MULTI-EXPOSURE IMAGE FUSION ALGORITHMS FOR HIGH DYNAMIC RANGE IMAGING

OĞUZHAN ULUCAN

Master's Thesis

Graduate School

Izmir University of Economics

İzmir

2020

MULTI-EXPOSURE IMAGE FUSION ALGORITHMS FOR HIGH DYNAMIC RANGE IMAGING

OĞUZHAN ULUCAN

A Thesis Submitted to

The Graduate School of Izmir University of Economics

Master's Program in Electrical and Electronics Engineering.

İzmir

2020

ABSTRACT

MULTI-EXPOSURE IMAGE FUSION ALGORITHMS FOR HIGH DYNAMIC RANGE IMAGING

Ulucan, Oğuzhan

M.Sc. in Electrical and Electronics Engineering

Advisor: Asst. Prof. Dr. Mehmet Türkan

August, 2020

High dynamic range imaging (HDRI) is a challenging technology but yet demanding for modern imaging applications. Low-cost image sensors have limited dynamic range, and it is not always possible to capture and display natural scenes with high contrast and information loss in any exposure is inevitable. Three solutions for HDRI are using expensive high dynamic range (HDR) cameras with HDR-compatible displays, tone mapping operators for low dynamic range (LDR) screens and capturing and fusing multiple exposures of the same LDR scene via image fusion algorithms. Companies that produce user grade devices prefer multi-exposure fusion (MEF) approaches to obtain HDR-like images for LDR screens due to its low cost. Hence, merging a stack of images containing different exposures of the same scene into a single informative image is an attractive research field. In this thesis, a novel, simple yet effective method is proposed for static scene MEF and state-of-the-art MEF techniques have been investigated. The developed technique is based on weight map extraction via linear embeddings (LE) and watershed masking (WSM). To the best of

available knowledge, this is the first time LE and WSM are employed in MEF. The comprehensive experimental comparisons demonstrate very strong visual and statistical results, and this approach should facilitate future MEF studies.

Keywords: Multi-exposure fusion, Linear embedding, Watershed masking, High dynamic range.

ÖZET

YÜKSEK DİNAMİK ARALIKLI GÖRÜNTÜLEME İÇİN ÇOKLU POZLAMAYLA GÖRÜNTÜ BİRLEŞTİRME ALGORİTMALARI

Ulucan, Oğuzhan

Elektrik ve Elektronik Mühendisliği Yüksek Lisans Programı

Tez Danışmanı: Dr. Öğr. Üyesi Mehmet Türkan

Ağustos, 2020

Yüksek dinamik aralıklı görüntüleme (YDAG) zorlu bir teknolojidir, ancak modern görüntüleme uygulamaları için gereklidir. Düşük maliyetli görüntüleme sensörleri sınırlı dinamik aralığa sahiptir ve yüksek zıtlıktaki doğal sahneleri yakalayıp ekrana yansıtmak her zaman mümkün değildir. Ayrıca herhangi bir pozda bilgi kaybetmek kaçınılmazdır. YDAG için üç çözüm yüksek dinamik aralığa (YDA) sahip pahalı kameralar ile YDA ile uyumlu monitörler kullanmak, düşük dinamik aralığa (DDA) sahip ekranlar için ton haritalama operatörleri kullanmak ve resim birleştirme algoritmalarıyla aynı DDA sahneye ait birden fazla pozu çekip birleştirmektir. Kullanıcı sınıfı cihazlar üreten şirketler, düşük maliyetleri nedeniyle DDA ekranlar için YDA benzeri görüntüler elde etmek adına çoklu pozlama füzyonu (ÇPF) yaklaşımlarını tercih etmektedir. Bu nedenle, aynı sahnenin farklı pozlarını içeren bir görüntü yığınını tek bir bilgilendirici görüntü oluşacak şekilde birleştirmek çekici bir araştırma alanıdır. Bu tezde statik sahne ÇPF için özgün, basit ama etkili bir yöntem önerilmiştir ve güncel ÇPF metotları incelenmiştir. Geliştirilen teknik, doğrusal

gömme (DG) ve nehir sınırı maskelemesi (NSM) yoluyla ağırlık haritasının çıkarılmasına dayanmaktadır. DG ve NSM bilindiği kadarıyla ÇPF için ilk kez kullanılmıştır. Kapsamlı deneysel karşılaştırmalar çok güçlü görsel ve istatistiksel sonuçlar göstermektedir ve bu yaklaşım gelecekteki MEF çalışmalarına yardım sağlayacaktır.

Anahtar Kelimeler: Çoklu pozlama füzyonu, Doğrusal gömme, Nehir sınırı maskesi, Yüksek dinamik aralık.

To my family…

ACKNOWLEDGEMENTS

I wish to express my deepest thanksgiving to my supervisor, Asst. Prof. Dr. Mehmet TÜRKAN for his tremendous assistance, supreme supervision and suggestions throughout my graduate studies. I feel very fortunate to become his student not only in my undergraduate years but also in my graduate years thanks to his light which enlighten my path to achieve my goal. He became not only a teacher but also, he became a role model for me.

Furthermore, I wish to express my deepest love to Diclehan KARAKAYA for walking with me in this difficult and exhausting road, for her continuous support, love and friendship through these years. Her motivation helped me to stay focused on my goal in this challenging road.

I also want to thank my parents Neşe Nevin ULUCAN, Osman ULUCAN, and my sister Melis YÜCEL for their love, continuous encouragement, understanding and moral support.

Finally, I would like to thank the jury members Prof. Dr. Cüneyt GÜZELİŞ and Prof. Dr. Türker İNCE for their constructive comments.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

We are living in the era of digital immortality; we capture every moment we love by using a camera. However, if scenes have high contrast which is the difference between intensities of the darkest and the brightest pixels present in an image, these moments cannot be saved as we remembered. The vividness of the colors might be damaged, or unwanted bright or dark areas can occur in images. Since, the contrast ratio of the human visual system (HVS) is around 16,000:1, it can capture all the high-level and low-level details, with vivid colors without any problem, but most of the user-grade devices cannot capture or demonstrate the images as we see (Hoefflinger, 2007). In order to deduce the reasons behind the information loss emerging in images, the concept of dynamic range should be scrutinized. The dynamic range has several distinct definitions for the HVS and imaging equipment (Eilertsen, 2018). While for cameras it is defined as the saturation-to-noise ratio and is dependent on the imaging sensor capacity, for screens it is the ratio between the darkest and brightest pixel values. On the other hand, for the HVS, dynamic range is the ratio between the smallest and largest perceptible luminance of the scene. Furthermore, dynamic range can be categorized as LDR and HDR. LDR images are 8 bit/color channel (255:1) and previously images containing more than 24 bit/pixel were considered as HDR images (Guan and Qiu, 2007). However, with the improvements in imaging technologies, currently 32 bit/color channel images are considered as HDR photographs. Important note here that, 96 bit/pixel is a high ratio for equipment but compared to outdoor sunlight or human eye it is significantly low (Mccollough, 2008). A detailed comparison for dynamic range is provided in Table 1.

Although in the near future HDR compatible equipment will become widely available to consumers, today most of the imaging devices are in LDR due to high economic costs. As a result of the big dynamic range gap between the standard equipment capacity and natural scenes (Table 1), the captured images through LDR cameras may contain undesirable outcomes, i.e., faulty exposures[1] might cause the

[1] While long exposures cover the details in dark regions, short exposures preserve the details in bright areas. Hence, long exposures result in over exposed and short exposures result in under exposed images (Fig.2).

Table 1. Detailed Dynamic Range Comparison (Source: Hoefflinger, 2007).

Equipment or Capture Type	Dynamic Range
Human Eye	1,000,000 : 1
Outdoor Sunlight	100,000 : 1
Computer Monitor	500 : 1
Digital Single Lens Reflex Camera	300 : 1
Compact Digital Camera	100 : 1
High Quality Matte Print	50 : 1

The three common approaches to prevent the detail loss are; *(i)* using HDR cameras with HDR-compatible displays, *(ii)* using HDR cameras to capture the scene and then applying tone mapping operators to project the image to LDR screens, *(iii)* capturing multiple exposures of the same LDR scene and then fuse them via image fusion algorithms. Although the first solution provides a simple way to overcome the problem, both the HDR cameras and the HDR screens are currently unaffordable to most of the consumers (examples of the currently available HDR compatible devices are given in Appendix A). The remaining two approaches which are based on software are more attractive to both researchers and user-grade technology manufacturers. The first method is tone mapping which is matching the real measured values in an image to the capabilities of the screens by preserving the colors and details in the image (Reinhard et al., 2002). There are two main approaches for tone mapping; *Global Tone Mapping* which is mapping each pixel's intensity and global image characteristics without considering its spatial location, and *Local Tone Mapping* which also considers the spatial location of a pixel. While local tone mapping preserves the fine details better than global tone mapping, global tone mapping is computationally efficient and tends to produce less artifacts compared to local tone mapping (Eilertsen, 2018).

There are successful implementations of tone mapping (Akyuz and Reinhard, 2006; Kiser et al., 2012; Eilertsen, Mantiuk and Unger, 2015; Kim and Lee, 2020), however several drawbacks exist such as, low subjective contrast and color saturation reduction which damage the image's visual quality (Akyuz and Reinhard, 2006; Kiser et al., 2012). Therefore, due to the high economic cost of HDR sensors and the problems regarding the tone mapping approach, companies producing user-grade devices mostly

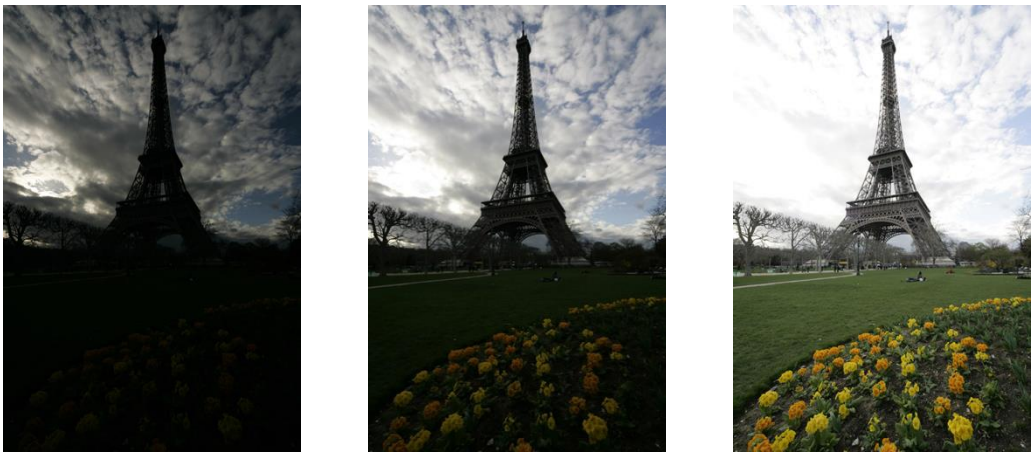
prefer the image fusion approach to obtain HDR-like images. Image fusion is a basic image processing technique that combines a stack of input images into a single output image which contains more information than any other inputs (Li, Kang and Hu, 2013). From medical imaging to military applications, image fusion algorithms are one of the well-known methods to create informative images for wide range purposes (Li et al., 2017). After the general idea of fusing a multi-exposure sequence was first introduced in (Burt and Kolczynski, 1993), image fusion has been actively used to acquire HDR-like content. There are two main approaches for image fusion; pixel-based or patch-(block-) based image fusion. In pixel-based image fusion, the output image is acquired by combining the values and intensities of the input images based on its average (Mertens, Kautz and Van Reeth, 2009; Raman and Chaudhuri, 2009). In block-wise image fusion, the features are extracted from a kernel, and then the fusion process is conducted patch-by-patch (Ma and Wang, 2015; Ma et al., 2017). Furthermore, the adoption of image fusion for multi-exposure sequences is called as MEF and it is defined as merging a stack of images containing different exposures of the same scene into a single image. These different exposures can either be static, which means the exposures can contain only very slight movement i.e. the motion of tree leaves, or dynamic where during capturing the images in a stack motion is present i.e. moving objects.

The general idea of MEF can be explained as follows. As demonstrated in Fig.1, each exposure has distinct parts which carry different information such that, the under exposed image has significant information in the sky, the flowers and the tower are visually more plausible in the over exposed image, and the normal exposed image carries small amount of general details of the image. The main aim of MEF is to keep the most informative parts, i.e. the best detailed regions, of the differently exposed images (e.g., under, normal and over) and blend them into a single informative image without damaging the fine details and color information. The acquired HDR-like image can be projected at any LDR screen without any significant information loss.

In recent decades, many successful MEF studies have been performed (Mertens, Kautz and Van Reeth, 2009; Li and Kang, 2012; Li, Kang and Hu, 2013; Ma and Wang, 2015; Paul, Sevcenco and Agathoklis, 2016; Lee, Park and Cho, 2018; Hayat and Imran, 2019). Although, the blending operation, which can be defined as the weighted sum of the input images, is nearly similar for each study, the MEF studies

mainly differ in the way of determining the fusion weight maps of different exposures. The measurement of fusion weights is indeed a challenging task and, for both static and dynamic contents, weight maps should be carefully considered since they enable us to keep the most informative parts of the distinct exposures. Furthermore, to eliminate artifacts such as halos in the sharp changeovers in the reconstructed images, the jitter effect in between exposures or to prevent ghosting[2] effect in the dynamic scenes, weight maps should be thoroughly formed.

In this thesis, MEF techniques are investigated and a simple yet effective MEF method based on LE (Roweis and Saul, 2000) and WSM (Beucher and Lantuejoul, 1979) is developed for static scenes. While initial fusion maps are extracted through LE of image pixel/patch spaces, a WSM procedure is used for adjusting these maps to refine informative parts of images for the final fusion step. Since, visual quality assessment is subjective, to present objective results the performance of the proposed approach is compared with the well-known MEF algorithms using the perceptual quality assessment method introduced in (Ma, Zeng and Wang, 2015). This thesis presents promising visual and statistical experimental results and it will enlighten the path of the future work and open new doors in this research domain.



a) Under Exposed b) Normal exposed c) Over exposed

Figure 1. Different Exposures of the *Tower* Sequence (Source: Ma and Wang, 2015).

[2] If during the capture of distinct exposures camera movement occurs or moving objects are present ghosting emerges (Wu et al., 2010).

The rest of the thesis is organized as follows. The following chapter of this thesis explains the well-known MEF studies, and gives information related to their advantages and disadvantages. The proposed novel MEF algorithm is explained in detail in Chapter 3. The dataset, which is used in this study, the experimental setup, the perceptual quality assessment method, and obtained experimental results are given comprehensively in Chapter 4. Finally, a conclusion of the thesis with a brief summary and possible future directions is given in Chapter 5.

