Enhancing color outdoor images with poor quality conditions by fusing visible and NIR information

by

M.Sc. María de los Ángeles Herrera Arellano

Dissertation submitted in partial fulfillment of the requirements for the degree of:

DOCTOR IN COMPUTER SCIENCE

at the

National Institute for Astrophysics, Optics and Electronics

March 2022

Tonantzintla, Puebla, México

Advisor:

Ph.D. Hayde Peregrina Barreto

Dra. Hayde Peregrina Barreto

To my mother and father, for all their support,

Yahir and Luna, for accompanying me on this way

Thank you for your endless love!.

Abstract

Outdoor images are often captured in weather conditions, such as fog, haze, or clouds, that cannot be controlled. These conditions degrade images causing loss of contrast and fine details in color images since dispersion has dimming and smoothing effects. The NIR band of the electromagnetic spectrum has become an important source for improving the quality of RGB images. NIR has high transmittance characteristics, which visually represent how materials absorb and reflect energy. So in NIR images, the contrast between foreground objects and the background is higher than in visible images. Thus, by fusing the RGB and NIR information, it is possible to complement and improve the quality of visual information. Although current V-NIR fusion methods allow image enhancement, some issues such as edge preservation and oversaturation need to be addressed. This thesis presents the results fusion method, which performs a selective V-NIR fusion of the most relevant structures through top-hat and bottom-hat morphological transformations. The performance of the methods is evaluated by quantifying added information, color similarity, and saturation. Experimental results show that it was possible to enhance image quality while color oversaturation is avoided. The proposed method showed to be competitive with other fusion methods. In addition, the proposed method can estimate the sizes of the structures that mainly must be enhanced, having the characteristics of adaping the enhancement according to the image information.

Keywords— V-NIR image fusion, image enhancement, mathematical morphology.

Resumen

Las imágenes en exteriores se suelen capturar en condiciones climáticas que no se pueden controlar, tales como neblina, niebla o nubes. Estas condiciones degradan las imágenes causando pérdida de contraste y detalles finos en las imágenes de color, ya que la dispersión tiene efectos de atenuación y suavizado. La banda del infrarrojo cercano (NIR) del espectro electromagnético se ha convertido en una fuente importante para mejorar la calidad de las imágenes RGB. El NIR tiene la característica de alta transmitancia, que permite representar visualmente cómo los materiales absorben y reflejan la energía. Por lo que en las imágenes NIR el contraste entre los objetos de primer plano y el fondo es mayor que en imágenes visibles.

Al realizar la fusión de la información contenida en visible y el NIR es posible rescatar información complementaria y así mejorar la calidad de las imágenes. A pesar de que los métodos de fusión V-NIR actuales logran la mejora de las imágenes, aún deben abordarse algunos problemas como la preservación de los bordes y la sobresaturación.

En esta tesis se reportan los resultados del método fusión selectiva V-NIR de las estructuras más relevantes a través de las transformaciones morfológicas top-hat y bottom-hat. El desempeño de los métodos se evalúa cuantificando la información agregada, similitud de colores y saturación. Los resultados experimentales muestran que fue posible mejorar la calidad de la imagen mientras se evita la sobresaturación. El método propuesto demostró ser competitivo con otros métodos de fusión. Además, de permitir estimar los tamaños de las estructuras que principalmente deben ser mejoradas, teniendo las características de adaptar el realce según la información de la imagen.

Palabras clave— Fusión de imágenes V-NIR, mejora de imágenes, morfología matemática.

Acknowledgments

It has been four years since I started this project; during this time, life has given me many good and bad surprises from which I have experienced moments of inspiration and gratefulness. Therefore, I want thanks from my heart to all people who have been with me in those different situations.

I want to express my deep and sincere gratitude to my research supervisor, Ph.D. Hayde Peregrina Barreto for allowing me to do research and providing invaluable guidance throughout this research. It was a great privilege and honor to work and study under her guidance. A special thanks to Ph.D. Ivan Terol for influencing my interest in image processing and choosing the thesis topic. Although he is no longer in this earthly world, his teachings were invaluable. Besides my advisors, I would like to thank the rest of my thesis committee Ph.D. Alfonso Martinez, Ph.D. Claudia Feregrino, Ph.D. Felipe Orihuela, Ph.D. Kelsey Ramirez, and Ph.D. Jose Crepo for providing guidance, encouragement, insightful comments, and feedback.

My endless gratitude to my family, especially my parents, for their love, caring, and sacrifices for educating and preparing me for my future. I am very much thankful to my husband and my children for their love, understanding and continuing support to complete this research work. Finally, I want to thank all my friends for their constant encouragement.

I appreciate the financial support from CONACyT through scholarship number 630325. Finally, my thanks go to all the people who have supported me to complete the research work directly or indirectly.

Herrera Arellano María, March 2022

Contents

List of Figures

List of Tables

Acronyms

NIR-Near Infrared

V-NIR-Visible-NIR

$l\alpha\beta$-luminance (l), Alpha (α), Beta(β) color space

CIE L* a* b*-L* a* b* color space

XYZ-XYZ color space

HSV-Hue, Saturation, Value color space

YCbCr-YC_BC_r color space

RGB-Red, Green, Blue color space

CNN-Convolutional Neural Network

CT-Computed Tomography

PET-Positron Emission Tomography

MM-Mathematical Morphology

PS-Pattern Spectrum

BF-Bilateral Filters

WLS-Weight Least Square filter

CIE-International Commission of Light, in French Commission International de l'Éclairage

ANOVA-Analysis of Variance

SSIM-Structural Similarity Index Measure

MSE-Mean Squared Error

PSNR-Peak Signal-to-Noise Ratio

Chapter 1

Introduction

The quality of outdoor color images may be degraded by several conditions such as haze, fog, poor light conditions, even pollution. These conditions often may cause loss of contrast and details in the image, as well as edges attenuation and smoothing effect [Ancuti and Ancuti, 2013, Sadeghipoor et al., 2015, Kudo and Kubota, 2018, Tan, 2008, Elliethy and Aly, 2017]. Figure 1.1 contains some examples of these problems: depicted how haze makes it difficult to observe the division between regions. It also indicates how fog can cause loss of information; other problems as shadows cause shape deformation of objects and loss of contrast between regions, degrading image quality. The objects and relevant information for the scene analysis are difficult to observe and analyze in such conditions. Moreover, the performance in tasks such as segmentation, remote sensing, surveillance, or other computer vision applications could be affected by processing images under these conditions [Kaur et al., 2021].

There are several methods in digital image processing to enhance color images by dehazing process, for example, [Fattal, 2008] method for estimating the optical transmission in hazy scenes given a single input image. Based on this estimation, the scattered light is attenuated. Also, [Cai et al., 2016] proposed a dehazing method based on convolutional neural network (CNN). [He et al., 2011] used image prior-dark channel prior to removing haze from a single input image. While [Zhu et al., 2015] uses color attenuation prior to haze removal. Other methods focus

(a) (b) (c)

Figure 1.1: Outdoor scenes under conditions of (a) haze, (b) fog, and (c) shadows; red rectangles highlight areas where information is lost.

on improving the contrast of color images for example [Barreto and Villalobos, 2011] proposed a morphological rational operator for contrast enhancement. The Retinex algorithm deals with the compensation of image illumination [Jobson, 2004].

Currently, image fusion methods have gained relevance in image improvement due to the high amount of image acquisition systems. Image fusion is a technique that allows combining information from several sources and preserving the detail of each to improve the quality of the information [Kaur et al., 2021]. According to [Zitová and Flusser, 2003], image fusion can be categorize into:

- **Multi-temporal:** Images of the same object are taken at different times to detect changes between them. Usually it is used for monitoring or tracking [Choi et al., 2017, Rad et al., 2007].
- **Multi-view:** Images are acquired from different viewpoints, e.g. three-dimensional microscopy images [Swoger et al., 2007, Trinidad et al., 2019].
- **Multi-sensor:** Images of the same scene come from different sensors to integrate both sources of information to obtain complementary scene representation for example, multispectral/panchromatic [Choi et al., 2005], computed tomography (CT)/positron emission tomography (PET) [Kostakoglu et al., 2004], visible/NIR [Ma et al., 2019].

Fusing another spectrum band, as NIR, can add information that is not captured directly in the visible band. Since NIR images have high transmittance, the contrast is higher in NIR images than in visible images; biomass and vegetation usually have high values in a NIR image. These are a complementary source of information. In recent years, information from the NIR has been used for the enhancement of visible images (V). V-NIR image fusion techniques are widely used in diverse fields including autonomous navigation of vehicles [Choe et al., 2018], security and surveillance [Salamati et al., 2011], aerial or landscape photography, dehazing, tone mapping, biometrics [Sharma and Gool, 2016], object recognition, image enhancement, and remote sensing [Ma et al., 2019].

Several approaches for image enhancement through V-NIR fusion have been proposed, including fusion by mapping and multi-scale transforms. The V-NIR fusion methods are focused on enhancing low-quality regions with information from the NIR, while high-quality regions are kept in the visible image [Sharma et al., 2017, Elliethy and Aly, 2017]; some other methods have a specific purpose such as dehazing [Ahn et al., 2011, Zhuo et al., 2010, Schaul et al., 2009, Fredembach and Süsstrunk, 2008, Kudo and Kubota, 2018, Jang and Park, 2017]. Although V-NIR fusion has proved to be a suitable method for improving image quality, some drawbacks, such as edge blurring and color oversaturation, remain and must be addressed.

1.1 Motivation

There are several challenging situations where the environmental or climatic conditions difficult images future applications, such as autonomous navigation of vehicles, e.g.low-light problems such as those at dusk or dawn, and poor-visibility conditions such as haze, and fog [Choe et al., 2018]. The study and analysis of different sources of images are considered pertinent to obtain information on the same scene. NIR information as a complement to the visible spectrum has become an important tool for image quality improvement. For instance, NIR is used to overcome the low quality of outdoors images and capture information despite climatic conditions.

As mention in [Ghassemian, 2016] the fusion method must preserve the relevant information from the sources, avoiding irrelevant information, noise, artifacts, and inconsistencies in the fused image. One way to evaluate fusion methods is to measure the information retrieved in the fused images, and in the case of color images, the color is also evaluated. Current fusion methods present two main drawbacks: blurred edges and color oversaturation. The blurred edges can be associated

with the type of used filter to select the information to fused. Color oversaturation emerges from the overlap of visible and NIR bands and affects the fused image, making it look unnatural. Thus in outdoor images, it is not common to see pure colors.
Performing a V-NIR fusion method in which as much NIR information as possible is recovered to complement the visible while maintaining the color of the reference image is one way to exploit the benefits of both sources.

1.2 Problem statement

Starting from an image $I(x, y)$ and $h(\bullet)$ where:

- $I(x, y)$ is defined by a bi-dimensional function whose value is n-dimensional $I : \mathbb{Z}^2 \rightarrow \mathbb{R}^2$.
- Each pixel $(x, y)\mathbb{Z}^2$ has an associated value $m = (c1, c2, c3, \cdots, ck)$. The range of values for m depends on the image bit-depth and the color space chosen for its representation [Gonzalez Rafael C. et al., 2009].
- Depending on the number of the n dimension, a image can be binary or grayscale ($n = 1$), color ($n = 3$) or multiespectral ($n > 3$).

Then, it is intended to obtain a new image $\hat{I}\,(x, y)$ where the value of $I(x, y)$ have been altered according to $h(\bullet)$ where:

$$\hat{I}(x, y) = h(I(x, y)) \tag{1.1}$$

Therefore, the problem is to propose and define the parameters of the transformation $h(\bullet)$ that makes the value of fused image contrast higher than original image contrast and satisfies the restrictions of fused image saturation as similar as possible from initial image saturation and, color similarity.
From now on, we denote V to $I(x, y)$, with the filters corresponding to visible (n=3), and N the ones with corresponding to NIR (n=1).

1.3 Research questions

The following questions will guide this doctoral research in order to achieve the scientific contributions in the thesis topic.

Q1. How could the information that complements V be selected from N to overcome the low contrast caused by capture in poor-quality conditions?

Q2. What criterion or method could follow the image fusion process to complement the visible information with the NIR information?

Q3. How can the information related to the color present in the visible be kept once the fusion is done?

1.4 Hypothesis

An image fusion method that selects relevant information from visible (V) and NIR (N) images could be able to enhance low-contrast regions while avoids color oversaturation generating a fused image ($\hat{I}$) of higher quality than V when evaluated of image information and color.

1.5 Objectives

To propose and validate a selective V-NIR fusion method that extracts the relevant information from the NIR image to complement the visible image to increase the image quality. Moreover, the fusion method will be able to avoid color oversaturation in the fused image. The results will be measure through information quality (contrast, entropy) and color (color similarity and saturation).

Specific Objectives

O1. To establish criteria for the selection of relevant information based on the NIR image.

O2. To develop and validate a method capable of fusing the V and NIR images to maximize the contrast in low-quality regions.

O3. To propose an algorithm that overcomes color oversaturation in the resulting fused image.

1.6 The rationale of the solution

Most of the fusion V-NIR methods process the source images in a general way because they assign fixed weights of how much every source contributes to the fusion. Usually, the NIR band has a higher weight since it is assumed that the image structures are better defined on it. However, this condition cannot be generalized because the low-quality regions affected by the haze in our study case do not always cover the entire image. Therefore, there could be regions in V with higher defined than in N, and they must be preserved. But the low-quality regions in V must be complemented with N to enhance them. In this way, current issues such as edges and color distortion could be overcome. Figure 1.2 shows an example of the luminance from V (l_V), a normalized N (l_N) and luminance of the method proposed by [Vanmali et al., 2015] ($l_{Vanmali2015}$). The distribution of the (l_V) has high-intensity values associated with haze in the environment. On the other hand, the (l_N) image has a broad distribution because it is less affected by weather conditions, thus higher contrast. The distribution of ($l_{Vanmali2015}$) is shortened compare to the V, and N when the goal is to take the well-contrasted information of both source images, V, and N, to complement each other and generate an enhanced image.

