figure A.12 Image Set 2 in Subjective Evaluation[1024x768]

Figure A.13 Image Set 3 in Subjective Evaluation[1024x768]

Figure A.14 Image Set 4 in Subjective Evaluation[1024x768]

Figure A.15 Image Set 5 in Subjective Evaluation[1024x768]

APPENDIX B

PARTICIPANT'S INSTRUCTIONS

In this test, the same steps have been followed to provide consistency. These steps are given as follows:

1. At the starting of the each subjective test, the lighting condition has been checked, and the properties of the notebook have been set to the same values.

2. Definition of the dynamic range has been given to the participants. The dynamic range has been defined as the ratio of radiances between the brightest and the darkest points in the scene.

3. Definition of the radiance has been given to the participants. In short, it has been defined as the amount of the light that is emitted from a surface or object.

4. The necessity of the image dynamic range enhancement has been explained by showing the example image to the participants.

5. The participants have been asked to score the images by considering two performance measures.

6. The first measure has been defined as the capability of recovering the details in saturated regions. For this purpose, the participants have been asked to evaluate

the outputs of the algorithms by considering the transferred details from the input images to the output image.

7. The second measure has been defined as the contrast of output images. The participants have been asked to evaluate outputs of the algorithms by considering their contrast.

8. At the last step, the participants have been asked to score the images without any breaks.