CHAPTER 2: LITERATURE REVIEW

In recent decades, many successful MEF studies have been performed, e.g., (Mertens, Kautz and Van Reeth, 2009; Li and Kang, 2012; Li, Kang and Hu, 2013; Ma and Wang, 2015; Paul, Sevcenco and Agathoklis, 2016; Lee, Park and Cho, 2018; Hayat and Imran, 2019). Since in this thesis traditional processing techniques are adopted for weight estimation, the comprehensive literature review focuses on the studies based on classical methods but also provides additional examples of weight estimation based on neural networks. Furthermore, only the milestone MEF methods in the literature are explained in detail and used for the comparison with the proposed method.

2.1 Mertens' Method: Exposure Fusion

Although, Mertens' method (Mertens, Kautz and Van Reeth, 2009) is one of the oldest MEF algorithm, the method still gives superior results compared to most of the remainder approaches. This pixel-based MEF approach is designed for the static scenes and does not require any post-processing or tone mapping after the obtainment of the fused image which is the main advantage of the study. Because of this advantage, most of the studies available today follow the weight map extraction scheme of this method.

As aforementioned, all MEF studies differ in the extraction of the weight maps. The main reason of calculating the weight maps is to keep the best parts in the stacks at the fusion stage. For instance, in input stacks, there might be some colorless or flat regions due to the over or under exposure. Hence, while to these non-informative parts less weights are assigned, detailed regions containing vivid colors with fine details should be assigned with higher weights. The extraction of distinct weight maps can carry out this assignment and Mertens introduced three different weights; contrast (C), saturation (S) and well-exposedness (WE) which carry different information and are called as "*quality measures*".

Covering the texture information is one of the challenging tasks in the MEF studies. Therefore, to keep the fine details such as textures or edges, Mertens calculates C by simply applying a Laplacian filter to the gray scaled version of each image in the stack.

During the calculation of *C*, while edges and textures take high importance, flat regions take less.

Another challenge in MEF methods is recovering the color information, since the colors in the over or under exposed images is damaged due to clipping. To acquire a more visually appealing fused image and to solve the toneless color problem, *S* is calculated via taking the standard deviation of each channel in *RGB* domain.

The final weight *WE* is calculated to keep the most informative intensities which are nonzero in both under and over exposed images in the stack. To do so, each intensity value is weighted according to its closeness to 0.5 via Gaussian curve. This Gaussian curve is applied to each channel separately then, *WE* is calculated by multiplication of resulted maps. To combine the information and to obtain a final scaler weight map as in Eqn. 1, each quality measure is multiplied and normalized so that they sum to one at each pixel.

$$W_{i,j,n} = C_{i,j,n} \times S_{i,j,n} \times WE_{i,j,n} \tag{1}$$

where, i, j is the corresponding pixel location on n^{th} input image (*I*) in image stack $I_n = \{I_1, I_2, \dots, I_n, \dots, I_N\}$.

As a general fusion approach, the weighted blending of input images is calculated as follows,

$$F_{i,j} = \sum_{n=1}^{N} W'_{i,j,n}\, I_{i,j,n} \tag{2}$$

where, $W'_{i,j,n} = \left[\sum_{n=1}^{N} W_{i,j,n}\right]^{-1} W_{i,j,n}$.

However, most of the time, this approach can produce undesirable outputs which contain disturbing seam effects or halo effects. To avoid this problem, Mertens adapted the pyramidal image decomposition technique into his approach. To blend the images, the Gaussian pyramid of the weight maps and Laplacian pyramid of the input images are taken. Then, fusion is carried out at each pyramid level (ℓ) separately, and the fused image is obtained by collapsing each ℓ of the fused image as follows,

$$L\{F\}^{\ell}{}_{i,j} = \sum_{n=1}^{N} \mathcal{G}\{W'\}^{\ell}{}_{i,j,n}\, L\{I\}^{\ell}{}_{i,j,n} \tag{3}$$

where, $\mathcal{G}$ and L denotes the Gaussian and Laplacian pyramids, respectively.

2.2 Raman's Method: Bilateral Filter Based Compositing for Variable Exposure Photography

In (Raman and Chaudhuri, 2009), the main aim is fusing the exposures while preserving the details in both over and under exposed images. The approach is based on an edge preserving filter which is known as the bilateral filter (Tomasi and Manduchi, 1998). As Mertens' approach, this method is also designed for static scenes. As mentioned before, preserving the texture of the scene is challenging and Raman aimed to solve this problem via a simple approach. The main goal of the study is to keep strong textures in the images while eliminating the weak ones. Therefore, the bilateral filter can be used effectively, since it assigns high weights to the strong regions (edges or textures) and smoothens non-informative weak regions. Thus, Raman created a function based on bilateral filter which detects both regions and give weights according to their importance. Furthermore, a sum-to-one constraint is applied to form each exposure's weight map as follows,

$$BF_{i,j,k} = \frac{(\tau + \left|I_{i,j,n} - I_{i,j,n}^{BF}\right|)}{\sum_{k=1}^{N}(\tau + \left|I_{i,j,n} - I_{i,j,n}^{BF}\right|)} \tag{4}$$

where, $I_{i,j,n}^{BF}$ is the filtered input images, and τ is a constant value which is selected as 70 during the study, and $\sum_{n=1}^{N} BF_{i,j,n} = 1$.

After finding the weight maps, the fused image is obtained via Eqn. 2. Although this method is simple and cost effective, Raman's approach presents statistically weak results when the image sequence contains multiple over or under exposed images. Moreover, since the algorithm focuses on covering the details, it produces toneless colored fused images.

2.3 Gu's Method: Gradient Filed Multi-Exposure Images Fusion for Dynamic Range Visualization

In (Gu et al., 2012), an approach which tolerates a little movement for dynamic image fusion yet usually results in ghosting effects, is proposed to fuse static images. The proposed method is based on the gradient field modification and the distance of intensities in feature space which is calculated via Euclidean metric.

This study demonstrates weak results in fusing LDR images. Its performance could be improved by choosing another effective metric based on HVS models. Moreover, since the method mostly focuses on detail enhancement, the algorithm fails when preserving the colors and causes artificial looking, low-light results.

2.4 Li's Method (Li12): Detail-Enhanced Exposure Fusion

In (Li, Zheng and Rahardja, 2012), a quadratic optimization-based MEF method extracts the fine details from each LDR image in the stack then fuses them by using Mertens' approach to create sharper edged HDR-like content. An initial fused image is obtained via Mertens' algorithm. Then, this initial image is enhanced with the vector of fine details which is formed from the gradients of each exposures' luma component. The largest absolute gradient value is used to find the most detailed position among other exposures but since this maximum value is commonly noisy, the desired vector field is adjusted as follows,

$$v_q = \frac{\sum_{n=1}^{N} W_{n,q} \nabla \log(\mathrm{Y_{n,q}})}{\sum_{n=1}^{N} W_{n,q}} \tag{5}$$

where, $\mathrm{Y_n}$ is the luma component of n^{th} image in the sequence, W is the weight of the gradient vector, and q = 1,2.

Then the fine details are extracted by solving a quadratic optimization problem given in Eqn. 6,

$$\min_{\mathcal{L}_d} \left[\|\mathcal{L}_d\|^2 + \left(\left\| \frac{v_1 - \frac{\partial \mathcal{L}_d}{\partial x}}{\sqrt{|v_1|^{1.2} + 10^{-4}}} \right\|_2^2 + \left\| \frac{v_2 - \frac{\partial \mathcal{L}_d}{\partial y}}{\sqrt{|v_2|^{1.2} + 10^{-4}}} \right\|_2^2 \right) \right] \tag{6}$$

where, $\mathcal{L}_d$ is vector containing fine details.

Finally, the initial fused image which is acquired via the method in (Mertens, Kautz and Van Reeth, 2009) is combined with the fine details obtained in Eqn. 6 to create sharper edged HDR-like images as follows,

$$F = F_{int} \exp(\mathcal{L}_d) \tag{7}$$

where, F_{int} is the initial fused image.

The main advantage of this study is that the method creates sharper looking images whose details are preserved while covering the colors. Since the initial fused images are based on the successful method of Mertens, it is not surprising that visually appealing images are obtained. However, in some cases artificial looking images are acquired because, the fine details significantly surpass the color information.

2.5 S. Li's Method: Fast Multi-exposure Image Fusion with Median Filter and Recursive Filter

The study (Li and Kang, 2012) which can fuse both static and dynamic scenes is carried out successfully. The following quality measurements are extracted: local contrast and brightness for static images, and color dissimilarity weight for dynamic images. The main advantage of this study its fastness and ability to preserve details in most of the images in a stack. Moreover, the method can be used for infrared image fusion. On the other hand, while eliminating the motion in images, artifacts occur which affect the color of the image and the statistical results. The reason behind this drawback is, since the approach takes *RGB* channels separately, the usage of the color information and preserving the correlation between the channels is hard to accomplish.

To preserve the fine details in each image in the sequence, local contrast is first calculated via the convolution of the gray image and high pass filter, then winner-take-all approach is applied to obtain a local contrast weight map ($\mathcal{A}$) as follows,

$$\mathcal{A}_{i,j,n} = I^{g}_{i,j,n} * h_{i,j}$$
$$\mathcal{A}'_{i,j,n} = \begin{cases} 1, & A_{i,j,n} = \max\{A_{i,j,n}, \quad n = 1,2,\dots.N\} \\ 0, & else \end{cases} \tag{8}$$

where, I^g is the gray image of an input image, h is high pass filter and * denotes the convolution operation.

To mainly cover the colors of the images in the stack, brightness of each pixel is determined to decide if a pixel is too dark or too bright. By using the hat function which is often employed in MEF methods to eliminate the negative effects of under or over exposed pixels, the created brightness weight map ($\mathfrak{B}$) is constructed by only pixels which are neither under nor over exposed. The hat function takes the value of 1 if the intensity value is in between the selected threshold value (β) or 0 if not. $\mathfrak{B}$ is acquired as follows,

$$\mathfrak{B}_{i,j,n} = \begin{cases} 1, & \beta < I^g_{i,j,n} < 255 - \beta \\ 0, & else \end{cases} \tag{9}$$

where, β is in between 10-30 during this study.

To prevent the ghosting effect in dynamic scenes, firstly histogram equalization[3] is performed onto the input images. Then, median filtering is applied to each image since it is widely used in motion detection methods (Li, Yu and Yang, 2007). Finally, the color dissimilarity weight map ($\mathcal{C}$) is calculated and adjusted by morphological operators to remove noise as follows,

$$\mathcal{C}_{i,j,n} = \left(\exp\left(\frac{\left(I^E_{i,j,n} - I^M_{i,j,n}\right)^2}{\delta^2} \right) \oplus s_1 \right) \ominus s_2 \tag{10}$$

where, I^E is the histogram equalized input image, I^M is the filtered input image, δ equals to 0.1 for controlling the Gaussian curve, s_1 and s_2 are *disk*-like structuring elements, $\oplus$ and $\ominus$ denote dilation and erosion operations, respectively.

After finding each weight map, the weight feature for the fusion stage is computed via multiplying each weight map. If the scene is static the $\mathcal{C}$ is not considered in the computations below. But if the scenes are dynamic then firstly, brightness and color dissimilarity features are multiplied to prevent weighting the same pixels of different

[3] Histogram equalization is a technique to adjust an image's intensity values (Gonzalez, Woods and Eddins, 2004).

exposures with motion as follows,

$$\mathcal{D}_n = \mathcal{C}_n \times \mathfrak{B}_n$$
$$\mathcal{D}'_{i,j,n} = \left[\sum_{n=1}^{N} \mathcal{D}_{i,j,n}\right]^{-1} \mathcal{D}_{i,j,n}\,. \tag{11}$$

Some of the pixels at the same location in the input stack may be labelled as both over and under exposed. To prevent this problem, only the pixels under an average score $(1/N)$ are labeled as zero (Eqn.12),

$$\mathcal{D}'_{i,j,n} = \begin{cases} 0, & \mathcal{D}'_{i,j,n} < 1/N \\ 1, & else \end{cases} \tag{12}$$

where, *N* is the number of images in the stack, and when $\mathcal{D}'_{i,j,n} = 0$, that pixel is from either under exposed or over exposed or from moving objects in n^{th} image in stack.

After preventing this problem, the final weight is obtained as follows,

$$W'_n = \mathcal{D}'_n \times \mathcal{A}_n\,. \tag{13}$$

However, the calculations above results in noisy and sharp images since they are formed only by 1s and 0s. That is why, the edge-preserving recursive filter is used to refine the final weight map. Then fusion is conducted via the general fusion approach (Eqn. 2).

2.6 S. Li' Method (Li13): Image Fusion with Guided Filtering

In another successful work of S. Li (Li, Kang and Hu, 2013), a novel method is designed for not only MEF but also for multi-spectral, multi-focus and multi-modal fusion which is the main advantage of the study. The research can be divided into two stages, *(i)* weight map extraction and adjustment through guided filtering, *(ii)* fusion of adjusted weight maps and layers, i.e. base layer and detail layer, via weighted average method. Two-scale decomposition is used in order to separate each source image into base layer and detail layer which contain the intensity and fine details, respectively. The base layer and detail layer of each exposure in the image stack is

obtained as,

$$B_n = I_n * Z$$
$$D_n = I_n - B_n \quad (14)$$

where, *B* is the base layer, Z is the average filter which has a size of *31 x 31*, and *D* is the detail layer.

After the obtainment of two layers, the weight maps are formed from the pixel saliency which demonstrates the pixel's unique quality. Then, they are adjusted via a guided filter which is an edge-preserving smoothing filter. The Laplacian filter is firstly applied to each input image in the stack, then the local means of absolute value is convolved with the Gaussian low-pass filter to form saliency maps of each exposure (*SM*) (Eqn. 15), and lastly weight maps are obtained via Eqn. 16.

$$SM_n = |I_n * L| * \mathcal{G}_{r,\sigma} \quad (15)$$

$$W_n = \begin{cases} 1, & SM_n = \max\limits_{1 \le n \le N} \{ SM_n \} \\ 0, & else \end{cases} \quad (16)$$

where, L is Laplacian filter, W is the weight map, and $\mathcal{G}$ is Gaussian filter with parameters size r, and σ is set to *0.5*.