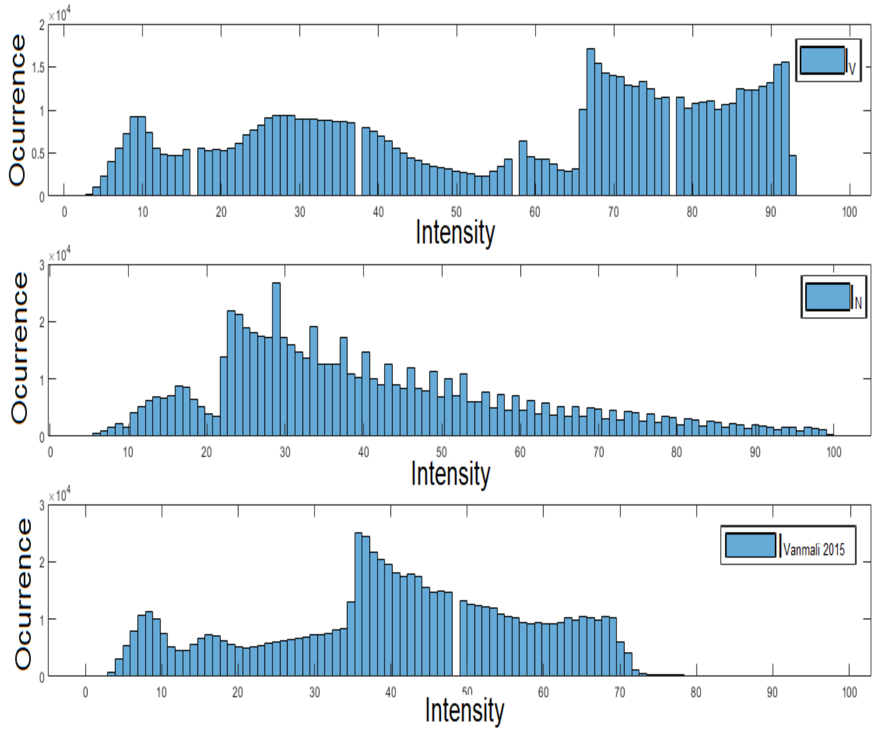


Figure 1.2: Luminances histograms of V, N and [Vanmali et al., 2015] method.

1.7 Scope and limitations

Despite some of the image quality problems that can be solved during the capture, we focus on the already registered images. Therefore, we are not working on enhancing the environment or the physical sensors where the images are taken. Instead, the enhancement of the images is realized from a computing approach.

Another consideration is that the images should be aligned spatially and temporarily for the fusion process. We are not working in real-time; all the processes are offline. The fusion method is designed and proved on outdoor images; we do not consider biomedical images.

1.8 Contribution and research products

The main contribution of this research is a fusion method that allows increasing the quality of images affected by haze conditions. Moreover, the proposed fusion method adds N information without generating blurring edges or color oversaturation. Obtained results showed that the proposed method reached better results than fusion methods in the state-of-the-art. The following works were derived from this research and have been published:

- JCR Journal Article
  - M.A. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, "Visible-NIR Image Fusion Based on Top-Hat Transform, "IEEE Transactions on Image Processing, vol. 30, pp. 4962–4972, 2021. Available: https://ieeexplore.ieee.org/document/9427055/ [Herrera-Arellano et al., 2021].
- Conference paper (Full length)
  - M. A. Herrera-Arellano, H. Peregrina-Barreto, and I. Terol-Villalobos, "Color outdoor image enhance-ment by V-NIR fusion and weighted luminance," in IEEE Autumn Meeting on Power, Electronics and Computing, 2019. ROPEC 2019. IEEE, Ixtapa, Mexico, 2019 [Herrera-Arellano et al., 2019].

1.9 Thesis outline

This chapter presents the introduction, research questions, hypothesis, and objectives that guided this research and the general thesis overview. The rest of the thesis is structured as follows: In chapter 2, the basics theory related to the thesis, beginning with an introduction to digital image processing, color spaces theory, a description of the morphological operators, and transformation applied to imaging processing are presented. Chapter 3 reports the works in the literature that frame the present research regarding the fusion methods of visible and NIR implemented in the state-of-the-art. Chapter 4 describes every step toward the development of the fusion method.

Chapter 5 presented the fusion results with the implemented algorithms, and a comparison is made with some state of art methods. Finally, chapter 6 describes the conclusions of this research and suggestions for future work derived from this thesis.

Chapter 2

Theoretical framework

This chapter presents an introduction to the main concepts used in this work. It seeks to explain the fundamental concepts necessary to understand the technical details of the thesis. First, concepts related to image processing are shown, such as visible and NIR images. Then, color spaces theories and Mathematical Morphology (MM), which are necessary to understand the fusion method. Finally, the functions used to evaluate the fusion methods are explained in detail.

2.1 Digital images

A digital image $I(x, y)$ is defined by a bi-dimensional function whose value is n-dimensional $I : \mathbb{Z}^2 \to \mathbb{R}^2$, where each pixel $(x, y)\mathbb{Z}^2$ has an associated value $m = (c1, c2, c3, \cdots, ck)$. Depending on the number of the n dimension, a image can be binary or grayscale ($n = 1$), color ($n = 3$) or multiespectral ($n > 3$). The range of values for m depends on the image bit-depth and the color space chosen for its representation [Gonzalez Rafael C. et al., 2009].

2.1.1 Visible and NIR images

The electromagnetic spectrum is defined as the graphical representation of electromagnetic waves arranged according to their wavelength. Thus, the electromagnetic spectrum spans the entire range of wavelengths of electromagnetic radiation from the shortest to the longest wavelength that can be generated physically (Figure 2.1) [Zwinkels, 2014]. The human eye is sensitive to the visible spectrum whose wavelength is between 350 nm and 780 nm, approximately. While NIR wavelengths, from 700 to 1100 nm approximately [Whetsel, 1968], have the characteristic of high transmittance, while optical transmittance refers to the amount of light that passes through a body at a specific wavelength; also, transmittance gives a physical measure of the ratio of incident and transmitted intensity as it passes through. As the wavelength lengthens, the transmittance increases so that the absorbed and refracted energy is less than that transmitted, making an object translucent at specific wavelengths. Thus, by this phenomenon, different objects can be observed between the visible and the NIR [Uzal, 2019]. For example, biomass* and vegetation usually have high values in a NIR image since, in outdoor environments, less energy is absorbed from the NIR than from the visible spectrum.

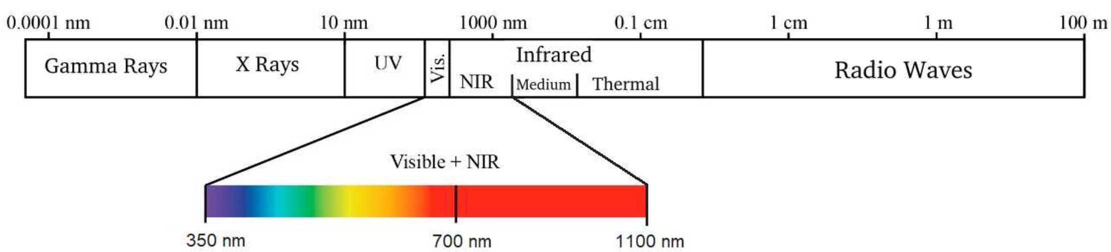
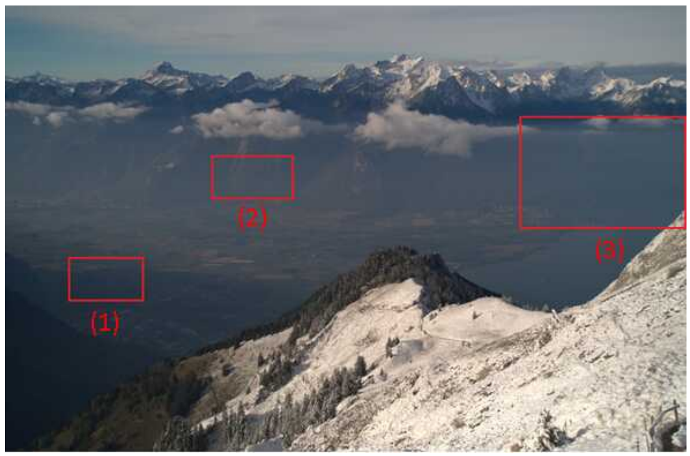


Figure 2.1: The electromagnetic spectrum, showing the range of the visible spectrum+NIR.

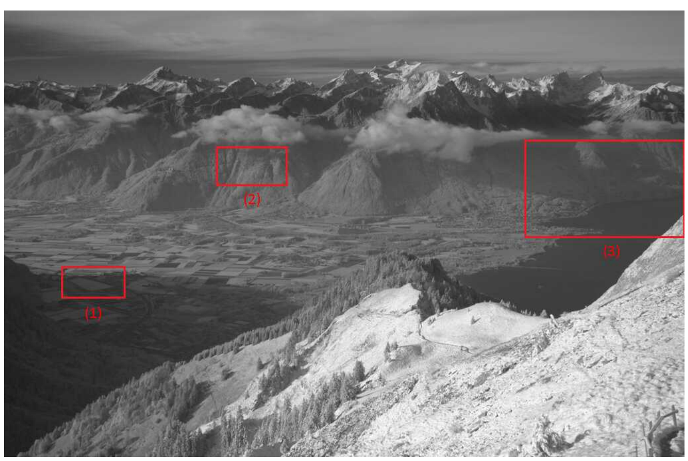
The high contrast in NIR images can be exploited to identify and highlight boundaries, e.g., landforms and shorelines. Moreover, NIR is less sensitive to atmospheric conditions, so it captures information through haze, fog, etc. Rayleigh scattering where the intensity of the scattered light is related to that of the incident light by two variables: the photon's wavelength λ and the scattering particle's size. Haze is formed when the aerosol particles are smaller than $\lambda/10$ and the scattering

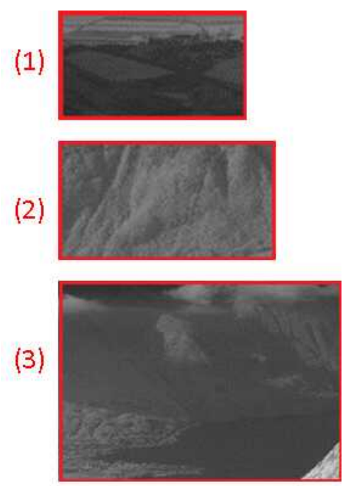
*Biomass is the organic matter generated in biological processes, e.g., the plant material derived from CO2(carbon dioxide) in the air, water, and sunlight via photosynthesis. The biomass's characteristics depend on the chemical and physical properties of the molecules from which it is made. [McKendry, 2002] follows Rayleigh's law: $E_s \propto \frac{E_0}{\lambda^4}$ [Schaul et al., 2009]. This scattering is why, in landscape images, distant objects become blurred and have a blue color cast. This phenomenon, atmospheric haze, is almost absent from NIR images, yielding haze-free images with a more considerable optical depth. [Firmenichy et al., 2011, Barsi et al., 2014, Sharma et al., 2017, Choe et al., 2018].

In Figure 2.2, a pair of visible and NIR images is shown. As observed, the NIR provides relevant information in certain regions that is not noticeable in the visible image.



(a)





(b) (c)

Figure 2.2: Pair of images captured from (a) the visible and (b) the NIR spectra; (c) red rectangles show a zoomed view where information through the haze can be found in NIR.

Visible images typically have a high spatial resolution, where spatial resolution measures the smallest discernible detail in an image. For example, the largest number of discernible line pairs per unit distance (e.g. 250 line pairs per mm) or dots per inch [Gonzalez Rafael C. et al., 2009]. Also, color information; thus, they are suitable for human visual perception. However, these images can be degraded by some conditions, such as poor illumination, fog, haze and other effects of bad weather. While, NIR images, which depict the thermal radiation of objects, overcome to these disturbances but typically have low resolution [Ma et al., 2019]. A color image is generally represented as a combination of three color channels: red, green and blue. However, it is not the only way to represent them, so color spaces are discussed below.

2.2 Color spaces

Color spaces (also named color model) are geometrical representations of color accepted by a convention or a standard. These representations correspond to n-dimensional arrangements of color (vectors of n components), typically color models have three or four color components [Gonzalez Rafael C. et al., 2009]. It is also defined as a system for measuring colors that can be perceived by a human and a process of combining different values as a set of primary colors [Ibraheem et al., 2012]. Finally, color spaces provide a method for specifying, ordering, and manipulating colors. There are numerous color spaces today.

2.2.1 RGB color space

RGB (red, green, and blue) color space in which the quantities of the three channels are mixed to obtain a color. This color space is based on the tristimulus model and additive, which means in RGB space, the color is specified by the positive sum of red, green, and blue, forming in 3D space the cube shown in the Figure 2.3. The range of each RGB color coordinate or component is usually [0,1], although, in multimedia and image processing, the specification in discrete quantities present in the interval [0,255] is more widespread. Images in the RGB color model are three independent image planes, each corresponding to a primary color. When inputting an RGB monitor, the three images are combined on the phosphor screen to produce a composite color image [Gonzalez Rafael C. et al., 2009]. There are some drawbacks with this color space, such as: At this color space is not possible to match all the colors by additive mixture, it is device-dependent and also the tristimulus values depend on the luminance. RGB is commonly used in virtually computer system as television, video etc. However, there will be times when the model is not the most suitable for image processing. For example, in a process or technique that is desirable to enhance the light intensity, a change at the individual channels may result in a color change. This color change is because all the RGB channels are highly correlated. A color model that represents intensity in a component will be the most suitable in this case [Hawkes, 2004].

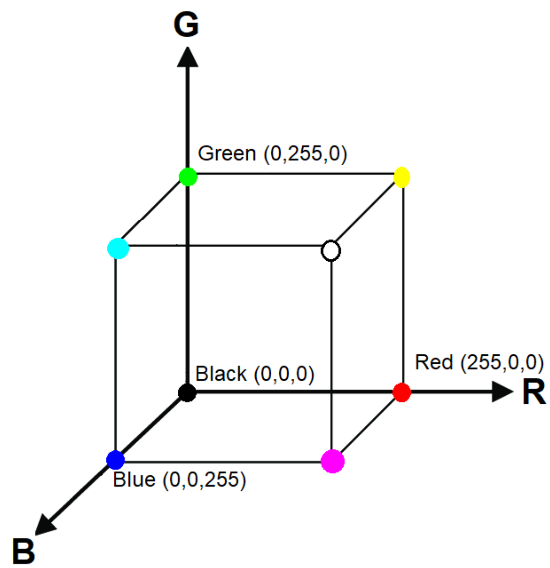


Figure 2.3: A 3D cube geometrically representing the RGB color space

2.2.2 XYZ color space

The XYZ color space is obtained by a linear transformation of the RGB system and was introduced by the International Commission on Illumination (CIE). In this case, XYZ's primary colors are imaginary i.e. they do not represent a physical light. The linear transformation matrix between the RGB and XYZ spaces is presented in Eq. 2.1. The representation of the spectrum is achieved in a 2D plane known as the x, y chromaticity diagram. Figure 2.4 shows the chromatic diagram according to CIE standard. In a chromatic diagram, a straight line connecting any two points defines the variations of the different colors obtained by combining the two primaries in an additive way. The XYZ color space is device independent and is used in applications where the color representation does not depend on the nature of the equipment or hardware. Likewise, the XYZ model is used for standard color specification and colorimetric calculations. It is not usual to find the XYZ space in digital image processing. However, it will be required as an intermediate space to transform the RGB system to other chromatic coordinates such as CIE L*a*b* or $l\alpha\beta$, every color space obtained using XYZ space is also regarded as device independent [Young et al., 1998, Kahu et al., 2019].