Since, each *W* of exposures is mostly noisy which causes artifacts in the fusion stage, the problem is solved by using spatial consistency. If two neighbor pixels' information such as brightness or color are close to each other, they are assigned similar weights. To carry out this assignment, instead of solving the problem via optimization-based methods, a guided filtering is applied to each *W* to obtain the final fusion weight maps of both layers. In this approach, the pixel's local variance is investigated. If the local variance is small, then the pixel is assumed to be in a flat region but if the local variance is large that means the pixel is on an edge. By this approach, the pixels with similar values take similar weights. After obtainment of the weight maps, fusion is carried out as follows,

$$F = \sum_{n=1}^{N} [(W_n^B \times B_n) + (W_n^D \times D_n)] \tag{17}$$

where, W^B and W^D are the filtered weight maps of the base layer and detail layer, respectively.

2.7 Paul's Method: Multi-Exposure and Multi-Focus Image Fusion in Gradient Domain

In (Paul, Sevcenco and Agathoklis, 2016), a novel approach is introduced for both MEF and multi-focus image fusion purposes in *YCbCr* color space. There are two main ideas behind taking this spatial domain for the image fusion process. The first one is that the HVS is more sensitive to the luminance channel than the chrominance channels. The second reason of selecting the *Y* channel is that, as proposed in Li12's study, the gradient magnitude of this channel in over or under exposed images is much lower than any image with better exposure in the image stack which implies that by taking the maximum magnitude of gradient at each pixel location, visually appealing images can be obtained. Therefore, the weight extraction and the reconstruction of each channel is conducted separately. Subsequently, fusion is conducted by combining each reconstructed channel.

As aforementioned, the reconstruction process of the *Y* channel is carried out by using the gradient information of an image in stack. After finding the gradient components in both directions, the magnitude is calculated to find the fused luminance gradient as follows,

$$\begin{gathered} \phi_{i,j,n}^{q} = I_{i+1,j,n} - I_{i,j,n} \\ z_{i,j} = \max_{1 \le n \le N} \sqrt{(\phi_{i,j,n}^{2})^2 + (\phi_{i,j,n}^{2})^2} \\ \phi_{i,j}^{q} = \phi_{z_{i,j}}^{q} \end{gathered} \tag{18}$$

where, $\phi_{i,j,n}^{g}$ is the gradient of the n^{th} images' *Y* channel in *q*-direction, z is the maximum gradient magnitude at pixel location, and $\phi_{i,j}^{g}$ is the fused luminance gradient.

Then by using the gradient reconstruction technique inspired from (Sevcenco, Hampton and Agathoklis, 2015), the *Y* channel is reconstructed. Afterwards, to prevent intensity clipping problems in the reconstructed *Y* channel, a post-processing step is performed via nonlinear mapping to ensure the outcomes of the pixel intensities of the reconstructed channel lies in the required range. To finalize the reconstruction of the *Y* channel, histogram equalization is applied.

Finally, the *Cb* and *Cr* channels are reconstructed via a straightforward pixel-based approach as follows,

$$Cb_{i,j} = \sum_{n=1}^{N} wb_{i,j,n} \times \left(Cb_{i,j,n} - 128\right) + 128$$
$$Cr_{i,j} = \sum_{n=1}^{N} wr_{i,j,n} \times \left(Cr_{i,j,n} - 128\right) + 128 \quad (19)$$

where, $wb_{i,j,n} = \frac{|Cb_{i,j,n} - 128|}{\sum_{n=1}^{N} |Cb_{i,j,n} - 128|}$, and $wr_{i,j,n} = \frac{|Cr_{i,j,n} - 128|}{\sum_{n=1}^{N} |Cr_{i,j,n} - 128|}$.

Then, the fused image is obtained by combining reconstructed channels, *Y*, *Cb*, *Cr*.

2.8 Ma's Method: Robust Multi-Exposure Image Fusion: A Structural Patch Decomposition Approach

This study is an update of the patch-wise MEF method introduced in (Ma and Wang, 2015) which is only for static scenes, but the study (Ma et al., 2017) modifies the previous version of the work by adding the motion detection algorithm for dynamic scenes. Important note here that, since the dynamic cases are out-of-context for this thesis, only the part of the algorithm designed for static cases is explained in detail.

The patches at the same locations for each image in the stack are extracted without separating the color channels to avoid the problems mentioned in Gu's and S. Li's method. Then, for each patch (x_n), three components are extracted which are signal strength (*cs*), signal structure (*ss*), and local mean intensity of the patch (μ_{x_n}) (Eqn.20),

$$x_n = cs_n \cdot ss_n + \mu_{x_n} \quad (20)$$

where, $x_n = \{x_n \mid 1 \leq n \leq N\}$, N is the number of exposures in the stack,

$cs_n = \left\|x_n - \mu_{x_n}\right\|$, $ss_n = \frac{x_n - \mu_{x_n}}{\|x_n - \mu_{x_n}\|}$, and $\|.\|$ denotes the l^2 norm.

Since, this decomposition is invertible, the patches of fused image are obtained via determining each component separately and inverting the decomposition. Subsequently, these patches are combined to obtain fused image.

Signal strength is related to local contrast and since, higher contrast shows better visibility, the highest *cs* among all exposure patches is selected for reconstruction of *cs* as follows,

$$\widehat{cs} = \max_{1 \leq n \leq N} cs_n \,. \tag{21}$$

For the reconstruction of *ss*, the power weighting function is used to adjust each patch's signal structure impact as follows,

$$\widehat{ss} = \frac{\sum_{n=1}^{N}(\left\|x_n - \mu_{x_n}\right\|^{\vartheta})\, ss_n}{\sum_{n=1}^{N}(\left\|x_n - \mu_{x_n}\right\|^{\vartheta})} \tag{22}$$

where, $\vartheta \geq 0$, and larger ϑ indicates larger strength.

Similarly, μ is adjusted by a weighting function which takes μ_{x_n} and the global mean intensity value of the exposure (μ_{I_n}) as inputs. This weighting function is the updated version of the *WE* function introduced in Mertens' method, and in this study, it is used as follows,

$$LG(\mu_I, \mu_x) = \exp\left(-\frac{(\mu_I - 0.5)^2}{2\sigma_g^2} - \frac{(\mu_x - 0.5)^2}{2\sigma_l^2}\right) \tag{23}$$

where, each σ controls the Gaussian curves of μ_I and μ_{x_n}, respectively.

Then, the reconstructed mean intensity of the local patch is obtained as follows,

$$\hat{\mu} = \frac{\sum_{n=1}^{N} LG\left(\mu_{I_n}, \mu_{x_n}\right)\mu_{x_n}}{\sum_{n=1}^{N} LG(\mu_{I_n}, \mu_{x_k})} \,. \tag{24}$$

Finally, the reconstructed components are used to form the fused patches (Eqn. 25).

$$\hat{x} = \widehat{cs} \cdot \widehat{ss} + \hat{\mu} \tag{25}$$

These patches are combined through a moving window with a fixed stride. The overlapped regions of patches are averaged to form the fused image. According to the experimental results, the designed method generates detailed, high quality color MEF images with very few ghosting effects which makes this study very successful.

2.9 H. Li's Method: Multi-Exposure Fusion with CNN Features

After presenting their potential with several image processing applications, in (Li and Zhang, 2018) convolutional neural networks (CNNs) are used to extract features and weight maps for both static and dynamic scene MEF. This study not only designs a MEF algorithm but also compares several CNNs models such as, denoising, super-resolution and classification networks to analyze their performance on feature extraction for MEF purposes. According to the reported results, rather than using the third layer of the denoising and the super-resolution networks, the first layer of the classification network should be employed since it is more efficient for feature extraction due to its lower computational cost. After determining the most effective CNNs model, two different weight maps; the feature vectors related to visibility measurement (*V*) and temporal consistency (*T*) are extracted via CNNs. Weight map *V* is calculated by taking the l^1 norm of the feature vector which is extracted from convolutional layer (Eqn. 26),

$$V_{i,j,n} = \|CNN(I_n)\|_1 \tag{26}$$

where, *CNN(.)* is the pre-trained deep classification network.

T is calculated for dynamic scenes. As previously mentioned, the ghosting effect is the biggest problem for dynamic scenes. To prevent this effect, motion detection is performed via calculating the Euclidean distance between two normalized feature vectors which are extracted from CNNs. When there is a motion in the image stack, this Euclidean distance will give smaller similarity value. To map the similarity between the features into a range of [0,1], a gaussian kernel is used and *T* is obtained

(Eqn.27),

$$t_{i,j,n}^2 = \left\| \widehat{CNN}\left(I_{i,j,n}\right) - \widehat{CNN}(I_{i,j,n'}) \right\|_2^2$$
$$T_{i,j,n} = \sum_{n=1}^{N} exp \frac{-t_{i,j,n}^2}{2\sigma^2} \tag{27}$$

where, n' is another input image in the stack and σ is the standard deviation which is determined as 0.05 during this study.

Then, mask ($\mathcal{M}$) is calculated via the hat function to simply adjust the weights according to the importance of the pixels (Eqn.28).

$$\mathcal{M}_{i,j,n} = \begin{cases} 1, & \beta < I_{i,j,n} < 1-\beta \\ 0, & else \end{cases} \tag{28}$$

Lastly, the final fusion weight map is calculated as follows,

$$W_{i,j,n} = \frac{V_{i,j,n} \times T_{i,j,n} \times \mathcal{M}_{i,j,n}}{\sum_{n=1}^{N} V_{i,j,n} \times T_{i,j,n} \times \mathcal{M}_{i,j,n} + \alpha} \tag{29}$$

where, α is a small coefficient which is added to prevent division by zero.

The fused image is obtained via linear weighted combination of images in the exposure stack as in Eqn. 2.

The most attractive part of this study is that CNNs can help to determine weight maps without solving any complicated optimization problems. The designed model produces desirable MEF images with low cost. However, there are two drawbacks of this study; the method is very slow and since CNNs are like black boxes, fine tuning the CNNs' parameters is a challenging task.

All the MEF approaches which are explained in detail are well-known MEF algorithms in the literature. The common part of all the studies is the fusing approach of the input stack. However, the weight map extraction differs and to the best of available knowledge, there is no study which can extract the weight maps via LE and watershed masks. Consequently, this thesis focuses on a novel MEF approach for static

scenes by taking advantage of LE and WSM. The proposed MEF algorithm which is called MDO-MEF is explained in the next Chapter of this thesis.

CHAPTER 3: PROPOSED MEF APPROACH

As mentioned previously, MEF studies are diverse in the fusion weight extraction stage and adjusting these weight maps is a challenging procedure. The proposed approach here depends on patch-based weight estimation (per pixel) via LE of images and WSM for obtaining global fusion maps. To the best of available knowledge, the proposed MEF method is a novel framework taking the advantage of LEs of image pixel/patch spaces and WSM of images. A simple flowchart of the proposed MEF algorithm is demonstrated in Fig. 2.

The block diagram in this figure illustrates two branches for extracting linear embedding weights (LEW) and WSM from the given stack of image exposures. These maps and masks are then combined to obtain global fusion masks in order to obtain a fused image. In the following, the developed method is detailed including the determination of three main exposures; LE of pixel/patch spaces of images for extracting weight maps; watershed masking together with the reasons why it is preferred rather than a binary masking; and the blending procedure of multi-exposure images to obtain HDR-like fusion, which is later to be enhanced in a final post-processing step.

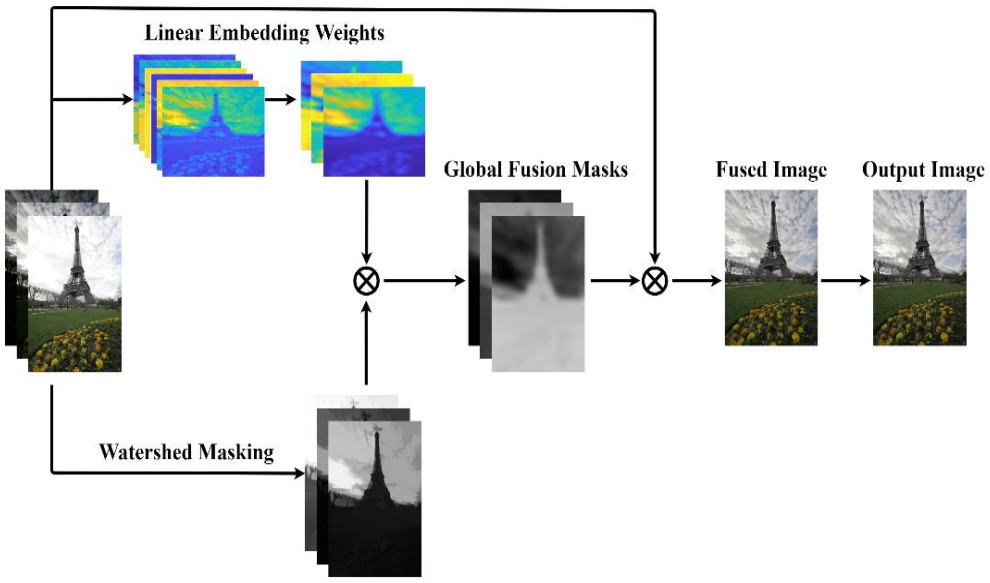


Figure 2. Proposed MDO-MEF Method's Simple Flowchart.

3.1 Determination of the Exposures

Although it can be extended to any number of exposures, the proposed MEF algorithm blends three different exposures to obtain a fused image. These exposures will be referred to as the “under exposed”, “normal exposed” and “over exposed” in remaining parts of this study. These exposures need to be carefully determined especially when the input stack has more than three images. The proposed method for determining the exposures consists of three steps. The first stage is to compute the histogram of each image in the sequence and extract the probability density function (PDF) of each histogram to form the feature vector. Secondly, a decision is made to label the images according to three clusters which are under, normal and over exposed through *k*-means clustering algorithm. Lastly, the images in the same cluster are averaged through a sliding window operation and the three exposures are obtained. In the following part of the thesis, the determination of the exposures is explained in detail. A simple flowchart of the determination of exposures is illustrated in Fig. 3 and the proposed algorithm is provided in Appendix B.

3.1.1 Histograms

Histogram are widely used in image processing and they are utilized to count the number of distinct pixel intensities. Since 8-bit images are employed in this thesis, the range of the histograms is between 0 and 255. As a straightforward approach, histograms of different exposures can be utilized effectively in separating the stack of input images into three clusters. Therefore, the PDFs of histograms are employed as features in *k*-means clustering to determine the three exposure clusters of images named as under, normal and over exposed. A PDF demonstrates the occurrence probability of a pixel in an image. The PDFs of histograms are used as features because each image in the input sequence has a unique pixel intensity distribution. For instance, while the majority of the pixels in the over exposed image in the sequence tend to have values near to 200-255, most of the pixel values in the under exposed image are around 0-50. On the other hand, the normal exposed image’s pixel intensities tend to spread nearly uniformly in the scale of 0-255. Based on these observations, all the images in the input sequence can be divided into 3 labels (under, normal, over) via a clustering algorithm.

Figure 3. Determination of the Exposures.

3.1.2 k-Means Clustering

k-means is an unsupervised clustering algorithm where *K* items are labeled into *k* number of clusters (Khan and Ahmad, 2004). The main goal is to maximize the similarities of data points in the same cluster and to minimize the resemblance of different clusters (Khan and Ahmad, 2004). In the *k*-means clustering algorithm, *K* data points are placed into clusters depending on the Euclidean distance between data points and the centroids of each cluster. Hence, data points closest with the same cluster centroid are placed into the same cluster (Jain, 2010). Since *k*-means is an unsupervised clustering method, the first assignment of the *k* data clusters, and initial centroids are decided arbitrarily. It is an iterative algorithm with the update of centroids and the data point assignments to the nearest centroids, until there is no change in data points belonging to *k* distinct clusters (Basu, Davidson and Wagstaff, 2008).

In this thesis, the feature vector based on the PDFs of the histograms are given to the *k*-means clustering algorithm. The cluster number is assigned to 3 beforehand and a condition is added to ensure that each cluster contains at least one image. Since *k*-means clustering is initialized uniformly at random from the range of the feature vectors, the algorithm runs 1000 iterations 5 times to obtain consistent results. Moreover, instead of determining each centroid based on the Euclidean distance, according to practical observations one minus the sample correlation between data points is used as distance metric.

3.1.3 Finding the Exposures via Sliding Window

After grouping the exposures in the image stack, a sliding window based averaging technique is applied in each cluster to obtain three different exposures (i.e., under, normal, over) of images to be used in the fusion process. The sliding window technique is based on a simple averaging procedure to combine images in the same cluster. Instead of a pixel-based averaging in these clusters, the averaging operation here operates on a very small neighborhood of pixel patches (e.g., 5×5 pixels) and overlaps between adjacent patches are implicitly allowed which are finally averaged uniformly in these regions. This enforces all image patches to agree on the overlapped areas, hence satisfying local compatibility and smoothness while reducing the noise and artifacts in the combined image per exposure cluster.

3.2 Weight Maps via Linear Embeddings

After the exposures are determined, the weight map extraction begins with the novel LE approach.

In image processing, there is an observation which suggests that natural images are sampled from low-dimensional manifolds. Hence, densely sampled images, or rather small texture patches, can be successfully reconstructed as a linear combination of their neighbors. This is generally referred to as neighbor embedding in image processing tasks (Chang, Yeung and Xiong, 2004; Türkan, Thoreau and Guillotel, 2012, 2013, 2014), and is inspired by the manifold learning algorithms for dimensionality reduction (Roweis and Saul, 2000; Tenenbaum, De Silva and Langford, 2000; Donoho and Grimes, 2003). In this study, the manifold sampling assumption is unified with exposure images in a given stack, which results in a new framework for weight map extraction in MEF. The general idea originates from the well-known dimensionality reduction technique called locally linear embeddings (LLE) (Roweis and Saul, 2000), based on the assumption that each exposure image is sampled from a manifold structure and all these exposures should lie on or close to a locally linear patch of the underlying sampled manifold.

The LLE algorithm aims to map J-high dimensional data to lower dimensional data space while preserving the structure of the manifold. To do so, the algorithm firstly

takes a data consisting of η real-valued column vectors $\mathscr{b}_j$ and identifies the m nearest neighbors $\mathscr{b}_{j_k}$ by using Euclidean distance, where $k = 1, \dots, m$, and $j \in J$.

Each data point $\mathscr{b}_j$ can be approximated by its neighbors $\mathscr{b}_{j_k}$ with the selection of optimal weights W_{jk}. W_{jk} can be computed by solving the least-squares problem in Eqn. 30 with a sum-to-one constraint while minimizing the reconstruction error given in Eqn. 31. Important note here that, enforcing the weights with the sum-to-one constraint leads to invariance to translations, rescalings and rotations, hence, the intrinsic geometric characteristics in each neighborhood are preserved.

$$\mathbb{E}_j(W_{jk}) = \left\| \sum_k W_{jk}(\mathscr{b}_j - \mathscr{b}_{j_k}) \right\|_2^2 \qquad s.t \qquad \sum_k W_{jk} = 1, \ \forall j \tag{30}$$

$$\mathbb{E}(W) = \sum_j \left\| \mathscr{b}_j - \sum_k W_{jk} \mathscr{b}_{j_k} \right\|_2^2 \tag{31}$$

Then, the optimum weighting coefficients are calculated as follows,

$$\mathscr{w}_j^{\boldsymbol{T}} = \frac{\left[\widehat{\mathcal{B}}_j^T \widehat{\mathcal{B}}_j\right]^{-\mathbf{1}} \mathbf{1}}{\mathbf{1}^T \left[\widehat{\mathcal{B}}_j^T \widehat{\mathcal{B}}_j\right]^{-\mathbf{1}} \mathbf{1}} \tag{32}$$

where, $\widehat{\mathcal{B}}_j = \left[\mathscr{b}_{j1} - \mathscr{b}_j \,|\dots|\, \mathscr{b}_m - \mathscr{b}_j\right]$ and $\mathscr{w}_j = [W_{j1} \,|\dots|\, W_{jk}]$.

In this thesis, the weight maps are extracted based on this approach. After determining three exposure images, namely, $\mathcal{U}$ (under exposed), $\mathcal{N}$ (normal exposed) and $\mathcal{O}$ (over exposed) via clustering and sliding window, a patch-based scheme is designed for characterizing the intrinsic properties of manifold structures of spatially local pixel/patch spaces of these exposures. While representing three exposure images by matrices **U**, **N** and **O** of sizes $r \times c$ pixels, $n \times n$ image patches are extracted with lexical ordering of image pixels, i.e., $i = 1 \dots rc$. Note that $n \times n$ patches $\boldsymbol{u}_i$, $\boldsymbol{n}_i$ and $\boldsymbol{o}_i$ extracted from **U**, **N** and **O** respectively, are centered around the pixel indexed by i and they are all collocated patches originating from each different exposure. These image patches are later transformed into the stacked column vectors of size $n^2 \times 1$. The

parameter *n* is fixed to a sufficiently small neighborhood of size 5 pixels. Let us denote these patch triplets in a set as $\boldsymbol{\mathcal{P}} = \{\boldsymbol{u}_i, \boldsymbol{n}_i, \boldsymbol{o}_i\}_{\forall i}$.

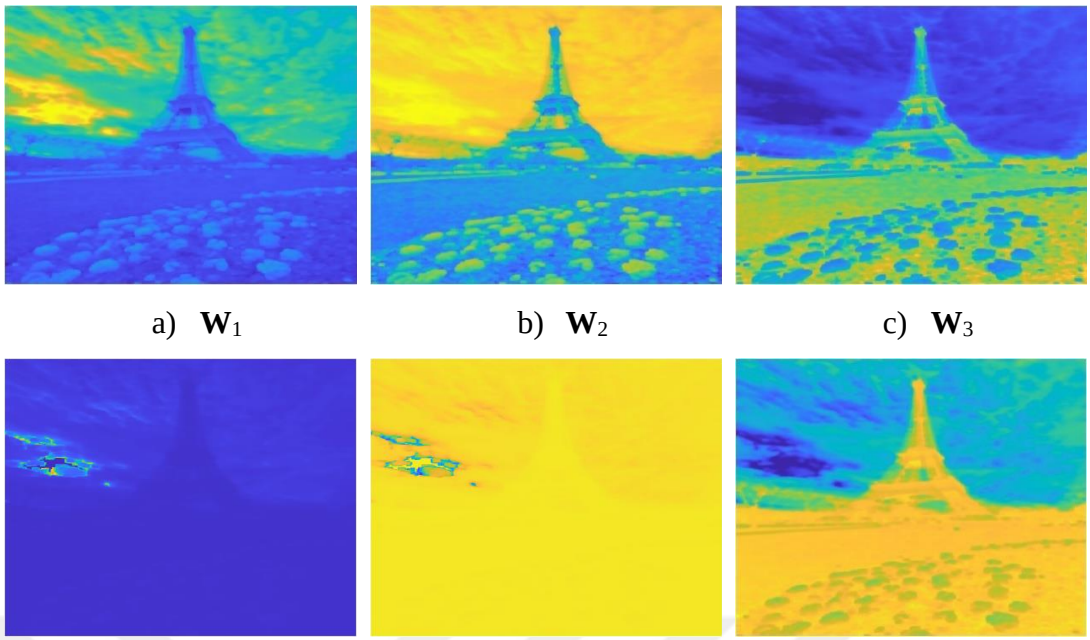
The main objective is to characterize point-based structures by means of patch manifolds through spatially collocated exposure patches. Local geometric properties of each $n \times n$ neighborhood indexed by i can be linearly characterized by solving three optimization problems given in Eqn. 33 as follows,

$$\begin{aligned}
\{\mathbf{W}_i^1, \mathbf{W}_i^2\} &= \arg\min_{\{w_1, w_2\}} \left\| \boldsymbol{o}_i - [\boldsymbol{u}_i \;\; \boldsymbol{n}_i] \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \right\|_2^2 \quad s.t \;\; w_1 + w_2 = 1 \\
\{\mathbf{W}_i^3, \mathbf{W}_i^4\} &= \arg\min_{\{w_3, w_4\}} \left\| \boldsymbol{u}_i - [\boldsymbol{n}_i \;\; \boldsymbol{o}_i] \begin{bmatrix} w_3 \\ w_4 \end{bmatrix} \right\|_2^2 \quad s.t \;\; w_3 + w_4 = 1 \\
\{\mathbf{W}_i^5, \mathbf{W}_i^6\} &= \arg\min_{\{w_5, w_6\}} \left\| \boldsymbol{n}_i - [\boldsymbol{u}_i \;\; \boldsymbol{o}_i] \begin{bmatrix} w_5 \\ w_6 \end{bmatrix} \right\|_2^2 \quad s.t \;\; w_5 + w_6 = 1
\end{aligned} \tag{33}$$

where each individual patch in the set $\boldsymbol{\mathcal{P}}$ is linearly embedded into remaining two exposures path subspaces leading to a set of weights $\{w_j\}_{j=1}^{6}$ and $\{\mathbf{W}_i^j\}_{j=1}^{6}$ denotes the set of linear embedding weight maps at each pixel location i, $\forall i$.

Note that there exists a sum-to-one constraint in each optimization in order to enforce the approximation to lie in the subspace of the patch to be embedded and also to provide invariance to translations. The optimization problems in Eqn. 33 can be easily solved by means of an inner product (Gram) matrix similar to Eqn. 32.

An example of the extracted six weight maps from *Tower* stack in given in Fig. 4. It can be clearly observed that each embedding weight map highlights specific parts of the exposures to be blend. In short, $\mathbf{W}^1$ and $\mathbf{W}^5$ originate from $\mathcal{U}$ for reconstructing $\mathcal{O}$ and $\mathcal{N}$, respectively. Similarly, $\mathbf{W}^2$ and $\mathbf{W}^3$ comes from the image $\mathcal{N}$ for $\mathcal{O}$ and $\mathcal{U}$; and $\mathbf{W}^4$ and $\mathbf{W}^6$ are extracted from $\mathcal{O}$ for $\mathcal{U}$ and $\mathcal{N}$. These weight map pairs obtained from the same exposures are later combined absolutely in Eqn. 34 to form fused linear embedding maps, namely $\mathbf{E}_1'$, $\mathbf{E}_2'$, and $\mathbf{E}_3'$ for $\mathcal{U}$, $\mathcal{N}$ and $\mathcal{O}$ images, respectively.



a) $\mathbf{W}_1$ b) $\mathbf{W}_2$ c) $\mathbf{W}_3$

d) $\mathbf{W}_4$ e) $\mathbf{W}_5$ f) $\mathbf{W}_6$

Figure 4. An Example of Extracted Weight Maps via LE.

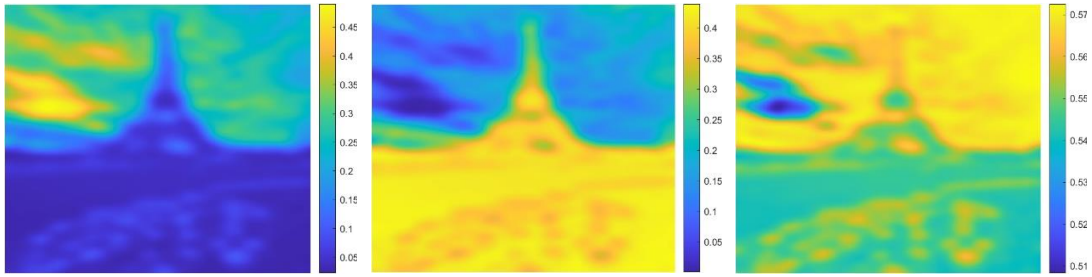
$$\begin{aligned} \mathbf{E}_1' &= |\mathbf{W}^1| + |\mathbf{W}^5| \\ \mathbf{E}_2' &= |\mathbf{W}^2| + |\mathbf{W}^3| \\ \mathbf{E}_3' &= |\mathbf{W}^4| + |\mathbf{W}^6| \end{aligned} \tag{34}$$

Since $\mathbf{E}_1'$, $\mathbf{E}_2'$, and $\mathbf{E}_3'$ are calculated from different exposures in the image stack, they are normalized to sum-to-one and then smoothed to provide local smoothness in the transition regions while avoiding possible noise and artifacts. The resulting embedding weights $\mathbf{E}^1$, $\mathbf{E}^2$, and $\mathbf{E}^3$ are obtained as follows,

$$\mathbf{E}_\mathrm{k} = \left(\mathbf{E}_\mathrm{k}' \otimes \left[\sum_{\mathrm{k}=1}^{3} \mathbf{E}_\mathrm{k}' \right]^{-1} \right) * \boldsymbol{\mathcal{G}} \tag{35}$$

where $\boldsymbol{\mathcal{G}}$ is a Gaussian smoothing kernel, $\otimes$ and $*$ denote the element-wise multiplication and the convolution operators, respectively.

Figure 5 exemplifies the obtained final linear embedding weight maps for the *Tower* stack. The procedure of acquiring the LE maps is provided as an algorithm in Appendix B.



a) $\mathbf{E}_1$ b) $\mathbf{E}_2$ c) $\mathbf{E}_3$

Figure 5. The Final Embedding Maps to be used in the Further Stages of the Algorithm.

3.3 Watershed Mask Construction based on Morphological Operations

A general strategy in MEF studies is to employ binary masks, e.g., hat function explained in Section 2.5. These masks often produce artifacts in the regions where sharp texture and color changes occur in the scene. Regular smoothing filters can be applied to avoid these artifacts but this process may cause another undesirable artifact, e.g., halo effects. Alternatively, edge-aware smoothing such as cross-bilateral filters can be employed; however, it is not trivial to control the spatial and intensity values via sufficient parameters in these types of filters.