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.6067 & 0.1736 & 0.2001 \\ 0.2988 & 0.5868 & 0.1143 \\ 0.0000 & 0.0661 & 1.1149 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.1}$$

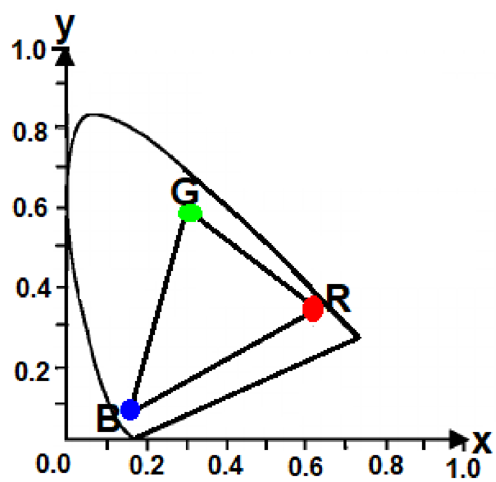


Figure 2.4: Chromatic diagram (x,y) according to the CIE standard.

2.2.3 CIE L*a*b* color space

The CIE standardizes the CIE L*a*b* color space in 1976 to achieve a perceptually uniform representation of the color. In this way, colors are represented at distances proportional to the visual differences between them. CIE L*a*b* is a color-opponent space where the component L* measures luminance, while the components a* and b* define magenta-green and yellow-cyan color, respectively. The spatial representation of the CIE L*a*b* model is illustrated in the Figure 2.5. The L* component ranges from 0 to 100 (dark to light). While if a* component is a positive value tends to be color redness. Suppose a negative value is the color greenness. Finally b* component approximately measures the color yellowness (positive value) or blueness (negative value) [Fairchild, 2005]. For the transformation of the RGB space to CIE L*a*b*, the values are calculated, through non-linear transformations, using the XYZ space (Appendix A.1). CIE L*a*b* space is a linear color space. That is, Euclidean distance between two colors is approximately proportional to their perceived difference by humans. And it is also a uniform space, and thus colors in the CIE L*a*b* space are perceptually more uniformly spaced than those in RGB or HSV. The CIE L*a*b* space is device-independent and used in applications that accurately measure the perceptual distance between two colors e.g., in the comparison of image filtering results, the distance metric to measure color similarity (ΔE). Another characteristic of the CIE L*a*b* is that the luminance component can be processed individually without affecting the chroma of the image. [Sharma and Trussell, 1997, Kahu et al., 2019].

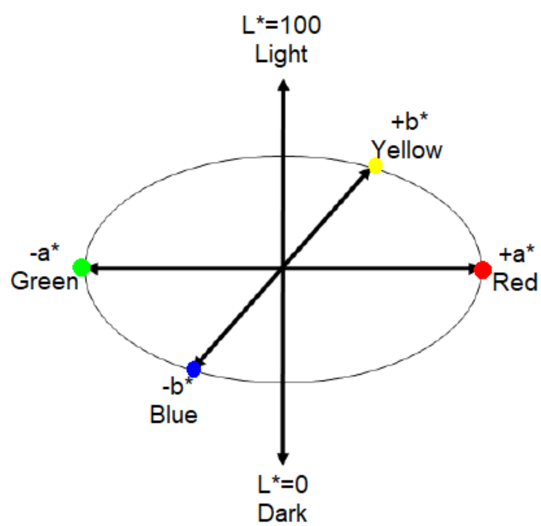


Figure 2.5: Geometrical representation of the CIE L*a*b* space.

2.2.4 YCbCr color space

YCbCr color space is an international standard for digital coding of TV pictures for 525 and 625 lines [BT.601-5, 1995]. The YCbCr color space is defined as one luminance (Y) and two chrominance components (Cb and Cr). In contrast to RGB, the YCbCr color space is luminance independent. The linear conversion between RGB color space and YCbCr color space is described in Appendix A.2. Figure 2.6 shows the YCrCb geometrical representation. As can be seen, the possible RGB colors occupy only part of the YCbCr color space limited by the nominal ranges. Therefore, many YCbCr combinations result in invalid RGB values. It color space is widely used for image and video compression [Kahu et al., 2019].

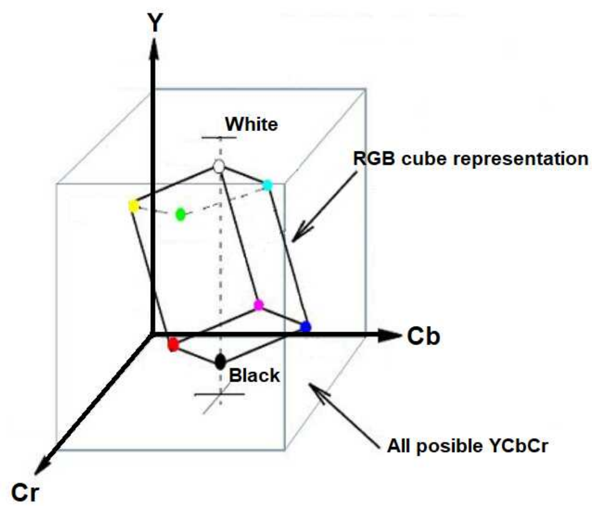


Figure 2.6: RGB color cube in the YCbCr space.

2.2.5 HSV color space

In image processing systems, it is common to specify colors that are compatible with the hardware used. In this sense, the RGB model stands out compared to the rest for being the closest to the computer vision hardware. However, this computationally practical system is not very useful in specifying and recognizing colors. The human being does not recognize color because it has a quantity of red, green, or blue components but uses perceptual attributes of luminance, saturation, and hue.

The HSV color space, proposed by [Smith, 1978], encodes color with the perceptual attributes of luminance (V), saturation (S), and hue (H), and are defined as intuitive, psychological, or user-oriented spaces, as they are optimal for human interaction. It has a cone shape (Figure 2.7) in which the values for the hue axis vary from $0°$ to $360°$, beginning and ending with red and running through all intermediary colors. The V is the vertical axis, and the vertex V=0 corresponds to black color. Like the YCbCr and CIE L* a* b* spaces, the luminance or intensity information is decoupled from the color information present in hue and saturation components. Hue, saturation, and value are calculated from RGB values. However, the HSV color space also has drawbacks e.g. the conversion between spaces is not linear (detailed extensively in Appendix A.3), high instability in saturation and hue values for reduced changes in RGB coordinates [Gonzalez Rafael C. et al., 2009].

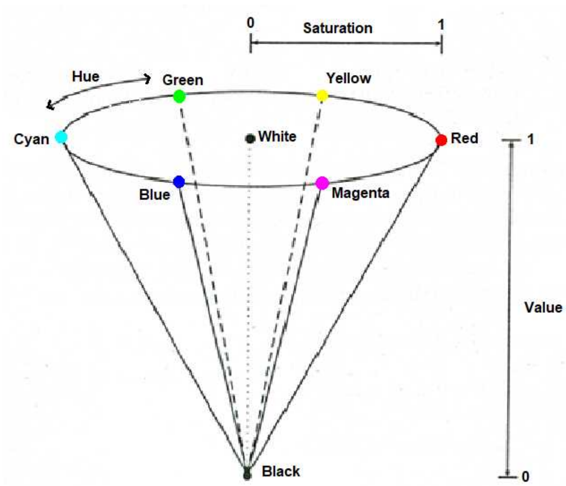


Figure 2.7: HSV space representation.

2.2.6 $l\alpha\beta$ color space

$l\alpha\beta$ perception-based logarithmic color space, proposed by [Ruderman et al., 1998], is expressed in the channels of luminance (l) and chroma ($\alpha\beta$). This color space was derived from a large ensemble of hyperspectral images of natural scenes using the first orders statistics of the images. It has some characteristics such as compactness, symmetry, and its principal characteristic is the decor-relation between components. The last one allows separating with some confidence in the absence of undesirable cross-channel artifacts [Reinhard et al., 2001]. The transform matrices are in the Appendix A.4. This is one of the spaces we considered using because the three channels are independently correlated it can help reduce color changes in the fusion.

Table 2.1: Summary of characteristics and applications of color spaces.

Color Space	**Characteristics**	**Aplications**
RGB	Orthogonal and hardware friendly, not uniform, based on trichromatic theory, high correlation between components.	Color monitors, computer graphics and color video cameras.
XYZ	Device independent color space, all the colors defined by X, Y and Z cannot be realized by actual color stimuli.	Colorimetric calculations, and is a reference color space for device calibration.
CIE L*a*b*	Device independent color space.	Industrial color measurement systems, evaluation in the color difference, and image analysis.
YCbCr	No uniform.	Widely used for image and video compression. Skin color extraction and classification.
HSV	User oriented.	Human visual perception, computer graphics, processing, computer vision, and image analysis.
$l\alpha\beta$	Symmetrical, and decorrelated.	Image analysis.

In this section, the specification of color in different spaces or systems has been indicated. The concept of color space has been defined as a codification of sensations chromatics in three-dimensional vectors that represent spatial coordinates. In digital image processing, there is no one optimal color space for all applications. The choice of a color space depends on the properties of the model and the characteristics of the application. Table 2.1 summarizes the spaces of color commented in this chapter and the applications for which they are used. Regarding obtaining chromatic coordinates of the models, it has been commented that most are derived from direct transformations of the RGB space. However, some spaces require intermediate use of XYZ color space.

2.3 Mathematical morphology

Mathematical morphology (MM), extended by Matheron and Serra [Serra, 1986], has been widely developed in image processing. MM is defined as the theory of spatial analysis of structures. The basis of MM is set theory operations such as union, intersection, complementation, etc. The notation for sets will be upper case (X, Y, Z, ...), and lower case the elements contained in it (x, y, z, ...). A morphological transformation $\Psi(f)$ is obtained by a relationship between the pixel values of the image f and the values of a subset called the structuring element ($B\lambda$). The shape (morphology) of $B\lambda$ is commonly chosen a priory by taking into account the morphology of the set on which it will interact, also, according to the extraction of shapes to be obtained. The shapes can be square, circular, linear, hexagonal, rhomboid, among others; Basic examples of structuring elements used in practice are shown in Figure 2.8. However, due to its ease of application, the square shape is commonly used [Fang et al., 2020]. The square shape is defined

(a) (b) (c)

Figure 2.8: Example of basic flat shapes of $B\lambda$: (a) circle, (b) square, (c) diamond.

as $B\lambda = (2\lambda + 1) \times (2\lambda + 1)$ where $\lambda \in \mathbb{N}$ is associated with the size of the transformation. At $\Psi_{B\lambda}(f)$ the origin of $B\lambda$ overlaps with the analyzed pixel in image $f(x)$, analyzing all the subyacent image pixels in f according to Ψ [Soille, 2013].

2.3.1 Dilation and erosion

The fundamental operations in MM are dilation $\delta_{B\lambda}$ and erosion $\varepsilon_{B\lambda}$. In the case of functions or gray level images the erosion of a set f by a structuring element $B\lambda$ is defined as shown in Eq. 2.2, where x_n is the value of the image at the point $x + k$, which means $f(x + k)$.

$$\varepsilon_{B\lambda}(f(x)) = max\,\{f(y) : y\epsilon B_{\lambda x}\} \tag{2.2}$$

Erosion and dilation have opposite effects on images. More formally they are dual by complementation. The dilation of a set f by $B\lambda$ is defined as (Eq. 2.3).

$$\delta_{B\lambda}(f(x)) = min\,\{f(y) : y\epsilon B_{\lambda x}\} \tag{2.3}$$

On the one hand, morphological dilation generates the growth of local regions with higher gray levels, allowing the connection of regions reached by $B\lambda$. The result of erosion in gray scale images is a darker image. On the other hand, morphological erosion generates a reduction of the regions, by setting them to a lower gray level, and also it increases the distance among them. For images in gray levels, δ generates an image perceived lighter than the original f; ε has the contrary effect [Haralick et al., 1987]. Figure 2.10 shows an example of the morphological erode and dilation operations applied to a gray image. Figure 2.10 (b) shows that when $\varepsilon_{B\lambda}\,(f)$ is applied to disconnected regions (i.e. space between petals), the size of these regions is also increased in λ pixels. The opposite case occurs when applying $\delta_{B\lambda}(f)$ (Figure 2.10 (c)).

(a) (b) (c)

Figure 2.9: (a) Original image (f), examples of (b) erode $\varepsilon_{B\lambda}(f)$, (c) dilation $\delta_{B\lambda}(f)$; where $B\lambda$ is a disk of radius 5.

Morphological erosion and dilation operations comply the following properties:

- **Duality:** These two operations are dual to their complement. In other words, an erosion is equivalent to the complement of the dilation of the image that has been complemented with the same structural element and vice versa.

$$\varepsilon_{B\lambda}(f) = [\delta_{B\lambda}(f^C)]^C \tag{2.4}$$

- **Increasing:** These two operations are increasing; they respect the order present in the structure of the lattice. For two images f and g.

$$\text{if } f \leq g \Rightarrow \varepsilon(f) \leq \varepsilon(g) \tag{2.5}$$

$$\text{if } f \leq g \Rightarrow \delta(f) \leq \delta(g) \tag{2.6}$$

  Increasing property in erosion and dilation implies that these operators respect the order relationships between the different stacked sets of an image.

- **Extensiveness and anti-extensivity:** Dilation is an extensive operation, and erosion is an anti-extensive operation; the structuring element must contain the origin:

$$f \leq \delta(f) \tag{2.7}$$

$$\varepsilon(f) \leq f \tag{2.8}$$

In general, for structuring elements that contain their origin, it is true that: $\varepsilon(f) \leq f \leq \delta(f)$.

2.3.2 Opening and closing

The morphological opening of a signal f by a structuring element $B\lambda$ is denoted by $\gamma_{B\lambda}(f)$ and is defined as the erosion of f, followed by dilation by the same structuring element:

$$\gamma_{B\lambda}(f) = \delta_{B\lambda}[\varepsilon_{B\lambda}(f)] \tag{2.9}$$

The morphological closing of a signal f by a structuring element $B\lambda$ is denoted by $\varphi_{B\lambda}(f)$ and is defined as the dilation of f followed by erosion by the same structuring element.

$$\varphi_{B\lambda}(f) = \varepsilon_{B\lambda}[\delta_{B\lambda}(f)] \tag{2.10}$$

An application of closing and opening could be removing unwanted objects from an image e.g. Figs. 2.10 (b-c) show the morphological opening and closing, respectively; Both allow to eliminate those components that cannot contain the structural element. The opening does it inside the function, during the closing on the complement of the function.

(a) (b) (c)

Figure 2.10: (a) Original image (f), examples of (b) opening $\gamma_{B\lambda}(f)$, (c) closing $\varphi_{B\lambda}(f)$; where $B\lambda$ is a disk of radius 5.