In this thesis alongside LE weights, the weight maps acquired from morphological operations verified via watershed segmentation (Beucher and Lantuejoul, 1979) are used. WSM is adopted for mask extraction in order to acquire natural texture and color transitions while avoiding unwanted effects and artifacts in the fused image, i.e. to remove small blemishes without disturbing the overall structures. The aim is to avoid over segmentation results in the watershed transform. It is worth mentioning here that, to the best of available knowledge, it is the first time that the watershed segmentation is employed in the HDR-MEF problem.

Introduced in 1979, the watershed method is based on placing a water source and separating basins through the water flood while the image is considered as a landscape consisting of valleys and ridges. The distinct heights in an image are acquired mathematically through its gradient map. Dark regions in the image tend to have a lower height, whereas areas with brighter pixels correspond to the ridges. Each basin forms a region and the image is segmented into several regions in contrast to binary masks with two regions only.

The watershed algorithm has widely been used for segmentation tasks in various image processing problems (Chung and Khan, 2019; Khan et al., 2019; Zhang, 2019). In this work, the watershed segmentation is used to verify the effectiveness of the masks obtained from morphological operations.

Two main operations, erosion and dilation, are defined as follows,

$$I \ominus \mathcal{S} = \bigcap_{s \in \mathcal{S}} I_{-\mathcal{S}} \qquad\qquad I \oplus \mathcal{S} = \bigcup_{s \in \mathcal{S}} I_{\mathcal{S}} \tag{36}$$

where, I is the input image and $\mathcal{S}$ is the structuring element which represent the subset of pixels within the input window that we wish to consider, $\mathcal{S} = \{ s \mid s \in \mathcal{S} \}$, $\ominus$ and $\oplus$ denote erosion and dilation, respectively.

From these two fundamental spatial image processing filters, two operators used in this thesis; opening-by-reconstruction and closing-by-reconstruction, can be derived. Note here that, while opening is an erosion followed by a dilation, opening-by-reconstruction is an erosion followed by a morphological reconstruction which is proposed in (Vincent, 1993). Similarly, closing-by-reconstruction is a dilation followed by a morphological reconstruction.

The parameter selection (the shape, and neighborhood size) for opening and closing operations depends on the watershed segmentation and it is a practical process. If the regions of the created masks can be segmented via watershed segmentation, then the parameters can be used for the construction of the WSM. After this process, WSMs can be used as effective weight maps in the fusion process.

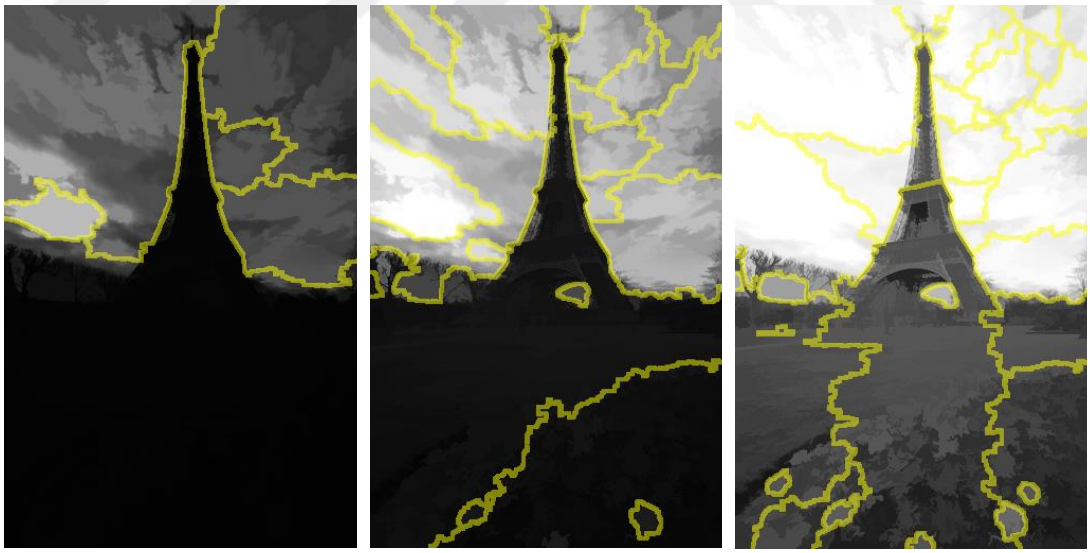
To carry out the morphological operations, the parameters of the structuring element need to be determined carefully to create efficient WSMs and this determination is based on the watershed segmentation as mentioned above. After several implementations a *disk*-shaped morphological structuring element with 11 neighborhood size is employed to perform morphological operations.

The prior process of the mask construction is the determination of the exposures, $\mathcal{U}$, $\mathcal{N}$, and $\mathcal{O}$. After finding each exposure in the sequence, the mask construction process begins with the conversion of each image to a grayscale image. After gathering the grayscale images, first the opening-by-reconstruction process is carried out.

Morphological reconstruction recovers all the details in the image, and high-intensity objects are identified while eliminating details smaller than the selected structuring element.

Afterwards, a dilation operation is performed on reconstructed image by opening with the same structuring element to expand the objects in the images by filling the object boundaries. Since, morphological operations are sensitive to the shapes, the low-level features and high-level features can be covered in the images' masks.

Lastly, the complements of the reconstructed images obtained by opening-by-reconstruction and the dilated version of it are used for the final morphological reconstruction. The complement of these maps is taken to form a watershed mask. For better understanding the outline of the algorithm to create WSMs is provided in Appendix B. As seen in Fig. 6, each region carries specific information such as clouds, sun and tower and these regions can be segmented via watershed segmentation. This concludes that, the specific parameters for morphological operations are effectively chosen.



a) $\mathbf{M_1}$ b) $\mathbf{M_2}$ c) $\mathbf{M_3}$

Figure 6. Watershed Masks of the *Tower* Stack.

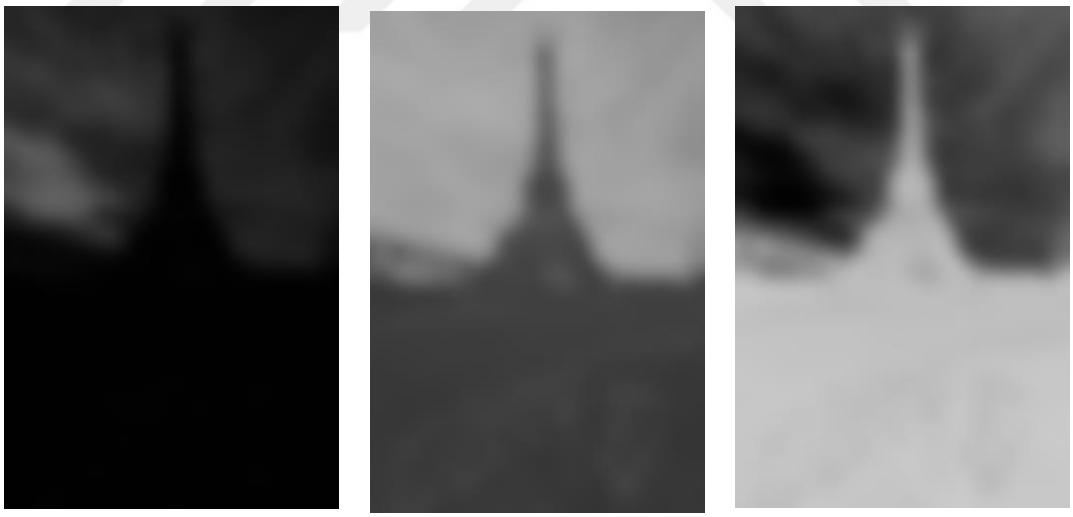
3.4 Exposure Fusion

The global fusion masks are obtained in Eqn. 37 via a linear combination of the extracted information contained in the WSMs and the LEW maps as follows,

$$\begin{aligned} \mathbf{G_1} &= \mathbf{M_1} \otimes \mathbf{E_3} \\ \mathbf{G_2} &= \mathbf{M_2} \otimes \mathbf{E_2} \\ \mathbf{G_3} &= \mathbf{M_3} \otimes \mathbf{E_1} \end{aligned} \tag{37}$$

where $\mathbf{G_1}$, $\mathbf{G_2}$, and $\mathbf{G_3}$ represent the global fusion masks for under, normal, and over exposed images, respectively.

Note here that global fusion masks are obtained in a way that the extracted LE information contained in over and under exposures are exchanged in between via the corresponding WSMs. Therefore, well-exposed regions in both under and over exposures are highlighted while normal exposure image contributes to both. These global fusion masks are demonstrated in Fig. 7 for the *Tower* stack.



a) $\mathbf{G_1}$ b) $\mathbf{G_2}$ c) $\mathbf{G_3}$

Figure 7. Global Fusion Masks of the Tower Stack, (a) for Under Exposed, (b) for Normal Exposed and (c) Over Exposed.

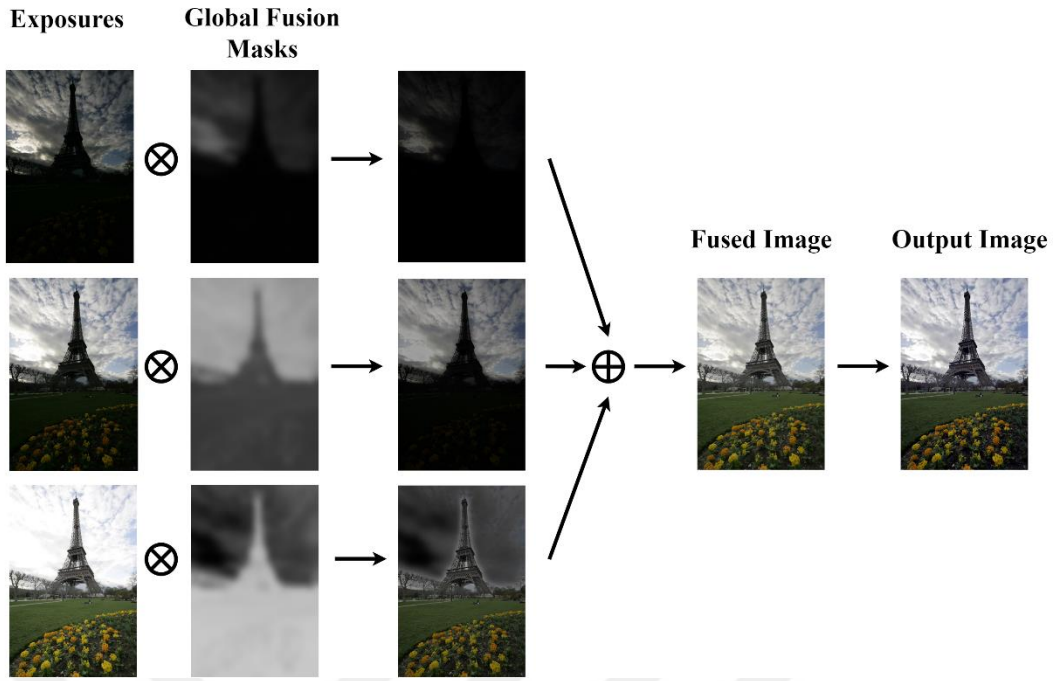


Figure 8. The Exposure Fusion Process.

The fused image **F** can be recovered through a weighted blending of input images with the corresponding global fusion masks as given in Eqn. 38 and illustrated in Fig. 8.

$$\mathbf{F} = \mathbf{U} \otimes \mathbf{G_1} + \mathbf{N} \otimes \mathbf{G_2} + \mathbf{O} \otimes \mathbf{G_3} \tag{38}$$

After obtaining the fused image **F**, a simple contrast enhancement based post-processing is employed in order to correct local low-light regions or unsatisfactory color intensities. To achieve this, the top 1% and the bottom 1% of all pixel values of the image are saturated to stretch the contrast of **F**. The final output obtained through this post-processing presents both statistically superior results and visually more plausible and natural-looking images. The algorithm for the fusion process is provided in Appendix B.

3.5 Failed Attempts During Algorithm Development

The algorithm developed in this thesis produces both statistically and visually satisfying results. However, its development was a troublesome and laborious process.

Although, while its development, several approaches resulted in unsatisfying outcomes, the experience gathered during these attempts led to form a successful novel algorithm. Therefore, in this part, the attempts which have significant impact on the development of the algorithm are mentioned.

During the determination of the exposures, firstly the images in the sequence were manually grouped and the mean of the exposures in the same group was taken to form $\mathcal{U}$, $\mathcal{N}$, and $\mathcal{O}$. However, manually separating the images in the stack caused too bright or too dark fused images. Therefore, this method was replaced with *k*-means clustering which is based on image features and results in objective outcomes. Moreover, after clustering the images, it was observed that a straightforward averaging approach results in noisy exposures. Therefore, instead of using a simple direct averaging method, a sliding window operation is employed which produces noiseless exposures.

Moreover, to produce a cost efficient LE method pixel-wise and patch-wise LE approaches were analyzed. It was observed that, weight extractions via the patch-wise LE method resulted in lower computational cost without reducing the statistical fusion outcomes. Furthermore, in order to obtain the optimal weights via LE, three different weight summation methods were analyzed which were uniformly summing (as in Eqn. 34), controlling the weights via their variance and exponential variance. According to the statistical fusion results, Eqn. 34 demonstrated the best outcomes. Thus, weight maps of the exposures are extracted via a patch-wise LE method where the LEW are formed via uniform summation.

To adjust the LE weights and give more significance to informative parts in each exposure, at first the hat function was adopted to create binary masks. However, binary masks produced artifacts and color distortions especially at edges where sharp changeovers were present. To deal with these problems, binary masks were smoothed at transition points. Even though smoothing improved the results, the halo effect problem could not be solved. Therefore, WSM, which is based on morphological operations are adopted to produce effective masks. In addition, during the creation of global fusion masks, the emergence of noise was observed. Therefore, gaussian functions are applied to $\mathbf{E}'$ to remove the noise.

CHAPTER 4: EXPERIMENTAL RESULTS

After the MEF algorithm is designed, both statistical and the visual analyses are conducted to observe the performance of the method. During these analyses, the output images of the proposed approach are compared with the fused images obtained from the MEF algorithms which are mentioned in Chapter 2.