Properties of morphological opening and closing:

- **Duality:** The opening of an image is equivalent to the complement of the closing of the complemented image. This means that the opening and closing are dual operations with respect to the complementation:

$$\varphi_{B\lambda}(f) = [\gamma_{B\lambda}(f^C)]^C \tag{2.11}$$

- **Order relations:** The closing operation is an extensive operation, while the morphological opening is anti-extensive. This means that the following order relationship can be presented between the original image f and transformations:

$$\gamma_{B\lambda}(f) \leq f \leq \varphi_{B\lambda}(f) \tag{2.12}$$

- **Increasing and idempotent:** These two operations are increasing and idempotent. The idempotent property is important in image filtering, ensuring that iterations of the transformation will not modify the image. Given two images f and g.

$$\text{if } f \leq g \Rightarrow \gamma(f) \leq \gamma(g) \tag{2.13}$$

$$\text{if } f \leq g \Rightarrow \varphi(f) \leq \varphi(g) \tag{2.14}$$

2.3.3 Top-hat and bottom-hat

The top-hat ($Tw_{B\lambda}(f)$) and bottom-hat ($Bt_{B\lambda}(I)$) transformations are residual filters, which use a combination of opening and closing operations. Top-hat is defined as the difference between the image input and its opening.

$$Tw_{B\lambda}(I) = f - \gamma_{B\lambda}(I) \tag{2.15}$$

Bottom-hat is defined as the difference between closing and the input image.

$$Bt_{B\lambda}(I) = \varphi_{B\lambda}(I) - f \tag{2.16}$$

The top-hat transformation is useful to discovering those image structures that have been eliminated in the opening or closing (filtering that with the choice of a structuring element of suitable shape, size, and orientation, it is possible to filter the image and eliminate certain elements in the original image) [Meyer, 1978]. A difference operation between the original and the filtering greatly increases the contrast of the removed areas . Thus, $Tw_{B\lambda}(f)$ is used to highlight bright regions of the image that are smaller than $B\lambda$. In comparison, $Tb_{B\lambda}(I)$ is used to extract dark regions of an image that are smaller than the structuring element [Dougherty and Lotufo, 2003].

For instance, consider that the relevant features in an image are those that stand out from the background. Figure 2.10 (a) shows a well-contrasted image since the main objects are distinguished. A feature may seem irrelevant from a general context (whole image), but locally (regions) it could be important. The top-hat transform deals with the identification of relevant local features (such as within flower petals or grass). In this way, the top-hat transform extracts the relevant light features from the image. Similarly, the bottom-hat extracts the dark features. To achieve this, the morphological opening $\gamma_{B\lambda}$ flattens the relevant features according to the $B\lambda$ size, which evens out the local background. This process generates a higher difference in gray values between the relevant features and the new background from $\gamma_{B\lambda}$, e.g., the contrast between them increases. Figures 4.9(a-b) show the differences between $Tw_{B\lambda}(I)$ and $Bt_{B\lambda}(f)$, respectively, by containing the outstanding features of each region. An example of the use of these features is to increase the contrast between regions as shown in Figure 4.9 (c) [Hassanpour et al., 2015].

(a) (b) (c)

Figure 2.11: Example of applying: (a) $Tw_{B\lambda}(f)$, (b) $Bt_{B\lambda}(f)$, and (c) the image improvement by using bottom and top-hat $f + Tw_{B\lambda}$.

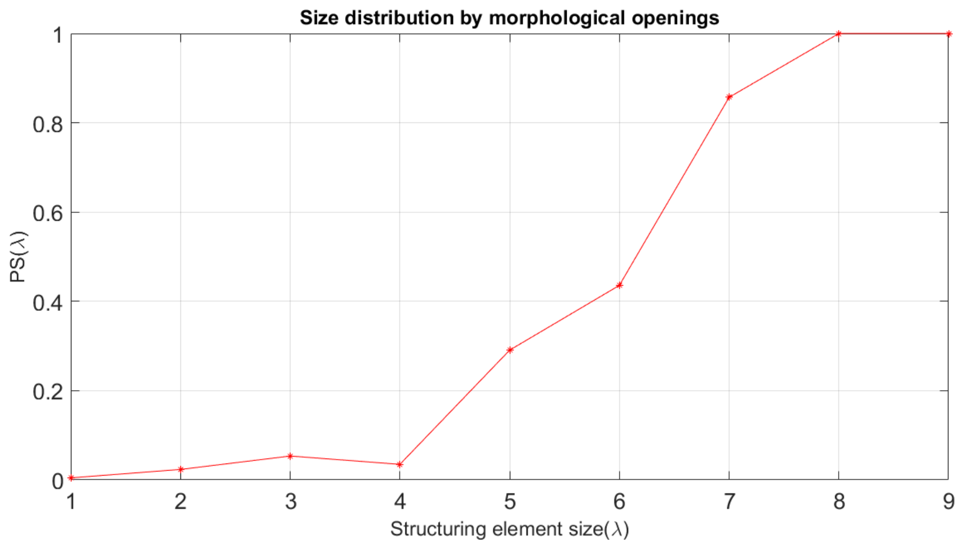
The choice of the structuring element used in top-hat operations depends on the morphology of the structures to be extracted, e.g., the bright objects detection in an image is achieved with a top-hat with a structuring element larger than the objects to be detected. On the other hand, the same utility is achieved with a bottom-hat in the case of dark objects.

2.3.4 Granulometry by openings

Granulometry and is used in several areas, such as imaging processing for feature extraction, size estimation, image segmentation, etc. The granulometric analysis of an image through a family of openings is an iterative process by increasing the scale of the transformation $\gamma_{\lambda+\Delta}(f)$ until it reaches the empty set. The measure associated with the granulometric is denoted Pattern Spectrum (PS) and is used to derive the granulometric curves (Eq. 2.17), where mes is the sum of all the pixel values of the image, and $\gamma_{\lambda}(f) = f$ when $\lambda = 0$. Thus, it is possible to measure the normalized area of the granulometric residues of the structures that are in the image [Santibañez, 2007, Vincent, 1994].

$$PS\,(f) = \frac{mes(\gamma_{\lambda}(f)) - mes(\gamma_{\lambda+\Delta}(f))}{mes(f)} \tag{2.17}$$

Where there is a high number of structures of λ size in f, the PS value is high; likewise, when the structures are few, the PS value decreases. Thus, the PS for a range of λ values suggests the most suitable structure sizes to use in a morphological transform. Figure 2.12 illustrates an example of the procedure followed to measure the particles size by using Eq. 2.17. The higher peaks indicate particles of similar sizes; in the example, the sizes are 3, 5 and 7.



(a)

(b)

(c)

(d)

(e)

Figure 2.12: (a) Granulometric analysis of: (b) the original image I and their openings with different sizes (c) $\gamma_3(f)$, (d) $\gamma_5(f)$ and (e) $\gamma_7(f)$.

2.4 Quality assessment for image fusion

The concept of image fusion has different definitions in the literature. Image fusion is a technique that allows combining information from several sources and preserving the detail of each to improve the quality of the information [Kaur et al., 2021]. Then [Bai, 2012] extends the image fusion and also refers to image fusion algorithm as ”The effective image fusion algorithm should extract the useful image regions and details for image fusion well, so that the final fusion image is clear and contains more image details”. Several quantitative methods and functions have been proposed for the evaluation of images resultant of fusion methods. These objective evaluation methods come in different types according to what they want to evaluate or the theory they take. Structural similarity index (SSIM) proposed by [Zhou Wang and Bovik, 2002] is used for measuring image quality. A higher value indicates better image quality. Root Mean Square Error (RMSE) denotes the dissimilarity between the fused and visible images. A small RMSE metric indicates that the fused image has a small amount of error and distortion [Jagalingam and Hegde, 2015]. The SSIM and mean squared RMSE are used to indicate the level of the resemblance of the final output to that of the original image [Kumar et al., 2019, Ma et al., 2019]. The peak signal-to-noise ratio (PSNR) metric is the ratio of peak value power and noise power in the fused image and thus reflects the distortion during the fusion process. The larger the PSNR, the closer the fused image is to the source images and the less distortion the fusion method produces [Ma et al., 2019, Jagalingam and Hegde, 2015]. The mutual information metric is a quality index that measures the amount of information that is transferred from source images to the fused image. A high mutual information value means that considerable information is transferred from source images to the fused image, which indicates a good fusion performance according to [Qu et al., 2002]. Table 2.2 shows some representative fusion evaluation measures and the references where they are used. In some cases, it is required to measure the characteristics of the original image to evaluate the improvement achieved by the fusion method applied [Ma et al., 2019]. The functions mentioned above are used in state-of-the-art works related to the fusion of visible and NIR. However, it is considered that when looking for an improvement, the image is not completely similar to the original image, as in the case of the Mutual information, PSNR, RMSE, and SSIM.

One of the problems while fusing images is preserving the images edges, which means it must preserve the important details of the components of the source images in the fused image. As mentioned above, these functions assess the quality of the fused information. However, in the particular case of color images, it is crucial to keep the information that color gives as this is the key to interpreting the image. Color is a descriptor that often is useful for object identification and extraction from a scene , then it is also important to assess the effect of image fusion over it [Russ and Neal, 2015, Gonzalez Rafael C. et al., 2009]. Therefore, it is relevant to assess and measure the added information and color distortion for image fusion methods. Each author defines how to measure image improvements; there is no standardized metric to measure these improvements. So we opted for the entropy and contrast functions to measure the information and quality of the image and the saturation and color similarity to measure the differences in color of the merged image with respect to the original. Therefore, it was decided to measure the quality of the images described below.

Table 2.2: Representative fusion evaluation measures and references.

Evaluation measure	**Description**	**Formula**	**Uses in**
Feature mutual information ↑	To measure information.	Appendix B.1	[Roberts et al., 2008]
Structural similarity index measure (SSIM) ↑	To measure the amount of edge information.	Appendix B.2	[Kumar et al., 2019, Jung et al., 2020]
Anisotropic Quality Index ↑	To reflects the distribution and contrast of the fused image: standard deviation.	Appendix B.3	[Vanmali et al., 2015, Vanmali and Gadre, 2017]
Mean squared error (MSE) ↓	To measure the dissimilarity, spatial frequency measure detail and texture of an image.	Appendix B.4	[Sharma and Gool, 2016, Sharma et al., 2017]
Peak signal-to-noise ratio (PSNR) ↓	To measure the distortion during the fusion process.	Appendix B.5	[Sharma et al., 2017, Park, 2020]
Entropy ↑	To measure information.	Equation 2.19	[Zhang Jingyun et al., 2016, Vanmali et al., 2015]
Contrast ↑	To measure the contrast.	Equation 2.18	[Son et al., 2015, Jang and Park, 2017]
Saturation =	To measure the level of saturation.	Equation 2.21	[Jang and Park, 2017]
Colorfulness =	To measure the color content in an image.	Appendix B.6	[Son et al., 2015]

*Value to look for fusion results(Higher ↑/Lower ↓)

2.4.1 Added information

Contrast is defined by [Burger and Burge, 2016] as ”a combination of the range of intensity values effectively used within a given image and the difference between the maximum and minimum pixel values”. This difference allows distinguishing objects in the background of an image.

$$C = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M}(f(x,y))^2 - \left(\frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M}f(x,y)\right)^2 \tag{2.18}$$

Equation 2.18 computes the deviation of gray levels where $f(x, y)$ is the intensity image. The more uniform the gray distribution of the image, the higher the value of C, indicated the more uniform the gray distribution of the image would be; therefore, the better the contrast of the image [ZHANG Chang Jiang, 2004]. It is expected that the fusion method generates a significant difference between the local values and the global representative value, which indicates an increase in contrast [Shi et al., 2017]. When an image has high contrast, areas are better differentiated. Therefore, in a low contrast image it is difficult to distinguish between objects. Hence, when performing a image fusion process, the contrast value of the resulting image must be greater than the value of the original image to assume improvement.

Shannon’s entropy measures the uncertainty of a source of information and is usually used to quantifies the image information content [Tang and Mat Isa, 2014]; entropy is defined as Eq. 2.19 where the term $log_2(P(g))$ implies the amount of gray values information obtained from an image with probability $P(g)$ of a pixel with gray level g and G is the maximum possible gray level (e.g., 255 for an 8-bit image). In this way, an image with a large number of pixels with the same intensity will have a low entropy, while with a more uniform distribution of gray values, a high entropy will be obtained. A high EN value (uncertainty, disorder) indicates higher information content [Ma et al., 2019, Son and Zhang, 2017]. For instance, a fog region in the image is associated with low quality since the similitude of the values reduces the contrast among the structures. Therefore, these results in low information content or low entropy in the region.

$$EN = -\sum_{g=0}^{G} P(g)log_2(P(g)) \tag{2.19}$$

2.4.2 Color similarity

There are many discrepancies in brightness and image structures between the NIR images and the visible color images. Due to this discrepancy, some color distortion problems emerge during V-NIR fusion. I.e., the original color significantly changes due to the image fusion, then false or oversaturated colors are generated [Son and Zhang, 2018, Elliethy and Aly, 2017]. Therefore, to measure how much the colors change after a fusion process the functions of color similarity (ΔE_{00}) and saturation ($\hat{S}$) are used. These functions are explained bellow.

Color similarity denoted as ΔE_{00} was defined by CIE L*a*b* 2000 where the Greek letter delta is used in mathematics to denote difference and the E comes from the German term Empfindung or Sensation. ΔE_{00} is used to measure the difference between two colors designated as two points in the CIE L*a*b* color space, as shown in Eq. 2.20 [Sharma et al., 2005]. The lightness ($\Delta L'$), chroma ($\Delta C'$) and hue ($\Delta H'$) differences of two color parameters being compared. Where K_L, K_S and K_H are parametric coefficients to compensate for the interference of external factors in the perception of color difference, with $K_L = K_c = K_H = 1$ for standard lighting conditions [Mokrzycki and Tatol, 2011]. While, S_L, S_c and S_H are compensation coefficients. Additional functions for calculating this metric are described in the formulas in Appendix B.7.

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{K_L S_L}\right)^2 + \left(\frac{\Delta C'}{K_c S_c}\right)^2 + \left(\frac{\Delta H'}{K_H S_H}\right)^2 + R_T \left(\frac{\Delta C'}{K_c S_c}\right)\left(\frac{\Delta H'}{K_H S_H}\right)} \tag{2.20}$$

The range of values for ΔE_{00} is from 0 to 100. The higher the value of ΔE_{00}, the lower the color equality, i.e. 100 indicates the two colors are opposite. According to CIE L*a*b*, the ΔE_{00} falls into five different ranges:

- $0 < \Delta E_{00} < 1$: the difference is not noticeable.
- $1 < \Delta E_{00} < 2$:the difference is only noticed by an experienced observer.
- $2 < \Delta E_{00} < 3.5$: the difference is noticed by an unexperienced observer.
- $3.5 < \Delta E_{00} < 5$: the difference is clearly noticeable.

- $5 < \Delta E_{00}$: these are two different colors.

Figure 2.13 shows examples of colors expressed in RGB, and the ΔE_{00} values. The first color pair varies in the component B from 137 to 131 the new color obtained a $\Delta E_{00} = 1.32$. This result means that the difference is only noticed by an experienced observer. In the following case, the value of $\Delta E_{00} = 2.12$, which even an unexperienced observer might notice. At the last example the is $\Delta E_{00} = 5.51$ where the observer note two different colors.

Figure 2.13: Color similitude between pairs of colors according to ΔE_{00}. From top to bottom: $\Delta E_{00} = 1.32$, $\Delta E_{00} = 2.12$, and $\Delta E_{00} = 5.51$ [Mokrzycki and Tatol, 2011].

Saturation refers to the intensity of a specific hue and is based on the purity of the color. A high saturated color has a vivid and intense color appearance while a low saturated color appears mute or grayish which means it is less colorful [Shi et al., 2017]. In the case of V-NIR image fusion it is expected that the saturation of the fused image is close to the original image. Thus, a saturation metric (Eq. 2.21) can be used to know in what extend the image fusion affects the color in the resulting image; in this metric, S is the saturation channel from the HSI color space.

$$\hat{S} = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} S(i,j) \tag{2.21}$$

In this section, the quality evaluation functions of the images resulting from V-NIR fusion were reviewed. In particular, the quantity and quality of the information and the measure of the difference in the color obtained. Evaluation functions are generally complemented by a subjective evaluation, for example, the amount of information. Since not all the fuse information is necessary information, artifacts could appear in the images that, although they increase the amount of information, visibly do not correspond to the structures of the images.

In this chapter, the basic principles to understand this thesis were explained. The main characteristics of visible and NIR images were described. Then the color spaces were mentioned with particular attention to the ones that separate the luminance from chroma. Then the principles of mathematical morphology were described, emphasizing that the top-hat and bottom-hat transformations were applied in the proposed methodology. An essential part of these transformations is the choice of the size of the structuring element, which is why the granulometry technique to select a size is described. Finally, a compendium of the main evaluation functions for images was made, focusing on information measurement and over-saturation functions.

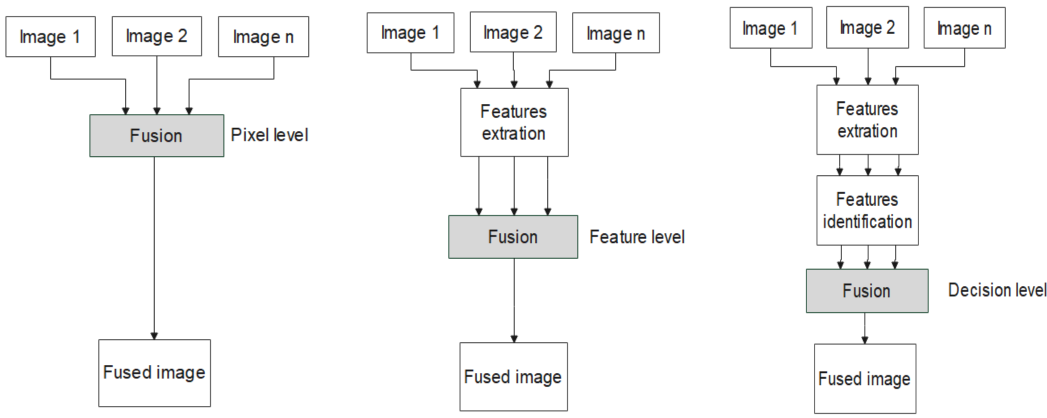
Chapter 3

V-NIR fusion methods

In this chapter, the related works associated with V-NIR fusion techniques are reviewed and discussed. Outdoor color images can be deteriorated by factors, mostly by weather conditions, that cannot be controlled, such as haze, lighting or fog. These factors may affect the visualization of edges, contrast, saturation. It is where the image fusion takes place to improve the quality of the images. As mentioned above image fusion integrates the complementary multi-temporal, multi-view, or multi-sensor information into a single image. In this way, image quality is improved and original features are keep.

Image fusion methods are divided according to three levels: pixel level, feature level, and decision level (Figure 3.1). The pixel-level techniques directly integrate the information from input images while feature-level techniques extracts relevant features, e.g., textures or edges that are compounded to create supplementary merged features; in decision level techniques, the input images are processed one at a time to extract information [Li et al., 2017, Kaur et al., 2021]. All the included works consider that source images, visible and NIR, were taken simultaneously from the same scene, and at the same time, the fusion process is multi-sensor. Most of the V-NIR fusion methods presented below are performed at pixel-level, combining pixels from source images to obtain the fused image.

The reviewed works take into account three main approaches: i) Fusion by mapping: these methods are based on the calculation of maps using the characteristics of the images, such as the level of haze, fog, or contrast level. ii) Multi-scale transform: these methods decompose the source images into n levels (layers), fuse corresponding layers with fusion rules, and reconstruct the target images. iii) Other approaches for V-NIR fusion: the methods that correspond to this category use mixed techniques or techniques not included above, such as CNN, fuzzy wavelet.



(a) (b) (c)

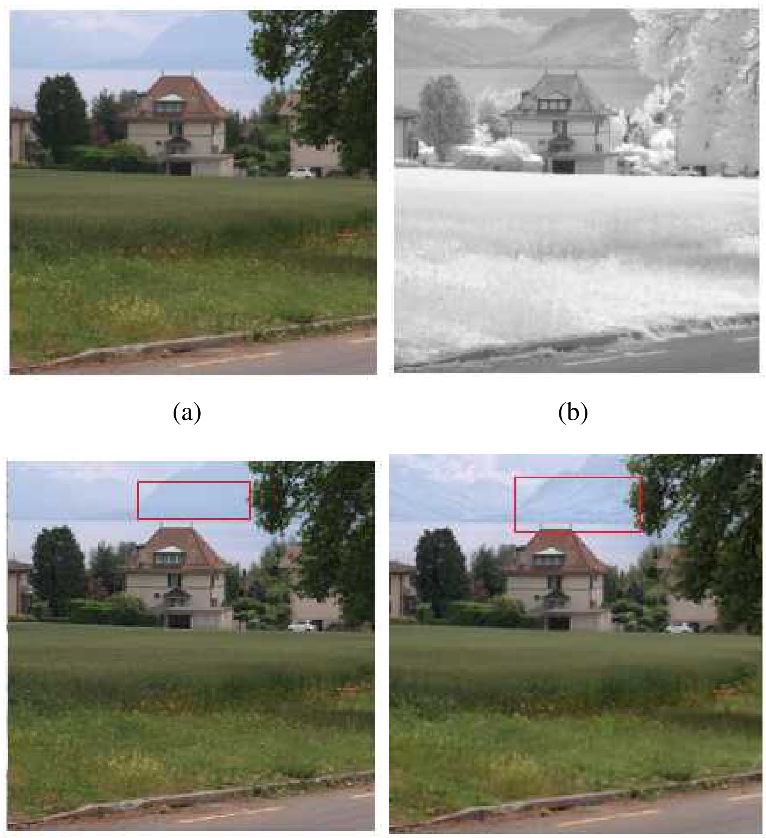
Figure 3.1: Image fusion according to the processing level (a) Pixel level, (b) feature level, and (c) decision level.

3.1 Fusion by mapping

This approach consists in taking the average of the two source images pixel by pixel. However, when this approach is performed directly, The information can be overlap; this causes the presence of artifacts or the deletion of the information. Therefore, it has been proposed to make fusion maps that search in the source images for some distinctive characteristics to fuse, such as contrast, haze, or salience maps. Those maps allow a weighted fusion, that take a linear weighted average of the two source images as shown in Eq.3.1, where W_A,W_B are the given weights for the source images $A(n,m)$ and $B(n,m)$ [Smith and Heather, 2005]. Next, some works that attempt to V-NIR fusion based on by mapping are discussed bellow.

$$I(n,m) = W_A A(n,m) + W_B B(n,m) \tag{3.1}$$

The method [Son et al., 2015] defines a relationship between NIR and visible through a preservation-contrast map to generate the fused image. This mapping model is used to correct the chrominance distribution of the visible images. The results of this method is shown in Figure 3.2 (c). Its main contribution is to avoid distorting the color during the fusion process; however, the results obtained when compared with other methods such as [Sharma et al., 2017] (Figure 3.2 (d)), it can be seen that it could fuse more information, also Figure 3.2 (a) and Figure 3.2 (c) seems similar.

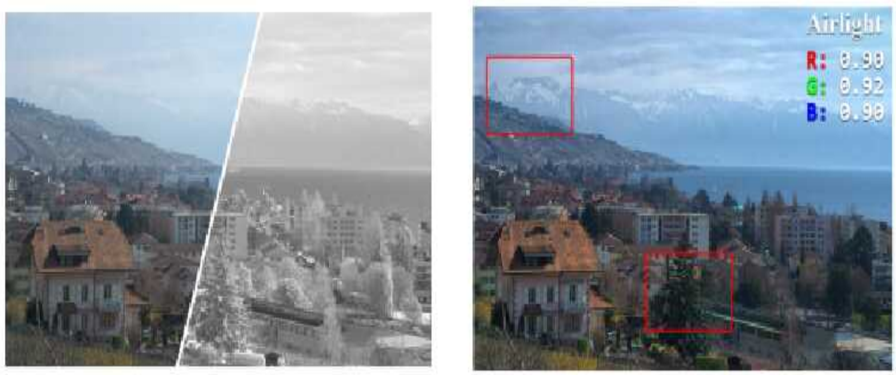


(a) (b)

(c) (d)

Figure 3.2: (a) Visible image, (b) NIR, and the result of the method proposed by: (c) [Son et al., 2015], (d) [Sharma et al., 2017]. Red rectangles show a particular area where different information can be found.

Figure 3.3 (b) shows the results of the proposed method by [Feng et al., 2013]. The method first consists of estimating the color of the aerial light, taking the differences between RGB and NIR to generate a haze map. Later they carry out a dehazing stage by enforcing the NIR gradient constraint through an optimization framework. From this approach, it is possible to fuse information corresponding to the NIR specialty to the zones in the red squares. The authors do not present any metrics or quantitative comparisons, and the image shown is the only one that they present in their results. So there is little that can be said about this method regarding its results. The only apparent backward could be Figure 3.3 (b) is more saturated than the original image Figure 3.3 (b).



(a) (b)

Figure 3.3: (a) Source visible and NIR images, and (b) result of the method proposed by [Feng et al., 2013].

Another method that proposed a haze map is [Zhang Jingyun et al., 2016] to addresses the fog problem. First, infrared-blue light intensity difference factor and dark channel prior* make a haze distribution map. Then according to the map, NIR information and visible information are used. Finally, a guide filter, fast fuzzy algorithm, and downsampling are applied. Their results show that the methods that use only the priority dark channel have less Entropy and Average gradient than when adding the optimization part composed of fast fuzzy algorithm + downsampling + fast guide filter.

The method overview proposed by [Elliethy and Aly, 2017] is presented at Figure 3.5. This method consist first in calculated the local contrasts for the NIR image I^{NIR} and the luminance plane (Y) of the visible image I^{RGB}. Then, the spatial details from the NIR are extracted using a high pass filter. Finally, the spatial details are weighted according to a fusion map (F)

*For outdoor images, in the non-sky local area, at least one color channel has very low intensity at some pixels and those pixels can form a so-call dark channel [He et al., 2009] to obtain the fused image (J^{RGB}). In their results, the authors are compared qualitatively with [Fredembach and Süsstrunk, 2008] and [Son et al., 2015]. The authors mention that the fused images ”show a great enhancement” to the scene details compared to the original preserving the spectral contents of visible, in addition to not modifying the color.

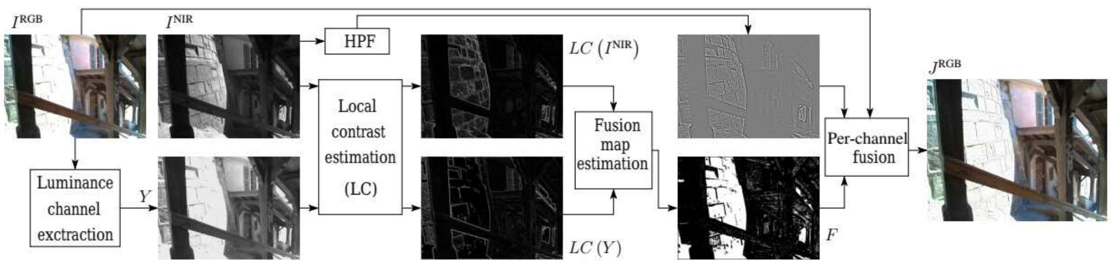
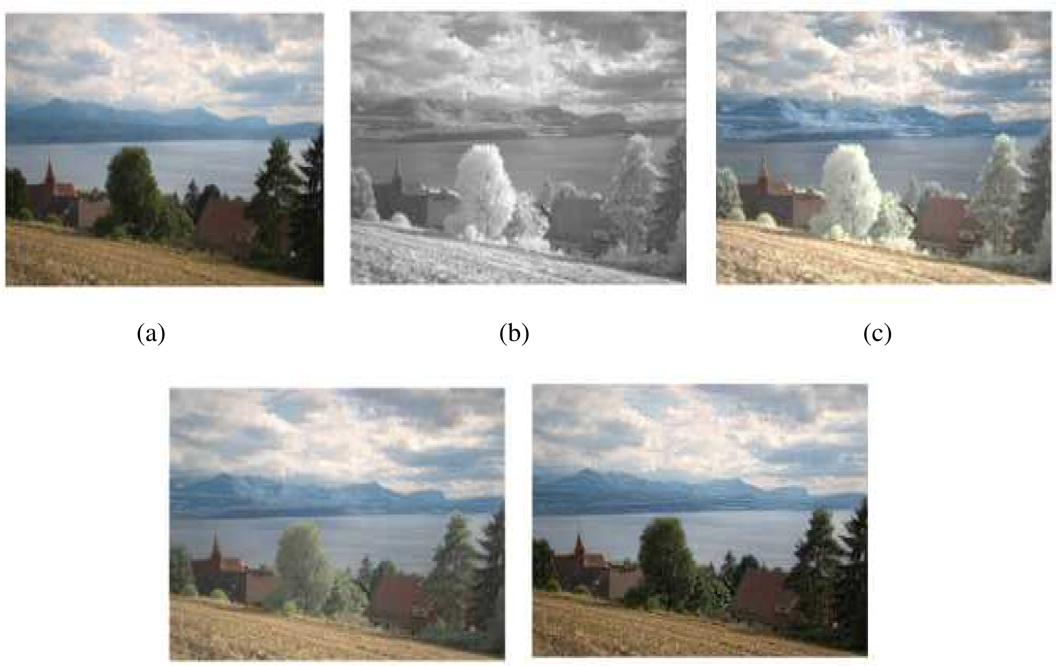


Figure 3.4: The overview of the method proposed by [Elliethy and Aly, 2017].



(a) (b) (c)

(d) (e)

Figure 3.5: (a) Original, (b) NIR and the proposed method by: (c) [Fredembach and Süsstrunk, 2008], (d) [Son et al., 2015], and (e) [Elliethy and Aly, 2017].

The main drawback in fusion by mapping is the calculation of the maps since this depends on the elements, variables or considerations taken to calculate the maps. Most of them physical considerations, such as haze distribution or the color of the aerial light. Thus, in the case of omission of some consideration in the map calculations, it may not be possible to rescue all the information belonging to source images.