4.1 Dataset and Experimental Setup

To perform a comprehensive analysis, 2 publicly available HDR datasets are used (Ma, Zeng and Wang, 2015; Merianos and Mitianoudis, 2019), hence two sets of experiments are conducted. While, in (Ma, Zeng and Wang, 2015), 13 different image sequences containing a diverse number of exposures are available, in (Merianos and Mitianoudis, 2019) 5 different high-resolution image stacks, each with 3 exposures are present. Although, the first dataset is adequate to analyze the proposed method's performance, the utilization reasons of the second dataset are; increasing the variety of the analysis and measuring the execution time of the algorithm in high-resolution images. The detailed information about the employed images are given in Table. 2 and all experiments are carried out on an AMD Ryzen(TM) 5 3600x CPU @ 3.80GHz 6-core 16GB RAM machine using MATLAB R2019b.

Table 2. Features of the 18 Image Stacks with Different Number of Exposures used in the Experiments.

Name	**Size**	**Name**	**Size**
Tower	*512 x 341 x 3*	*Balloons*	*339 x 512 x 9*
Venice	*341 x 512 x 3*	*Belgium*	*384 x 512 x 9*
Garden	*340 x 512 x 3*	*Desk*	*384 x 512 x 15*
Farmhouse	*341 x 512 x 3*	*M. Capitol*	*384 x 512 x 30*
Landscape	*341 x 512 x 3*	*Flowers*	*720 x 1080 x 3*
Lighthouse	*340 x 512 x 3*	*SeaRock*	*720 x 1080 x 3*
Kluki	*341 x 512 x 3*	*SecretBeach*	*720 x 1080 x 3*
Cave	*384 x 512 x 4*	*OldHouse*	*720 x 1080 x 3*
Office	*340 x 512 x 6*	*Rovinia*	*720 x 1080 x 3*

4.2 The Perceptual Quality Assessment Algorithm

The performance of the proposed MEF approach is compared statistically with other algorithms using the perceptual quality assessment method, which is a multi-scale structural similarity framework (MEF-SSIM) (Ma, Zeng and Wang, 2015). MEF-SSIM basically measures patch structural consistency for MEF and provides statistical analysis results in the range [0, 1], in which outcomes closer to 1 indicate better perceptual quality.

In order to assess the quality of the fused image, MEF-SSIM forms a multi-input (i.e., distinct exposures) structural comparison element (SCE) based on structural similarity index (SSIM) (Wang et al., 2004). While neglecting the patch luminance components because of under/over exposedness, SCE depends only on contrast and structure components of input images in as follows,

$$SCE(\{x_n\}, y) = \frac{2\,\sigma_{\hat{x}y} + \theta}{\sigma_{\hat{x}}^2 + \sigma_y^2 + \theta} \tag{39}$$

where, x_n represents collocated set of patches in all *N* images in the input stack and y is the corresponding patch in the fused image. $\hat{x} = \hat{c} \cdot \hat{s}$ denotes the desired output (fused) patch as a function of the desired contrast $\hat{c}$, i.e., the highest contrast of x_n, and the desired structure $\hat{s}$, i.e., a weighted average of the input structure vectors. $\sigma_{\hat{x}}^2$ and σ_y^2 demonstrate local variances of $\hat{x}$ and y respectively, $\sigma_{\hat{x}y}$ is the local covariance between $\hat{x}$ and y. θ is a small constant handling the low contrast saturation effects (Wang et al., 2004).

The MEF-SSIM comparison is applied on local patches across the entire image, resulting in a spatial quality map which gives an indication of the structural quality. These local values are then averaged to acquire the overall MEF-SSIM score of the fused image. The luminance consistency in the fused image is further considered with a multi-scale extension by a set of scale-level quality scores.

4.3 Comprehensive Experimental Results

In the first set of experiments, the proposed MEF algorithm is compared against eleven well-known approaches including Mertens, Gu, Li12, S.Li, Li13, Ma, H.Li,

Paul, and Raman. Important note here that, the default settings without any optimization for each algorithm are adopted for comparison and results provided for Ma, H. Li, and Paul are obtained by executing the code reached from the webpages (Ma Algorithm, 2017; Li CNN Algorithm, 2018; Paul Algorithm, 2016) In Table 2, LCE and GBE stand for two simple methods which linearly combine input exposures using local energy and global energy as weight maps, respectively. Furthermore, LEW and WSM denote linear embedding maps and watershed masking applied individually to the see their effectiveness in MEF problem as weight map, respectively.

The proposed algorithm produces highly competitive results among all MEF-SSIM scores, and it is able to outperform most of the state-of-the-art approaches on the average with given image sequences (Table 2). Moreover, LEM and WSM are employed individually in the fusion process in order to demonstrate the impact of the extracted weights in the proposed algorithm. Although these individual weights sometimes outperform their combination, the combined algorithm is clearly more effective on average and the results can be seen in Table 3.

It is also worth mentioning here that the proposed method produces visually more plausible results for several image sequences in cases when statistical results do not provide the best results reported in Table 3.

The fused images obtained for *Office, Garden* and *Venice* result in 0.991, 0.990, 0.977 MEF-SSIM scores respectively, which provide superior results when compared with the other methods. As observed in Fig. 9 for *Office*, the proposed MEF algorithm produces better visual details in the shelf region especially for the toys, when compared with Mertens and S.Li. Moreover, the specific features of the MathWorks Environment and the *Peppers* image can be seen more clearly on the computer screen. However, it is important to note that there is a slight saturation problem (e.g., the pathway behind the tree) in the proposed result.

The algorithm tends to lose information in cases with excessively over exposed images in the stack. Nevertheless, this method produces the best score statistically, with visually plausible output image for this stack.

Table 3. MEF-SSIM Scores for the First Set of Experiments.

	Algorithms													
	LCE	GBE	Mertens	Raman	Gu	Li12	S.Li	Li13	Paul	Ma	H.Li	LEW	WSM	Proposed
Tower	*0.898*	*0.912*	*0.986*	*0.895*	*0.931*	*0.950*	*0.984*	*0.986*	*0.977*	*0.986*	*0.981*	*0.959*	*0.961*	*0.981*
Garden	*0.917*	*0.928*	*0.989*	*0.911*	*0.927*	*0.951*	*0.982*	*0.984*	*0.982*	*0.985*	*0.977*	*0.975*	*0.989*	*0.990*
Venice	*0.845*	*0.913*	*0.966*	*0.892*	*0.889*	*0.937*	*0.951*	*0.954*	*0.954*	*0.940*	*0.947*	*0.967*	*0.964*	*0.977*
Farmhouse	*0.941*	*0.916*	*0.981*	*0.877*	*0.932*	*0.958*	*0.977*	*0.985*	*0.971*	*0.984*	*0.974*	*0.929*	*0.979*	*0.979*
Landscape	*0.901*	*0.961*	*0.976*	*0.953*	*0.941*	*0.948*	*0.972*	*0.942*	*0.972*	*0.993*	*0.954*	*0.993*	*0.991*	*0.986*
Lighthouse	*0.793*	*0.944*	*0.980*	*0.938*	*0.934*	*0.968*	*0.953*	*0.950*	*0.965*	*0.970*	*0.962*	*0.978*	*0.962*	*0.974*
Kluki	*0.851*	*0.907*	*0.980*	*0.902*	*0.922*	*0.948*	*0.965*	*0.968*	*0.952*	*0.970*	*0.957*	*0.955*	*0.930*	*0.960*
Cave	*0.861*	*0.837*	*0.974*	*0.693*	*0.933*	*0.923*	*0.961*	*0.978*	*0.964*	*0.948*	*0.959*	*0.930*	*0.953*	*0.968*
Office	*0.831*	*0.955*	*0.984*	*0.907*	*0.899*	*0.954*	*0.972*	*0,967*	*0.973*	*0.988*	*0.970*	*0.985*	*0.989*	*0.991*
Balloons	*0.770*	*0.862*	*0.969*	*0.768*	*0.913*	*0.941*	*0.944*	*0.948*	*0.893*	*0.965*	*0.949*	*0.941*	*0.949*	*0.964*
Belgium	*0.732*	*0.874*	*0.971*	*0.809*	*0.896*	*0.954*	*0.947*	*0.964*	*0.899*	*0,973*	*0.945*	*0.923*	*0.961*	*0.969*
Desk	*0.577*	*0.836*	*0.969*	*0.729*	*0.875*	*0.945*	*0.964*	*0.928*	*0.851*	*0.956*	*0.916*	*0.889*	*0.932*	*0.932*
M. Capitol	*0.779*	*0.886*	*0.977*	*0.763*	*0.864*	*0.949*	*0.918*	*0.968*	*0.932*	*0.983*	*0.936*	*0.969*	*0.970*	*0.981*
Average	*0.822*	*0.902*	*0.977*	*0.849*	*0.912*	*0.948*	*0.961*	*0.967*	*0.945*	*0.973*	*0.956*	*0.953*	*0.963*	*0.973*

a) *Office* stack has 6 input exposures.

b) Mertens (0.984) c) S. Li (0.972) d) Proposed (0.991)

Figure 9. Visual Comparison of the Proposed Method with Mertens and S.Li for *Office* (Source: Ma, Zeng and Wang, 2015).

As seen in Fig. 10, specific features in *Garden* are better preserved in the output of the proposed algorithm. While the colors for the sky appear artificial in Mertens, both Ma and the proposed method uncover a natural sky scene. Additionally, the reflections on the water are more informative and visually plausible in the proposed output. However, the details on the rooftop and ivies seem to have a lower contrast when compared to Ma.

In the fused *Venice* image obtained via the proposed method, the details are greatly preserved as shown in Fig. 11. Even though the algorithm produces a brighter image than the remaining methods, the fine details are generally much more effectively recovered. In particular, the sky is more natural, and clouds are more vivid. The specific features on the boat can be easily distinguished and the details are more visible on the scene, while some details are lost (e.g., of the boat) in Mertens which has the second best MEF-SSIM score.

As it can be seen in Table 3, the proposed algorithm has the second best MEF-SSIM scores for *Tower*, *M. Capitol* and *Landscape* stacks. The visual comparison of the fused Tower stack is given in Fig. 12. Although MEF-SSIM scores of Mertens, Ma and Li13 are slightly better than the proposed method, lack-of-contrast regions and detail-loss in several areas are present in the outputs of these algorithms. Since the Eiffel Tower has low contrast, highlights are almost non-existent in the dark regions in Mertens and Li13. Also, there are lost details mainly on the upper side of the tower in these methods. The proposed algorithm in contrast shows better visual quality with

brighter details of the tower and more natural looking clouds.

a) *Garden* stack has 3 input exposures.

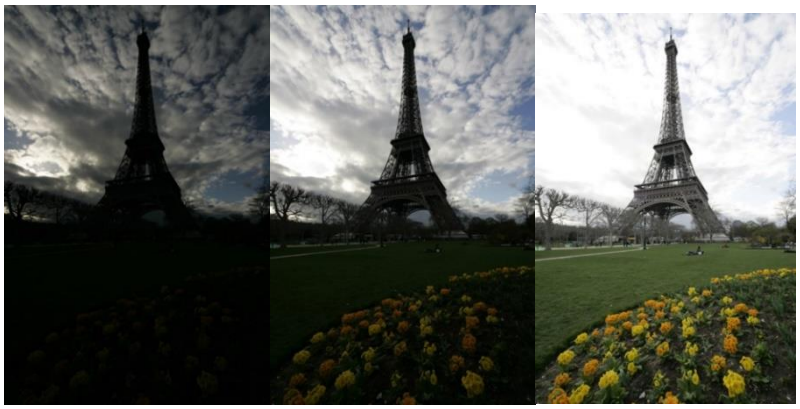
b) Mertens (0.989) c) Ma (0.985) d) Proposed (0.990)

Figure 10. Visual Comparison of the Proposed Method with Mertens and Ma for *Garden* (Source: Ma, Zeng and Wang, 2015).

a) *Venice* stack has 3 input exposures.

b) Mertens (0.966) c) Li13 (0.954) d) Proposed (0.978)

Figure 11. Visual Comparison of the Proposed Method with Mertens and Li13 for *Venice* (Source: Ma, Zeng and Wang, 2015).

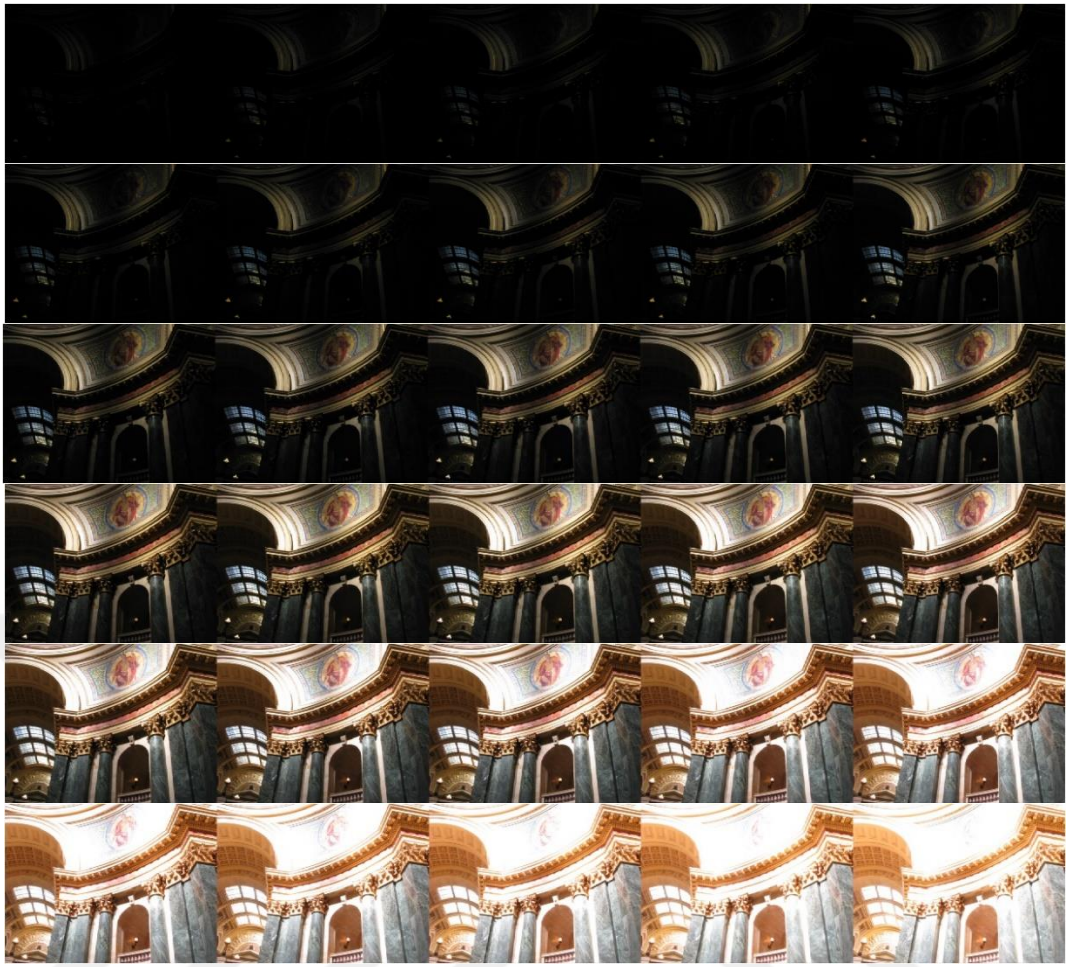


a) *Tower* stack has 3 input exposures

b) Mertens (0.986) c) Li13 (0.986) d) Proposed (0.981)

Figure 12. Visual Comparison of the Proposed Method with Mertens and Li13 for *Tower* (Source: Ma, Zeng and Wang, 2015).