3.2 Multi-scale fusion approaches

Methods that do not use maps to perform the fusion are presented below, e.g., fusion techniques use a multi-resolution signal decomposition scheme, such as Laplacian pyramid and Gaussian pyramids, edge-preserving filters, and weighted least squares (WLS). Multi-scale V-NIR image fusion approach comprises three steps (Figure 3.6). First, the source images are decomposed into a series of multi-scale representations. Then, the multi-scale representations of the source image are fused according to a given fusion rule. Finally, the fused image is done by using the corresponding inverse multi-scale transforms on the fused representations [Kaur et al., 2021].

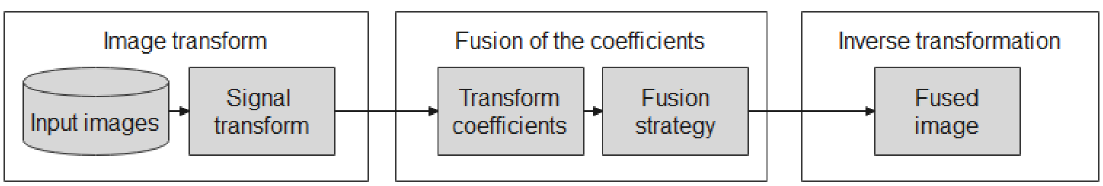


Figure 3.6: Generic multi-scale decomposition based image fusion scheme.

3.2.1 Laplacian and Gaussian pyramids

A pyramid is a sequence of images where each image is filtered, with a low pass filter, representing a subsampled copy of its predecessors. Figure 3.7 shows the relationship between Gaussian and Laplacian pyramids. Where Gaussian pyramid is a set of images $\{F_L\}$, called levels, which represent progressively lower-resolution versions of the image G. Thus, the pyramid base contains the image with the highest resolution ($F_0 = G$), and the lowest resolution is found at the top. $F_L + 1 = REDUCE(F_L)$ is a low value pass version of F_L with half the weight and height. The filtering and the decimating process are repeated n times.

The Laplacian pyramid is related construction, whose levels $\{L_L\}$ represent details on different spatial scales, decomposing the image into separate frequency bands. The levels of the Laplacian pyramid are defined by the details that distinguish the successive levels of the Gaussian pyramid, $L_L = F_L - EXPAND(F_L + 1)$, where $EXPAND(\cdot)$ is an operator that doubles the size of the image in each dimension using a smooth kernel. The highest level of the Laplacian pyramid, also called the dual residue, is defined as $L_n = F_n$ and corresponds to a tiny version of the image. To reconstruct the original image by recursively applying $F_L = L_L + EXPAND(F_L + 1)$ until $F_0 = G$ is recovered [Burt and Adelson, 1987].

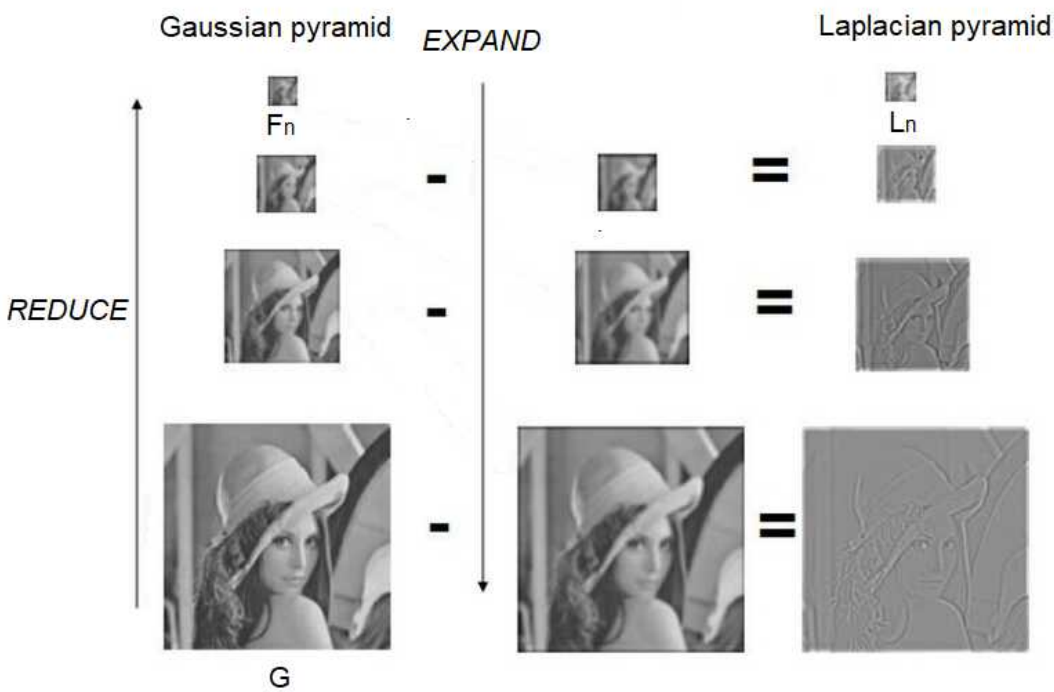


Figure 3.7: Relationship between Gaussian and Laplacian Pyramids.

[Vanmali and Gadre, 2017] use a Laplacian–Gaussian pyramid based multi-resolution fusion process, guided by weight maps generated using local entropy, local contrast and visibility as metrics to control the fusion. The method proposed by [Jang and Park, 2017], first decomposes the input images by a Gaussian low pass filter. Then, to carry out the weight map from a guided filter proposed by [He et al., 2009], which gives rise to what the authors call a fusion layer. Finally, the fusion is carried out by joining the layer without fog with the layer of details. Its contribution is to increase the details in the hazy regions without distorting the color and producing artifacts at the edges. However, after observing the results presented (Figure 3.8), It is visibly noticeable that could be retrieved more information.

The pyramids are commonly regarded as poor for applications in which image edges play an important role. Thus, the use of edge-preserving filters is used to prevent blurred edges at fusion. The bilateral filter in its direct form has the limitation on simultaneously edge-preserving and image smoothing. It can introduce various image artifacts, e.g., the introduction of false edges in the images. Also, it can produce halos around some edges due to unwanted smoothing of these edges [Zhengguo Li et al., 2013].

3.2.2 Edge-preserving filters

The aim of edge-preserving filter is to preserves the edges and their information while blurring an image. The base layer is obtained using an edge-preserving filter, e.g., bilateral filter [Tomasi and Manduchi, 1998], anisotropic diffusion filter [Perona and Malik, 1990], guided filter, and WLS. On the image, edge-preserving filter potentially captures the intensity changes. The detail layers comprise a series of different images that can preserve details at various progressively scales. For more detailed information on edge preservation filters refers to [Pal et al., 2015]. The

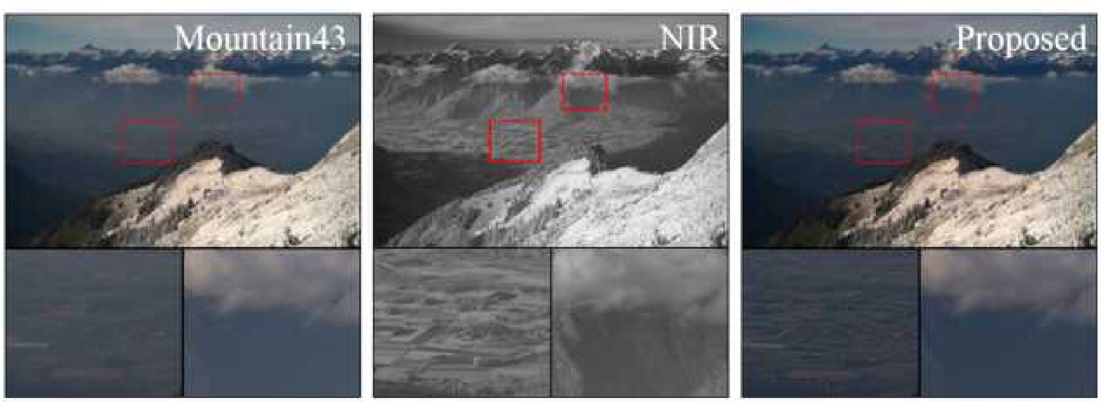


Figure 3.8: Result of the method proposed by [Jang and Park, 2017]

method proposed by [Kumar et al., 2019] is based on an edge-preserving smoothing filter with a multiresolution decomposition based on contrast images. The results of this method (Figure 3.9) show a desaturation, which is reflected as a change of color.

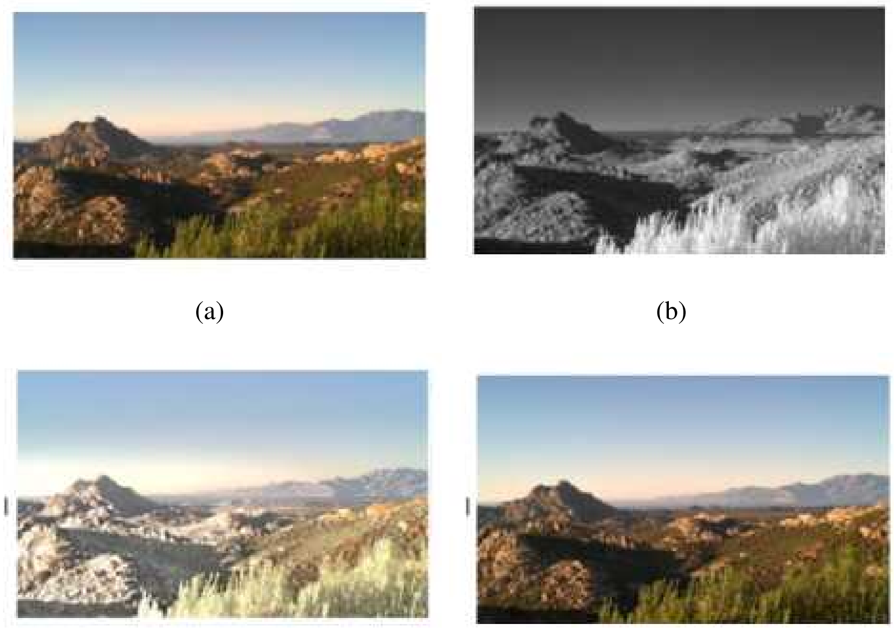
(a) (b) (c)

Figure 3.9: (a) Original, (b) NIR, (c) results of the proposed method by [Kumar et al., 2019]

Weighted least squares (WLS)

An approach that use WLS is proposed by [Schaul et al., 2009]. Figure 3.10(c) shows the results of this proposed method. The fusion of the NIR detail information and the color images is done using the WLS framework using a multiresolution approach. Results obtained show less haze and with more details compared to the original. However, the authors mention that colors might change radically in extreme luminance changes due to only work with luminance. Also, [Park, 2020] proposed a V-NIR image fusion method that decomposed the visible and NIR images by using WLS into detail and base images, respectively. Nevertheless, the NIR detail image was obtained by boosting in an analytical manner, which can change the result according to this factor. Figure 3.10 (d) shows the results of these method. The main advantage of this method is not to need calculated haze, airlight detection, or the generation of depth maps. Therefore no assumptions are needed to generate a correct depth map. However, there are still some limitations, such the manual boosting part.

Bilateral filters (BF) uses two kernel filters: Spatial Kernel is the distance between the image pixels (Euclidean distance), and range kernel is the intensity similarity between two pixels in the image. [Sharma and Gool, 2016, Sharma et al., 2017] combined a WLS filter with a BF. Where BF filter preserves edges and can extract details at a fine spatial scale but cannot extract details at arbitrary scales. At the same time, the WLS filter is used to preserve fine details at arbitrary scales.



(a) (b)

(c) (d)

Figure 3.10: (a) Original, (b) NIR, and results of the method proposed by: (c) [Schaul et al., 2009] and (d) [Park, 2020] using WLS filter.

Taking an average between the two filters gets the details from both. The method proposed by [Ahn et al., 2011] also uses a BF in a multi-resolution approach obtaining fused images enhanced with detail information and high image contrast than the visible one. However, the main drawback is the smooth of the borders that make lose information about the limits of the regions.

3.3 Other approaches for V-NIR fusion

This section shows methods that cannot be classified as the previous ones since their approaches are hybrid or are not based on maps or multiscale filters. The [Kudo and Kubota, 2018] method uses an hybrid method. The method uses a transmission map where the amount of the light that is not scatted and reaches by the camera is set, using the method of [He et al., 2009]. This map is used to give weights to the visible image. To later decompose the luminance and NIR image using a Laplacian pyramid. The authors show as the main contribution that they do not overemphasize the contrast in the haze-free regions. However, results show that they cannot wholly remove the haze, which is an opportunity for improvement. If the NIR images contain information without haze, it seeks to recover as much information as possible with the fusion.

[Jung et al., 2020] proposed a net for fusion of RGB and NIR images based on two stage CNN as can be see at Figure 3.11. The method focuses on removing noise from visible images through fusion V-NIR. Tor training set, added Gaussian noise to the images and used the original images as ground truth. However, CNN has challenges when applied to image fusion. First, training a network requires high-labeled data. Second, the artificially designed image fusion rules are challenging to realize the end-to-end model network. Thus, some errors will be mixed in the feature reconstruction process, affecting the image's feature final reconstruction [Sun et al., 2020].

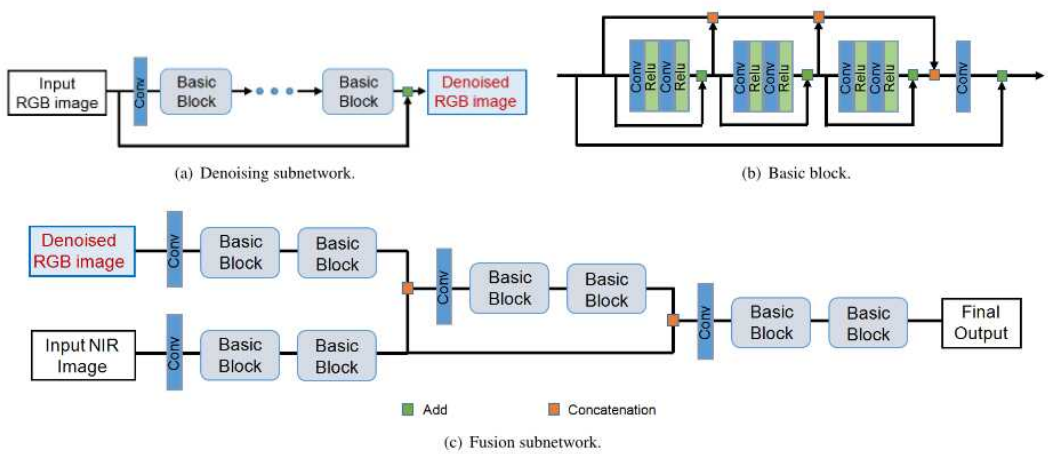


Figure 3.11: Proposed method by [Jung et al., 2020].