Compared to the other stacks, the *M. Capitol* has the most exposures with 30 input images. Since there are many exposures, even though the over exposed image is too bright and the under exposed image is too dark, the proposed algorithm successfully fused the input images which can be seen in Fig. 13. The proposed algorithm's MEF-SSIM score is not high as Ma's score, but it manages to preserve the fine details at marbles, roof window and the lamp on the balcony. Although, Mertens's algorithm is statistically the most successful MEF method on average, colors at the roof windows look artificial which look more natural in both Ma's algorithm and the proposed method. Moreover, while the wall near the lamp in the small balcony is too bright in Ma's and Mertens' output which causes an artificial looking region, the fused image via the proposed method covers the details at this region without any artifacts.



a) *M.Capitol* stack has 30 input exposures.

b) Ma (0.983)

c) Mertens (0.977)

d) Proposed (0.981)

Figure 13. Visual Comparison of the Proposed Method with Mertens and Ma for *M.Capitol* (Source: Ma, Zeng and Wang, 2015).

For the *Landscape* stack in Fig. 14, the colors of the fused image via the proposed method look more vivid, while both the output images of Ma and Mertens appear generally toneless in the image.

a) *Landscape* stack has 3 input exposures.

b) Ma (0.993) c) Mertens (0.976) d) Proposed (0.986)

Figure 14. Visual Comparison of the Proposed Method with Mertens and Ma for *Landscape* (Source: Ma, Zeng and Wang, 2015).

The algorithm gives the third best MEF-SSIM scores in *Belgium*, *Lighthouse*, *Cave* and *Balloons* stacks when compared to other methods. The visual comparison and the input exposures of *Belgium* are demonstrated in Fig. 15. Even though, the fused images of each method are visually plausible, the writings on the black board are more readable in both Mertens and Ma, and the colors are more natural in the proposed method's output. The details outside the house can be seen in all of the resulted images, but in the fused image of the proposed method, the details are more clearly observable, and the colors are more naturally looking.

For *Lighthouse* in Fig. 16, the specific features and fine details in low light regions, i.e., sand, the rocks and house, are more visible in the proposed method compared to Mertens. On the other hand, Mertens and Li12 preserve more details on the sea and sky whereas some saturation problems exist in these regions in the proposed output. As aforementioned, this undesired outcome occurs when the number of input image is not much, and the stack has extremely over exposed image.

a) *Belgium* stack has 9 input exposures.

b) Ma (0.973)

c) Mertens (0.971)

d) Proposed (0.969)

Figure 15. Visual Comparison of the Proposed Method with Mertens and Ma for *Belgium* (Source: Ma, Zeng and Wang, 2015).

a) *Lighthouse* stack has 3 input exposures.

b) Mertens (0.980)

c) Li12 (0.968)

d) Proposed (0.974)

Figure 16. Visual Comparison of the Proposed Method with Mertens and Li12 for *Lighthouse* (Source: Ma, Zeng and Wang, 2015).

Fig. 17 further compares the fused *Cave* stack with four different exposures. Li13 presents the best MEF-SSIM score for this image and Mertens has a similar statistical score to the proposed algorithm. The left-side of the cave entrance is artificially dark in Li13, but more natural in the proposed result. In addition, in Mertens, the fused image has less contrast especially at the top of the cave and at the rock on the right.

A visual comparison of the fused outputs for the *Balloons* stack containing nine

exposures can be seen in Fig. 18. For the *Balloons* stack, the MEF-SSIM scores are close to each other in the top three MEF approaches but the visual qualities are distinct. Although, Ma's score is slightly higher than the proposed method's, there are artifacts which can be clearly seen in the sun and the clouds in Ma's result. Since, the input sequence contains an extremely over exposed image, the clipped regions which are in the surroundings of the sun can be seen in the proposed image. But still, the proposed method offers a visually plausible output with natural looking colors.

a) *Cave* stack has 4 input exposures.

b) Mertens (0.974)

c) Li12 (0.978)

d) Proposed (0.968)

Figure 17. Visual Comparison of the Proposed Method with Mertens and Li13 for *Cave* (Source: Ma, Zeng and Wang, 2015).

a) *Balloons* stack has 9 input exposures.

b) Mertens (0.969)

c) Ma (0.965)

d) Proposed (0.964)

Figure 18. Visual Comparison of Different Methods for *Balloons* (Source: Ma, Zeng and Wang, 2015).

In the final part of the first set of the experiments, *Kluki*, *Farmhouse* and *Desk* stacks are analyzed. Among all other MEF methods which are compared in this thesis, the proposed method was 4th among them for these stacks. The main reason of this decrease in the statistical performance is that all these stacks contain images which have extremely over exposed regions in most informative parts. For instance, the Fig. 19 demonstrates the exposures of the *Kluki* sequence. Both normal exposed and over exposed images have extreme brightness in the sky region. Although the sun and clouds are still visible in the outputs, these areas still look over exposed for both Ma's algorithm and the proposed method. For this stack, Mertens covers all the details in the clouds and sunlight. On the other hand, while the grass is too dark in Ma's output and too bright in Mertens' outcome, it is natural looking in the result of the proposed method.

For the *Farmhouse* stack (Fig. 20), preserving the details inside the room and outside the windows is a challenging task since the room is dark for both under and normal exposed images, and outside the window is too bright for over and normal exposed images in stack. In Li13's method it is observable that, the details both inside the room and the outside the lower window can be seen clearly, and the details outside the window and the objects in the room look sharp and the colors are vivid. Also, in Ma's result the inside of the room is brighter than the proposed algorithm's result. On the other hand, the details outside the window are still too bright for both Ma's and proposed method's outcome.

The *Desk* sequence (Fig. 21) has 15 exposures and preserving the details on the paper and inside the lamp is troublesome. The fused image via the proposed algorithm appears to be bright and there is an artificial looking dark spot on the middle part of the paper. Hence, the statistical result is low, and the visual quality is not plausible as Mertens' method or as S. Li's algorithm.

As a final note for the first set of experiments, all experimental results given above are fused in 1.76s on the average, ranging between 1.5s to 2.5s.

a) *Kluki* stack has 3 input exposures.

b) Mertens (0.980) c) Ma (0.970) d) Proposed (0.960)

Figure 19. Visual Comparison of the Proposed Method with Mertens and Ma for *Kluki* (Source: Ma, Zeng and Wang, 2015).

a) *Farmhouse* stack has 3 input exposures.

b) Li13 (0.985) c) Ma (0.984) d) Proposed (0.979)

Figure 20. Visual Comparison of the Proposed Method with Li13 and Ma for *Farmhouse* (Source: Ma, Zeng and Wang, 2015).

a) *Desk* stack has 15 input exposures.

b) Mertens (0.969) c) S.Li (0.964) d) Proposed (0.932)

Figure 21. Visual Comparison of the Proposed Method with Mertens and S.Li for *Desk* (Source: Ma, Zeng and Wang, 2015).

As mentioned previously, the proposed MEF algorithm is tested with high-resolution images in order to increase the variety of the analysis and to observe the resolution's impact on the execution time. The dataset in this second setup includes five different static image stacks, namely *Flowers*, *SeaRock*, *SecretBeach*, *OldHouse*, *Rovinia*, of sizes 720×1080 pixels with three exposures each as it can be seen in Table 1. While the best MEF-SSIM scores in Table 4 are reached via the proposed approach (0.969 on average), the average execution times of the best three algorithms are 0.50s, 7.85s and 6.37s for Mertens, Ma and the proposed method, respectively. The MEF-SSIM scores for Mertens, Raman and S.Li are aligned with (Merianos and Mitianoudis, 2019) and the results provided for Ma are obtained through the code available at (Ma Algorithm, 2017) .

Table 4. MEF-SSIM Scores for each Stack (Source: Merianos and Mitianoudis, 2019).

	Algorithms				
	Mertens	Raman	S.Li	Ma	Proposed
Flowers	*0.964*	*0.906*	*0.921*	*0.987*	*0.989*
SeaRock	*0.932*	*0.896*	*0.913*	*0.933*	*0.958*
SecretBeach	*0.951*	*0.927*	*0.888*	*0.899*	*0.963*
OldHouse	*0.974*	*0.959*	*0.907*	*0.987*	*0.991*
Rovinia	*0.934*	*0.881*	*0.913*	*0.935*	*0.943*
Average	0.951	0.914	0.908	0.948	0.969

The visual comparisons are also provided in Figs. 22, 23, 24, 25, 26 for *Flowers*, *SeaRock*, *SecretBeach*, *OldHouse*, *Rovinia*, respectively. It can be observed from these illustrations that the details and colors are better preserved in the proposed technique which results in more natural-looking outputs. Furthermore, the obtained images are more vivid and visually appealing. This mainly results from the successful usage of LE and WSM to fuse an image stack, which has a significant potential.

a) *Flowers* stack has 3 input exposures.

b) Mertens (0.964) c) Ma (0.987) d) Proposed (0.989)

Figure 22. Visual Comparison of Different Methods for *Flowers* (Source: Merianos and Mitianoudis, 2019).

a) *SeaRock* stack has 3 input exposures.

b) Mertens (0.932) c) Ma (0.933) d) Proposed (0.958)

Figure 23. Visual Comparison of Different Methods for *SeaRock* (Source: Merianos and Mitianoudis, 2019).

a) *SecretBeach* stack has 3 input exposures.

b) Mertens (0.951) c) Ma (0.899) d) Proposed (0.963)

Figure 24. Visual Comparison of Different Methods for *SecretBeach* (Source: Merianos and Mitianoudis, 2019).

a) *OldHouse* stack has 3 input exposures.

b) Mertens (0.974) c) Ma (0.987) d) Proposed (0.991)

Figure 25.Visual Comparison of Different Methods for *OldHouse* (Source: Merianos and Mitianoudis, 2019).

a) *Rovinia* stack has 3 input exposures.

b) Mertens (0.934) c) Ma (0.935) d) Proposed (0.943)

Figure 26. Visual Comparison of Different Methods for *Rovinia* (Source: Merianos and Mitianoudis, 2019).

CHAPTER 5: CONCLUSION

While capturing or projecting HDR content, preserving the information in highlights and shadows is a challenging task due to the dynamic range gap between equipment. In order to obtain high quality images containing vivid colors and fine details, HDR compatible devices can be used. While in the near future, HDR compatible equipment is expected to be widely accessible, currently due to their high cost, they are not preferred on user grade level. Hence, currently imaging technology manufacturers seek the solution not in hardware but in software. Therefore, two main approaches used in HDRI are tone mapping and MEF. However, due to problems in tone mapping such as low subjective contrast, ghosting and color saturation reduction which damage the overall quality of the image, most of the current capturing devices use the MEF approach. Hence, HDR-like image reconstruction through MEF is a common study field in image processing and computer vision, and weight map extraction typically presents the novel part of different algorithms. Therefore, this thesis initially analysis the most impactful studies in the literature then proposes a novel MEF algorithm based on LE of pixel/patch spaces of images and WSM. In the developed method, LEW are extracted from differently exposed images and the corresponding WSM are used to adjust these maps according to the informative parts of input images for the final fusion step. After a fused image is acquired, the low-light areas and unsatisfactory color intensities are corrected via a simple local brightness enhancement algorithm. To the best of available knowledge, the proposed framework in this study is novel since it exploits LE of image spaces and WSM of images for the first time in MEF. The proposed algorithm is compared with several state-of-the-art techniques. In the comprehensive comparison in Chapter 4, it is observed that the proposed algorithm produces natural-looking HDR-like contents with mostly superior statistical results compared to existing techniques.

The main drawback of the proposed method is that the visual quality and statistical scores tend to decrease slightly when the input stack contains excessively over exposed or excessively dark under exposed images. This problem can simply be solved by manually discarding extreme exposures; alternatively, automated techniques can be envisaged by means of, for example, outlier detection techniques. Since this process was not within the scope of this thesis, such extension is designated as future work.

Furthermore, the proposed algorithm can be adapted for dynamic image stacks in the future. As a final remark, the main computational complexity lies on the three optimization problems given in Eqn. 33 for LE. This run-time complexity can be greatly reduced by parallel implementations on GPU processors.

REFERENCES

Akyuz, A. O. and Reinhard, E. (2006) *Color appearance in high-dynamic-range imaging*, Journal of Electronic Imaging. International Society for Optics and Photonics, Vol.15(3), pp. 33001.

Basu, S., Davidson, I. and Wagstaff, K. (2008) *Constrained Clustering: Advances in Algorithms, Theory, and Applications*. 1st edition. Boca Ratton: CRC Press.

Beucher, S. and Lantuejoul, C. *Use of Watersheds in Contour Detection*. *Proceedings of the International Workshop on Image Processing*. CCETT/IRISA, Rennes. September 1979.

Burt, P. J. and Kolczynski, R. J. *Enhanced Image Capture through Fusion*. *International Conference on Computer Vision*. Berlin, Germany. 11-14 May 1993.

Chang, H., Yeung, D.-Y. and Xiong, Y. *Super-Resolution through Neighbor Embedding. IEEE Conference on Computer Vision and Pattern Recognition.* Washington DC, USA. 27 June - 2 July 2004.

Chung, T. D. and Khan, M. K. A. A. *Watershed-based Real-Time Image Processing for Multi-Potholes Detection on Asphalt Road*. *IEEE International Conference on System Engineering and Technology.* Shah Alam, Malaysia. 7 October 2019.

Donoho, D. L. and Grimes, C. (2003) *Hessian eigenmaps: locally linear embedding techniques for high-dimensional data*, Proceedings on the National Academy of Sciences, Vol. 100(10), pp. 5591–5596.

Eilertsen, G. (2018) *The High Dynamic Range Imaging Pipeline*. Linköping: Linköping University.

Eilertsen, G., Mantiuk, R. and Unger, J. (2015) *Real-time noise-aware tone mapping*, ACM Transactions on Graphics, Vol. 34(6), pp. 1–15.

Gonzalez, R.C., Woods, R.E. and Eddins, S. (2004) *Digital image processing using MATLAB*. New York, USA: Prentice Hall.

Gu, B., Li, W., Wong, J., Zhu, M. and Wang, M. (2012) *Gradient field multi-exposure images fusion for high dynamic range image visualization*, Journal of Visual Communication and Image Representation, Vol. 23(4), pp. 604–610.

Guan, J. and Qiu, G. *Display HDR Image using a Gain Map*. *IEEE International Conference on Image Processing*. San Antonio, TX, USA. 16 September - 19 October 2007.

Hayat, N. and Imran, M. (2019) *Ghost-free multi exposure image fusion technique using dense SIFT descriptor and guided filter*, Journal of Visual Communication and Image Representation, Vol. 62(-), pp. 295–308.

Hoefflinger, B. (2007) *High-dynamic-range (HDR) vision*. Berlin, Heidelberg: Springer.

Jain, A. K. (2010) *Data clustering: 50 years beyond K-means*, Pattern Recognition Letters, Vol. 31(8), pp. 651–666.

Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M., Saba, T., Zahoor, S. and Akram, T. (2019) *Brain tumor detection and classification: a framework of marker-based watershed algorithm and multilevel priority features selection*, Microscopy Research and Technique, Vol. 82(6), pp. 909–922.

Khan, S. S. and Ahmad, A. (2004) *Cluster center initialization algorithm for K-means clustering*, Pattern Recognition Letters, Vol. 25(11), pp. 1293–1302.

Kim, J. and Lee, S. (2020) *Information measure based tone mapping of outdoor LDR image for maximum scale-invariant feature transform extraction*, Electronics Letters, Vol. 56(11), pp. 544–546.

Kiser, C., Reinhard, E., Tocci, M. and Tocci, N. *Real Time Automated Tone Mapping System for HDR Video*. *IEEE International Conference on Image Processing*. Orlando, Florida. 30 September - 3 October 2012.

Lee, S., Park, J. S. and Cho, N. I. *A Multi-Exposure Image Fusion Based on the Adaptive Weights Reflecting the Relative Pixel Intensity and Global Gradient*. *IEEE International Conference on Image Processing.* Athens, Greece. 7-10 October 2018.

Li, H. and Zhang, L. *Multi-Exposure Fusion with CNN Features*. *IEEE International Conference on Image Processing.* Athens, Greece. 7-10 October 2018.

Li, H. (2018) *Li CNN Algorithm* [Online]. Avaliable at: https://github.com/xiaohuiben/MEF-CNN-feature (Accessed 23 June 2020).

Li, S., Kang, X., Fang, L., Hu, J. and Yin, H. (2017) *Pixel-level image fusion: A survey of the state of the art*, Information Fusion, Vol. 33(-), pp. 100–112.

Li, S. and Kang, X. (2012) *Fast multi-exposure image fusion with median filter and recursive filter*, IEEE Transactions on Consumer Electronics, Vol. 58(2), pp. 626–632.

Li, S., Kang, X. and Hu, J. (2013) *Image fusion with guided filtering*, IEEE Transactions on Image Processing, Vol. 22(7), pp. 2864–2875.

Li, R., Yu, S., and Yang, X. (2007). *Efficient spatio-temporal segmentation for extracting moving objects in video sequences*, IEEE Transactions on Consumer Electronics, Vol. 53(3), pp. 1161-1167.

Li, Z. G., Zheng, J. H. and Rahardja, S. (2012) *Detail-enhanced exposure fusion*, IEEE Transactions on Image Processing, Vol. 21(11), pp. 4672–4676.

Ma, K. (2017) *Ma17Code* [Online]. Available at: https://ece.uwaterloo.ca/~k29ma/codes/. (Accessed 23 June 2020).

Ma, K., Li, H., Yong, H., Wang, Z., Meng, D. and Zhang, L. (2017) *Robust multi-exposure image fusion: a structural patch decomposition approach*, IEEE Transactions on Image Processing, Vol. 26(5), pp. 2519–2532.

Ma, K. and Wang, Z. *Multi-Exposure Image Fusion: a Patch-Wise Approach*. *IEEE International Conference on Image Processing.* Quebec City, QC, Canada. 27-30 September 2015.

Ma, K., Zeng, K. and Wang, Z. (2015) *Perceptual quality assessment for multi-exposure image fusion*, IEEE Transactions on Image Processing, Vol. 24(11), pp. 3345–3356.

Mccollough, F. (2008) *Complete guide to high dynamic range digital photography*. 1st edition. New York: Lark Books.

Merianos, I. and Mitianoudis, N. (2019) *Multiple-exposure image fusion for HDR image synthesis using learned analysis transformations*, Journal of Imaging, Vol. 5(3), pp. 32.

Mertens, T., Kautz, J. and Van Reeth, F. (2009) *Exposure fusion: a simple and practical alternative to high dynamic range photography*, Computer Graphics forum, Vol. 28(1), pp. 161–171.

Paul, S., Sevcenco, I. S. and Agathoklis, P. (2016) *Multi-exposure and multi-focus image fusion in gradient domain*, Journal of Circuit, Systems and Computers, Vol. 25(10), pp. 1650123.

Paul, S. (2016) *Paul Algorithm* [Online]. Avaliable at: https://github.com/sujoyp/gradient-domain-imagefusion . (Accessed 23 June 2020).

Raman, S. and Chaudhuri, S. *Bilateral Filter based Compositing for Variable Exposure Photography*. *Eurographics.* Munich, Germany. 30 March - 3 April 2009.

Reinhard, E., Stark, M., Shirley, P. and Ferwerda, J. *Photographic Tone Reproduction for Digital Images. Conference on Computer Graphics and Interactive Techniques.* San Antonio, Texas. July 2002.

Roweis, S. T. and Saul, L. K. (2000) *Nonlinear dimensionality reduction by locally linear embedding*, Science, Vol. 290(5500), pp. 2323–2326.

Sevcenco, I. S., Hampton, P. J. and Agathoklis, P. (2015) *A wavelet based method for image reconstruction from gradient data with applications*, *Multidimensional Systems and Signal Processing*. Springer, Vol. 26(3), pp. 717–737.

Tenenbaum, J. B., De Silva, V. and Langford, J. C. (2000) *A global geometric framework for nonlinear dimensionality reduction*, Science, Vol. 290(5500), pp. 2319–2323.

Tomasi, C. and Manduchi, R. *Bilateral Filtering for Gray and Color Images*. *International Conference on Computer Vision.* Bombay, India. 7 January 1998.

Türkan, M., Thoreau, D. and Guillotel, P. *Self-Content Super-Resolution for Ultra-HD Up-Sampling*. *European Conference on Visual Media Production.* London, United Kingdom. December 2012.

Türkan, M., Thoreau, D. and Guillotel, P. *Optimized Neighbor Embeddings for Single-Image Super-Resolution*. *IEEE International Conference on Image Processing*. Melbourne, VIC, Australia. 15 - 18 September 2013.

Türkan, M., Thoreau, D. and Guillotel, P. *Iterated Neighbor-Embeddings for Image Super-Resolution*. *IEEE International Conference on Image Processing*. Paris, France. 27 - 30 October 2014.

Vincent, L. (1993) *Morphological grayscale reconstruction in image analysis: applications and efficient algorithms*, IEEE Transactions on Image Processing, Vol. 2(2), pp. 176–201.

Wang, Z., Bovik, A. C., Sheikh, H. R. and Simoncelli, E. P. (2004) *Image quality assessment: from error visibility to structural similarity*, IEEE Transactions on Image Processing, Vol. 13(4), pp. 600–612.

Wu, S., Xie, S., Rahardja, S. and Li, Z. *A Robust and Fast Anti-Ghosting Algorithm for High Dynamic Range Imaging*. *IEEE International Conference on Image Processing*. Hong Kong, China. 26 - 29 September 2010.

Zhang, X. (2019) *Online defect detection method of array optical sensor based on watershed image processing algorithm*, Multimedia Tools and Applications [online]. Available at https://link.springer.com/article/10.1007/s11042-019-7340-y (Accessed 30 June 2020).

Appendix A. Examples of Currently Available HDR Compatible Devices

Company	Samsung	Sony	Nikon	Fujifilm	Canon	Kodak	Panasonic	Pentax	Olympus
Digital Camera	NX1	SLT A99	D5	GFX100	5D Mark III	AZ528	LUMIX GH5	KP	E-M10 Mark II
	NX30	SLT A77II	D600	GFX50R	6D DSLR	AZ401	LUMIX G7	K-70	E-M10 Mark III
	WB2200F	ALT A57	D800	X-T4	650D	AZ901	LUMIX GX9	645Z	Stylus SP-100

Company	Samsung	Sony	Google	Huawei	Xiomi	LG	Apple	Nokia	OnePlus
Smart Phone	Galaxy S20+	Xperia XZ3	Pixel 3	P40 Pro	Mi 9T	G6	Iphone 8 Plus	Nokia 6.2	7
	Galaxy Note 10+	Xperia XZ2	Pixel 3a XL	P30 Pro	Mi 10 Pro	X5	Iphone X	Nokia 8.1	7T
	Galaxy Z Flip	Xperia 1 II	Pixel 4 XL	Mate Xs	Red Mi K20	V40	Iphone 11 Pro Max	Nokia 9 PureView	Reno Z

Company	Samsung	Sony	Vestel	Beko	Arçelik	LG	Toshiba	Asus	MSI
Display	The Frame QLED TV	Master Series Z9G	75U9400	B55 OLED 9890 5B	A55L 8900 5A Diamond Pro	86UN85006LA	65X9763DAT	ROG PG27UQ	Optix MAG277CQR
	Q900R QLED	Master Series A9G	65O9900	B55L 8900 5A Crystal Pro	A75L 8870 5B Diamond	70UN71006LA	65UL3063DT	ProArt PA32UCX-K	Optix MAG322CR
	RU7300 Curved	Z8H	65UD9860	B49L 8900 5A Crystal Pro	A65L 9785 5S ULTIMA	75UN71006LC	43V6863DAT	PA329C	Optix MAG273R

Algorithm 1 Determination of Exposures

Inputs: Image stack $\{I_n\}$, $n = 1 \dots N$

Outputs: Determined exposures $\{\mathcal{U}, \mathcal{N}, \mathcal{O}\}$

for all Image $I_n \in \{I_n\}$ **do**
    $I_g = rgb2gray(I_n)$
    $\mathbf{v}_n = histogram(I_g)$
$label(i) = kMeans(\{\mathbf{v}_n\})$, $i = 1, 2, 3$
$\{\mathcal{U}, \mathcal{N}, \mathcal{O}\} = SlidingWindow(\forall I_n \in label(i))$

Algorithm 2 Weight Maps via Linear Embeddings

Inputs: Determined exposures $\{\mathbf{U}, \mathbf{N}, \mathbf{O}\} \leftarrow \{\mathcal{U}, \mathcal{N}, \mathcal{O}\}$

Outputs: Embedding maps $\{\mathbf{E}_k\}$, $k = 1, 2, 3$

Create patch triplet set $\mathcal{P} = \{\mathbf{u}_i, \mathbf{n}_i, \mathbf{o}_i\}_{\forall i}$
for all Triplet $(\mathbf{u}_i, \mathbf{n}_i, \mathbf{o}_i) \in \{\mathbf{u}_i, \mathbf{n}_i, \mathbf{o}_i\}$ **do**
    Solve Eqn. (33) to obtain $\{\mathbf{W}_i^j\}$, $j = 1 \dots 6$
Blend derived weight maps via Eqn. (34) to obtain $\{\mathbf{E}'_k\}$
Normalize and smooth $\{\mathbf{E}'_k\}$ via Eqn. (35) to obtain $\{\mathbf{E}_k\}$

Algorithm 3 Watershed Masking

Inputs: Input image $I \in \{\mathcal{U}, \mathcal{N}, \mathcal{O}\}$ and structuring element $\mathbf{SE}$

Outputs: Watershed mask $\mathbf{M}$

$I_g = rgb2gray(I)$
$I_e = imErode(I_g, \mathbf{SE})$
$I_{obr} = imReconstruct(I_e, I_g)$
$I_{obrd} = imDilate(I_{obr}, \mathbf{SE})$
$I_{obrcbr} = imReconstruct(\sim I_{obrd}, \sim I_{obr})$
$\mathbf{M} = \sim I_{obrcbr}$

Algorithm 4 Exposure Fusion

Inputs: $\{\mathbf{U}, \mathbf{N}, \mathbf{O}\}$, $\{\mathbf{E}_k\}$, $\{\mathbf{M}_k\}$, $k = 1, 2, 3$

Outputs: Fused image $\mathbf{F}$

Derive global masks $\{\mathbf{G}_k\}$ via Eqn. (37)
Apply Eqn. (38) to obtain $\mathbf{F}$

Appendix C. The Publications during Master's Degree

Published in or submitted to journals:

- O.Ulucan, D.Karakaya and M. Turkan. (2020) *Multi-exposure image fusion based on linear embeddings and watershed masking*. Signal Processing, (In Press).
- D.Karakaya, O. Ulucan and M. Turkan (2020) *Image declipping: Saturation Correction in Single Images*. Computer Graphics Forum, (Submitted).
- D.Karakaya, O. Ulucan and M. Turkan (2020) *Electronic nose and its applications: a survey*. International Journal of Automation and Computing, Vol. 17(2), pp. 177 – 209.

Published in or submitted to indexed conference proceedings:

- D. Karakaya, O. Ulucan and M. Turkan. *A Comparative Study on Electronic Nose Data Analysis Tools. The Innovations in Intelligent Systems and Applications Conference.* Marmara University, Istanbul. 15 – 17 October (Accepted).
- O. Ulucan, D. Karakaya and M. Turkan. *A Large-Scale Dataset for Fish Segmentation and Classification. The Innovations in Intelligent Systems and Applications Conference.* Marmara University, Istanbul. 15 – 17 October 2020 (Accepted).
- Y. Oktar, O. Ulucan, D. Karakaya, E. O. Ersoy and M. Turkan. *Binocular Vision based Convolutional Networks*. *IEEE Conference on Signal Processing and Communications Applications.* Gaziantep. 5 – 7 October 2020 (Accepted).
- O. Ulucan, D. Karakaya and M. Turkan. *Meat Quality Assessment based on Deep Learning. The Innovations in Intelligent Systems and Applications Conference.* Yasar University, Izmir. 31 October – 2 November 2019.
- D. Karakaya, O. Ulucan and M. Turkan. *A Comparative Analysis on Fruit Freshness Classification. The Innovations in Intelligent Systems and Applications Conference.* Yasar University, Izmir. 31 October – 2 November 2019.