[Fredembach and Süsstrunk, 2008] propose two approaches; the first approach separates the visible and the NIR images according to their frequencies. Then, a low-pass Gaussian filter is applied to the images to obtain their low frequencies. High-frequency images are calculated as the original image minus its low frequencies. The fusion is performed once the high and low-frequency images have been found for the visible and NIR images. The second approach uses wavelet transforms. First, the authors decompose the luminance of the visible and NIR images using symmetric 8-band wavelet filter and 8 levels of the wavelet decomposition. Then, the wavelet coefficients of each decomposition for each pixel, the maximum coefficient between the NIR and the visible image for the high frequencies, and averaged these coefficients for the low frequencies parts of the image. The improvement is not perceived; most regions are not affected, or too substantial, color changes.

Figure 3.12: Results of the proposed method by [Fredembach and Süsstrunk, 2008].

Table 3.1 summarizes the comparisons of the methods of state of the art. Most of them separate the chroma from the luminance and work only with luminance. As an exception, [Jang and Park, 2017], decided to fuse the NIR directly with the red channel of the RGB color space. Even though several methods mention some quantitative measures, the results of the experiments obtained indicate them based on a few images, from 1 to 5 images such as [Jang and Park, 2017, Son and Zhang, 2017]. Some suggest that their quantitative measures were made to a set of images and give the best results ([Son et al., 2015]). Or the authors provide a value that could be interpreted as an average ([Zhang Jingyun et al., 2016]). In any of these cases, the information is not sufficient for statistical comparison. Of all the methods reviewed, only [Son et al., 2015] evaluates the color by using the colorfulness metric, proposed by [Hasler and Susstrunk, 2003], but does not mention the value obtained. As seen during this chapter, some methods over-saturate or de-saturate the images; this causes the color changes to the original image. Therefore, part of the limitation for the design of the method is to avoid this color change. To make a quantitative comparison in the results section, it was necessary to replicate some methods. Then analyzed using the same database and compared with the selected quantitative functions.

Table 3.1: Review of methods reported in literature.

Work	**Method**	**Color space**	**Quantitative functions or metrics reported**
[Kudo and Kubota, 2018]	Transmission map	*NR	NR
[Ahn et al., 2011]	Local contrast estimation map	NR	NR
[Jang and Park, 2017]	Weight map	RGB	Contrast, Saturation and Restored edges
[Feng et al., 2013]	Haze distribution map	NR	NR
[Zhang Jingyun et al., 2016]	Haze map	YUV	Entropy, Gray variance and Average gradient
[Son et al., 2015]	Contrast-Preserving Map	$l\alpha\beta$	Colorfulness, Spatial frequency, Entropy and Contrast
[Son and Zhang, 2017]	Contrast-Preserving Map + Detail layer transfer	$l\alpha\beta$	Colorfulness, Spatial frequency, Entropy and Contrast
[Elliethy and Aly, 2017]	Local contrast Map	YcbCr	NR
[Sharma et al., 2017]	Laplacian-Gaussian pyiramid to generate weight maps	YCbCr	RMSE and PSNR
[Sharma and Gool, 2016]	BF	YCbCr	RMSE and PSNR
[Ahn et al., 2011]	BF	NR	NR
[Schaul et al., 2009]	WLS	NR	NR
[Park, 2020]	WLS	YCbCr	PSNR
[Vanmali and Gadre, 2017]	weight maps, calculated on the basis of local entropy, local contrast and visibility, Laplacian–Gaussian pyramid fusion	HSV	Entropy, Quality index, Quality-aware clustering, Edges, Visual information fidelity
[Fredembach and Süsstrunk, 2008]	Gaussian filter and the fast Fourier transform	HSV/ YCbCr	NR
[Vanmali et al., 2015]	Depth map and airlight estimation	HSV	Entropy and Quality index
[Jung et al., 2020]	CNN	CIE L* a* b*	PSNR, SSIM,BIQE
[Kumar et al., 2019]	Fuzzy-wavelet	CIE L* a* b*	Contrast restoration, Newly visible edges, Naturalness Image Quality Evaluator(NIQE), Blind/ Referenceless Image Spatial Quality Evaluator (BRISQUE), SSIM and MSE

***NR=No reported.**

Chapter 4

Proposed method

The proposed method aims to complement the visual image with information of the structures obtained from the NIR image through a selective fusion process that avoids color over-saturation. The proposed approach is designed based on the following assumptions: i) commonly, in outdoor images, there are significant differences between the content highlighted in the visible and NIR bands that can improve the information of the visible; ii) the added information is selected through a criterion to assess if it is relevant for being fused; iii) the source images, V and N, are properly spatially and temporally aligned. The general fusion method is described in four phases or main

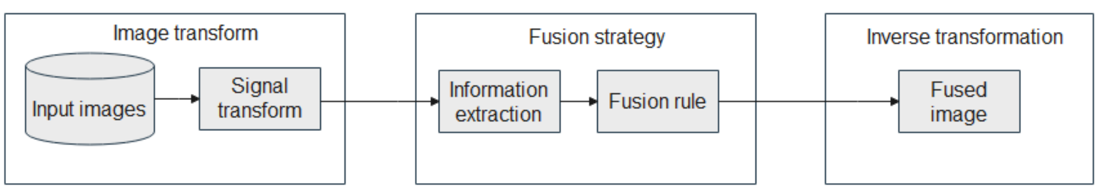


Figure 4.1: General description of the V-NIR fusion method.

stages as shown in Figure 4.1.

1. **Image transform:** To carry out the image fusion, it is proposed to separate the chroma from the luminance of the visible image (V) in order to work with the latter without affecting the original colors significantly. In V-NIR fusion, these color spaces have been used:

YUV [Zhang Jingyun et al., 2016], YCbCr [Elliethy and Aly, 2017, Sharma et al., 2017, Fredembach and Süsstrunk, 2008], HSV [Vanmali et al., 2015, Vanmali and Gadre, 2017, Fredembach and Süsstrunk, 2008], and $l\alpha\beta$ [Son et al., 2015]. The one decided to use is the $l\alpha\beta$ given its low correlation between the channels. This transform is done by converting RGB to $l\alpha\beta$ color space using the equations presented by [Reinhard et al., 2001].

2. **Information extraction:** The information is extracted selectively using a difference map or morphological transformations. This selection allows only relevant information to be brought into the fusion and avoids unwanted artifacts.

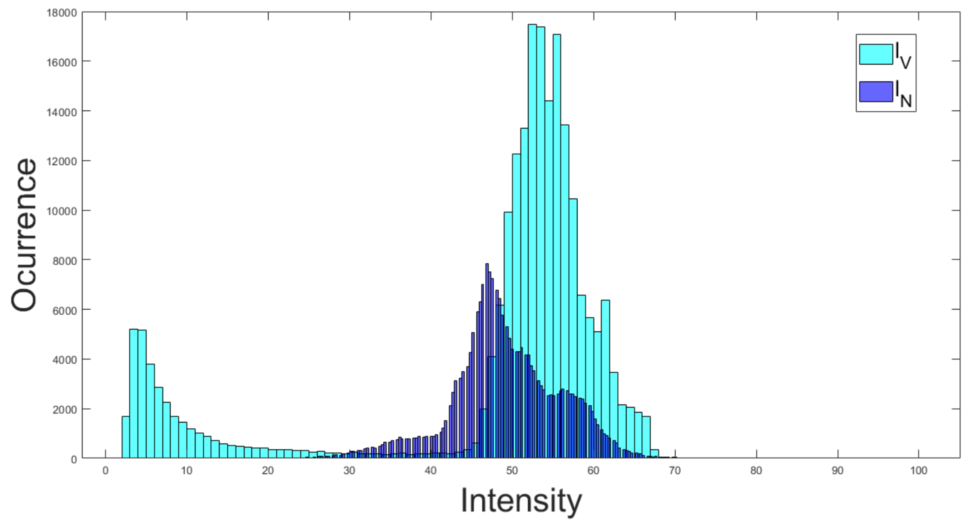
3. **Fusion rule:** Once the information is obtained from the sources, a fusing criterion is proposed to establish the combination or proportion in which this information will be fused. Not all the proposed criteria give the same result.

4. **Inverse transformation:** Finally the information is fused in the luminance band, an inverse transformation of the color space is carried out from $l\alpha\beta$ to RGB to get the final fused image and visualize the results.

4.1 Visible luminance and NIR distributions

This section analyzes the luminance and NIR distributions to observe the characteristics as help to propose a way to extract the information. By analyzing the luminances of V and N and plotting them in a histogram, complementary distributions can be found as shown in Figures 4.2 (a-b) the red squares remarks the areas where complementary information can be found. Figure 4.2 (c), the distribution of the V image is concentrated on the high-intensity values since the image has a whitish appearance associated with the presence of fog in the environment. On the other hand, the N distribution covers a wider range of V. Since the NIR image is less affected by weather conditions, it distinguishes the structures present in the fog region from the visible bands. Therefore, there is a greater difference between the regions translated into a wider range and better contrast. Then the goal is to take the well-contrasted information from both source images, V and N, to complement each other. The main idea of the proposed algorithm is to use the information from both images to obtain a fused image with the best characteristics of the input images.

(a) (b)



(c)

Figure 4.2: (a) Visible image, (b) NIR image,(c) their corresponding luminance distributions (l_V, l_N), red squares show the possible areas where complementary information can be found.

4.2 Fusion by weighted luminance

In the first approach of the fusion method, it is assumed that the information to be fused by finding differences between the luminances of V images and N. So direct subtraction is performed to find these differences. Then, the fusion criterion is based on image fusion technique weighted pixel averaging by the weight assignation of luminances. The overview of the proposed method is presented in Figure 4.3 and is establish at the Algorithm 1.

First, transform the V image into a grayscale (Gv) (Algorithm 1 line 1) for ranging the values of V image to correspond to the range of the N image. Then, to find the differences

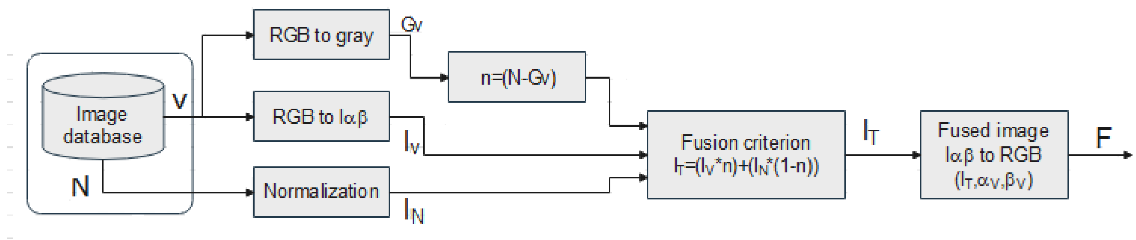


Figure 4.3: Overview of the fusion by weighted luminance.

Algorithm 1 V-NIR fusion method based on weighted luminance

Input: N: Near-infrared image,V:visible image
Result: F: Fused image
1 $Gv \leftarrow$ RGB to gray (V)
2 $n \leftarrow (N - Gv)$
3 $[l_V, \alpha_V, \beta_V] \leftarrow$ RGB to $l\alpha\beta(V)$
4 $l_N \leftarrow$ Normalize N
5 $l_T \leftarrow (l_V * n) + (l_N * (1 - n))$
6 $F \leftarrow l\alpha\beta$ to RGB(l_T, α_V, β_V)

between these two images, a subtraction is performed. The results of this subtraction will be found in the matrix n; where n contains the normalized differences between the sources images, ranging from -1 to 1. The normalization is by dividing the value of n by the maximum value of bits used to define each pixel of the images, which in this particular case is 255. There are two possible extremes cases where the range comes from:

- Case 1: When $N = 255$ and $Gv = 0$; therefore $n = (255 - 0)/255 = 1$.
- Case 2: When $N = 0$ and $Gv = 255$ in this case result $n = (0 - 255)/255 = -1$.

Figure 4.4 shows an example of Gv, N and n where the differences between the two images can be observed. The more considerable difference when N value is higher than Gv, the value of n tends to 1, which is reflected in the image as a lighter pixel. Conversely, when the value of Gv is greater than N, the value of n tends to -1, so it is displayed darker. This differences map is use later to assign weights to the luminances in the fusion rule.

Next transform the color image V to $l\alpha\beta$ (Algorithm 1 line 3) to obtain the luminance values also l_N which is the normalized N image (Algorithm 1 line 4). The normalization is carried out

(a)

(b)

(c)

Figure 4.4: (a) Gray image from visible Gv, (b) NIR image N and (c) differences value n matrix.

by substituting the value of N in place of the matrix of R, G, B in the transformation Eqs. (A.21-A.23) so that it is in the same range of values as l_V. The idea of the fusion rule starts from the general method of assigning weights, where it is assumed that both images contribute the same. However, the proposed method starts from the assumption that the input images can contribute in different proportions. And this ratio can be controlled with the difference map n that has just been calculated. In Eq. 4.1 the proportions in which l_v and l_N will contribute are established. This rule is used to calculate the new luminance value (l_T)(Algorithm 1 line 5). Where the operator $A * B$ is the element-by-element product of A and B.

$$l_T = (n * l_V) + ((1 - n) * l_N) \tag{4.1}$$

In the fusion rule, there may be three possible cases:

- Case 1: When $N = Gv$ in this case $n = N - Gv = 0$; therefore $l_T = (l_V * 0) + l_N(1 - 0) =$

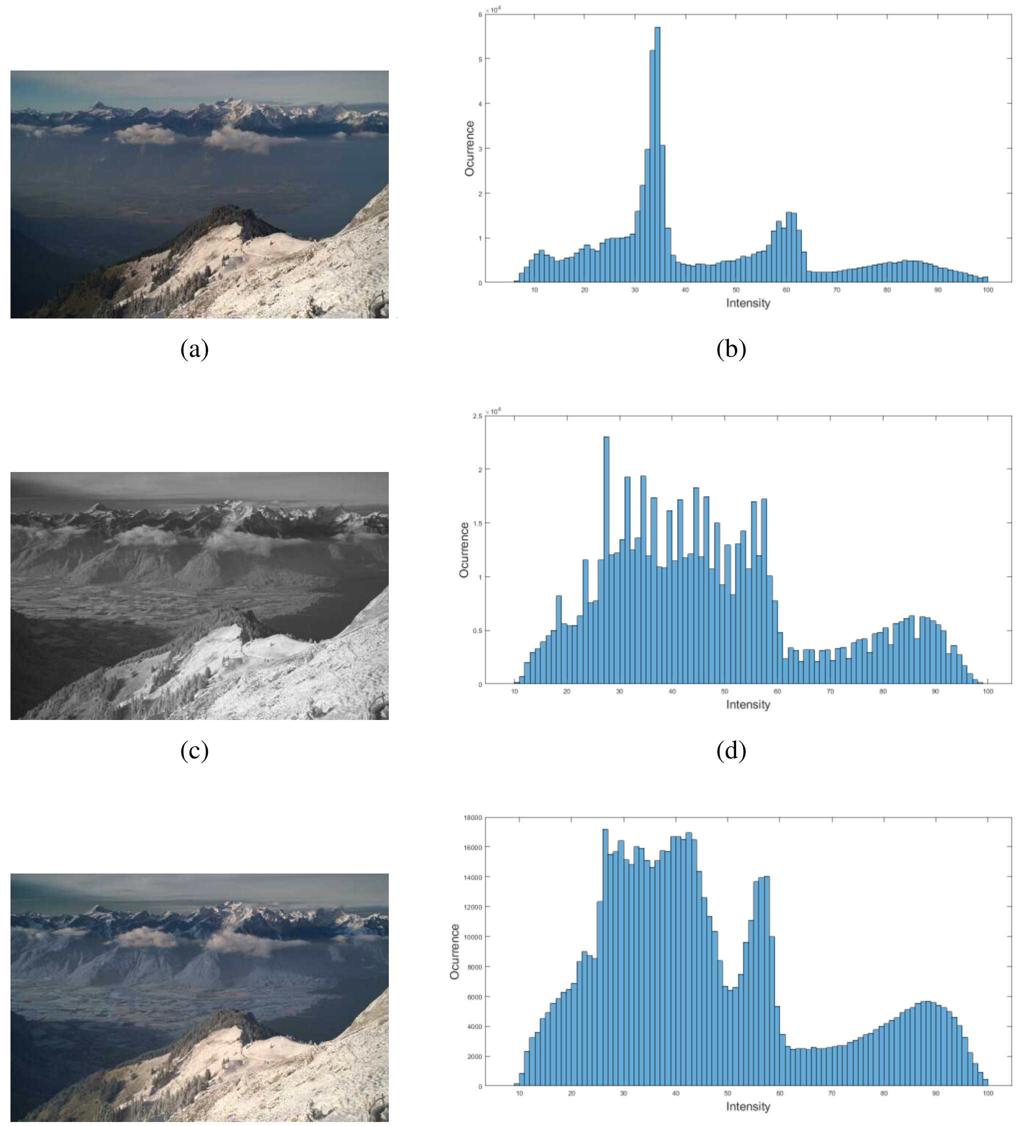
l_N the value of the luminance of the NIR (l_N) is taken. In this case it does not matter which one contributes more or less.

- Case 2: When $N > Gv$ in this case $n = N - Gv = +n$ the result of n will have a positive sign, therefore $l_T = nl_V + l_N(1 - n) = n(l_V - l_N) + l_N$.

- Case 3: When $Gv > N$ in this case $n = Gv - N = -n$ the result of n will have a negative sign therefore $l_T = -nl_V + l_N(1 + n) = n(l_N - l_V) + l_N$.

Finally, the fused image (F) is obtained by using the channels l_T, α_v and β_v for returning back to the RGB color space (Algorithm 1 line 6). As can be seen in Figure 4.5 the fused image, information that could not be seen in the visible can be recovered, especially in the region covered with haze. And as expected, the histogram of the fused image (Figure 4.5 (f)) shows how the information is complemented.

The fusion carried out in this section was based on the assignment of weights of the two luminances through a difference map, from now on is mentioned as [Herrera-Arellano et al., 2019]. Although it was possible to fuse information, it is still not possible to improve the contrast comparing to the original, and in some cases, the images are desaturated. So it was decided to explore morphological transforms to select information.



(a) (b)

(c) (d)

(e) (f)

Figure 4.5: (a) Original V, (b) luminance histogram l_V, (c) NIR image N, (d) luminance histogram l_N, (e) fused image (F) using the proposed method and (e) luminance histogram l_T.

4.3 Fusion by top-hat and bottom-hat

In this fusion method, the selection of information is carried out using the top-hat, and bottom-hat transforms. Assuming that using the top-hat transformation is possible to extract the information from the light regions and the bottom-hat, the information of dark regions is selected.

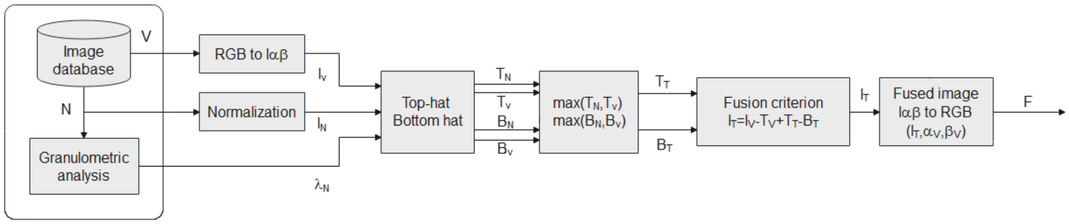


Figure 4.6: Framework of the proposed V-NIR fused method.

The fusion method (Figure 4.6), is performed between the luminance component of V and N input images. First, it is analyzed to determine the optimal λ scale for the top-hat transformation through a granulometric process for the data set. This ensures that both transformations, T_V and T_N, have the same target structure size. Then the information to fuse is selected by top-hat (T_T) and bottom-hat (B_T). Next, a fusion criterion is applied by considering the luminance l_V and the higher improvement reached by top-hat (T_T) and bottom-hat (B_T). Finally, l_V is replaced by the new fused luminance l_T in the inverse transform to return F to the fused color image F.

Algorithm 2 V-NIR fusion method based on top-hat and bottom-hat

Input: N: Near-infrared image,V:visible image,$\lambda = 9$
Result: F: Fused image
$[l_V, \alpha_V, \beta_V] \leftarrow$ RGB to $l\alpha\beta(V)$
$[l_N] \leftarrow$ Normalize N
$T_N \leftarrow Tw_\lambda\,(l_N)$
$T_V \leftarrow Tw_\lambda\,(l_V)$
$B_N \leftarrow Tb_\lambda\,(l_N)$
$B_V \leftarrow Tb_\lambda\,(l_V)$
$T_T \leftarrow max(T_N, T_V)$
$B_T \leftarrow max(B_N, B_V)$
$l_T \leftarrow l_V - T_V + T_T - B_T$
$F \leftarrow l\alpha\beta$ to RGB(l_T, α_V, β_V)

4.3.1 Information selection

As part of selecting the information is the use of the top-hat transform in which the setting of structural element to use is important in this type of transformation. Thus it is essential to know which structure sizes (λ) are relevant to use. The selection of the structuring element is carried out using the entire database, and once the structuring element is selected, this variable is set for the subsequent fusion process. The idea is to find a representative size of the entire database that can be left fixed to simplify the complexity of the processing. However, it is also possible to select the element for each of the images.

Structural element selection.

In Figure 4.7, it can see how the result of the fusion is affected by the structuring element depending on the size and shapes that it can present. Therefore, a granulometric analysis was carried out to establish the size. Then define the shape to be used.

Granulometric analysis to determined structural element size.

The NIR image contains well-contrasted regions that complement the visible information. Thus, the NIR image is processed, through a granulometric analysis, to identify the regions with a higher contribution. To determine the most suitable size for Tw_{B_λ}, a granulometric analysis was performed by using (Eq. 2.17), the parameter λ controls the size of the operation $\gamma_{B\lambda}$ applied to the image. The range of values $[1, 50]$ with an increment $\Delta = 1$ for λ was used. Figure 4.8 illustrates an example of the pattern spectrum obtained through the granulometric analysis. The higher PS values in the plot are associated with λ sizes from 1 to 9 that cause a significant change in the image, either because they match with large structures or there is a higher occurrence of λ size structures. Small λ values are common in textured images because of the high variation of gray levels. Since the dark regions are smaller than the light regions, the λ value for the bottom-hat is fixed at 3. For this work, a λ value of 9 was used since it showed a reliable performance during testing. The analysis performed for the used database indicates that 133 of the 305 tested images get the same λ value due to their similar textural characteristics. The others tried images that have closer values to λ, but they are not the majority by themselves. A fixed scale could be relevant when seeking to reduce the processing, such as in real-time applications. However, it is important to highlight that our methodology can be easily adapted by analyzing individual images as input instead of the complete database. This characteristic allows a wide range of applications, for example, when working with images that have different characteristics of each other.

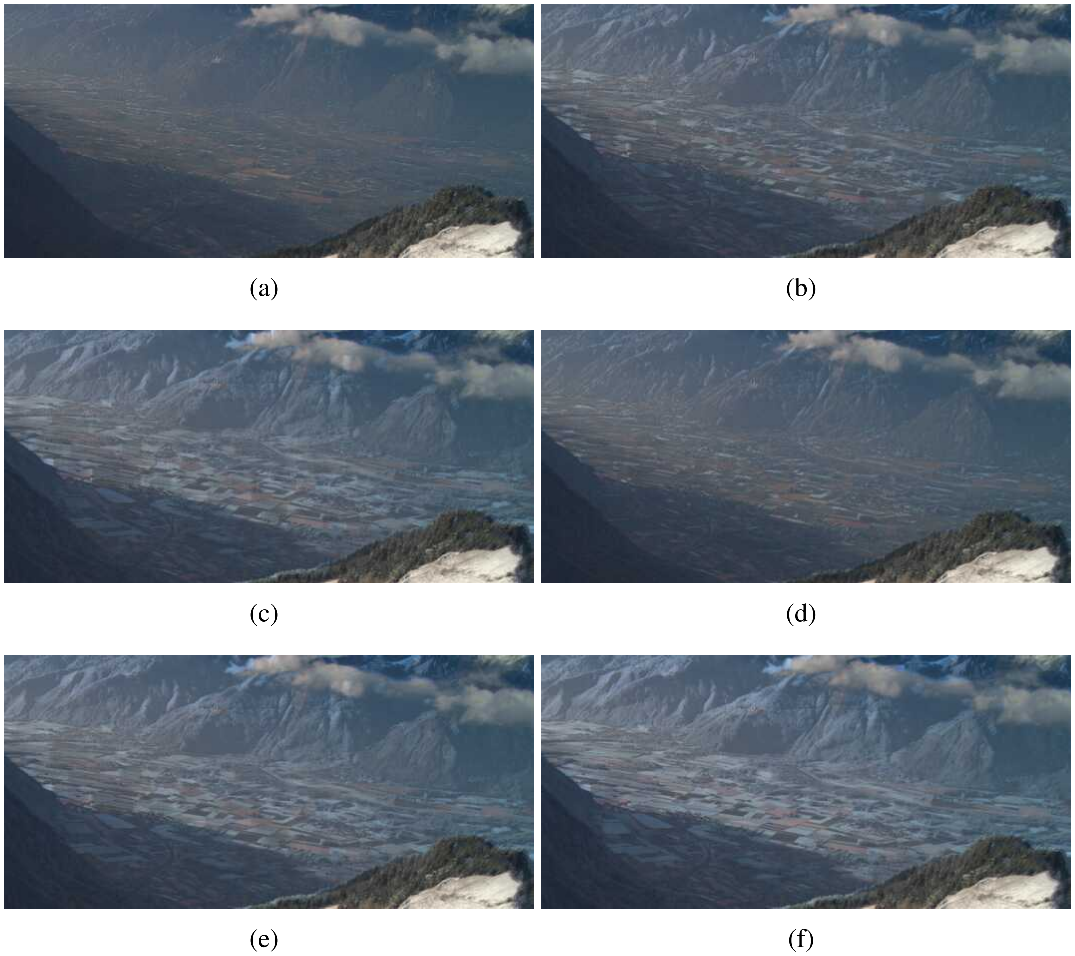


Figure 4.7: Difference among structural element shapes and sizes. (a) Square shape $\lambda = 3$, (b) Square shape $\lambda = 9$, (c) Square shape $\lambda = 15$, (d) Disk shape with $\lambda = 3$, (e) Disk shape with $\lambda = 9$, (f) Disk shape $\lambda = 15$.

Top-hat and bottom-hat.

Once the structural element is chosen, the bottom-hat and top-hat transform are applied to extract the information to be fused. First the N values are normalized, as l_N, to compare them with the luminance values (algorithm 2, lines 1-2). The normalization is carried out by substituting the value of N in place of the matrix of R, G, B in the transformation Eqs. (A.21-A.23) so that it is in the same range of values as l_V. Then the top-hat is applied to l_V and l_N as $T_V = Tw_{B\lambda}(l_V)$ and $T_N = Tw_{B\lambda}(l_N)$, respectively (algorithm 2, lines 3-4). The top-hat transform extracts image regions of size B_λ. Also, the bottom-hat is applied to l_V and l_N as $B_V = Tb_{B_3}(l_V)$ and

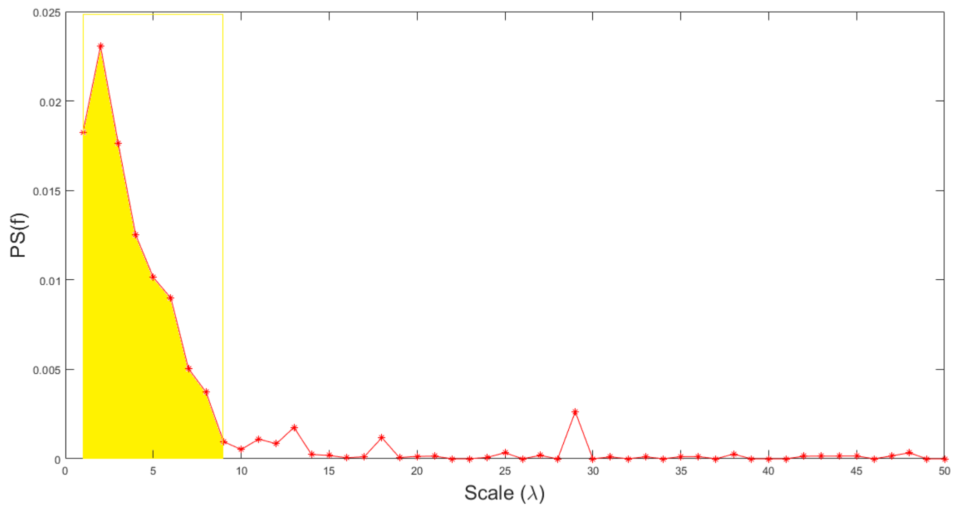


Figure 4.8: Pattern spectrum obtained from the granulometric analysis for the λ range [1, 50] with $\Delta = 1$ to select the λ value to use in the top-hat.

$B_N = Tb_{B_3}(l_N)$, respectively (algorithm 2, lines 5-6). Figure 4.9 shows the information obtained after applying the top-hat and bottom-hat transforms where the four images show that they contain information in common, which is the information contained in V and N but also have different information, which is the complementary information. So in the following step consist of recovering this information to keep the similar information and add the different one. The selection of relevant structures for the fusion imaging process is performed though the maximum value T_T and B_T (Algorithm 2, lines 7-8). This is the selective part of fusing the information by taking the maximum values of both transformations; similar information is kept while the different information is added (Figure 4.10).

4.3.2 Fusion strategy

The next step is the selection of the final luminance (l_T) (Algorithm 2, line 9). The fusion criterion is based on the following reasoning. First, the luminance from the visible l_V must be increased to a particular value to reach a contrast improvement. Second, the difference obtained through top-hat transformation allows the structures reached by the λ scale and from bottom-hat to outstand from the background. Then, the most outstanding value is passed to the fusion criterion as the

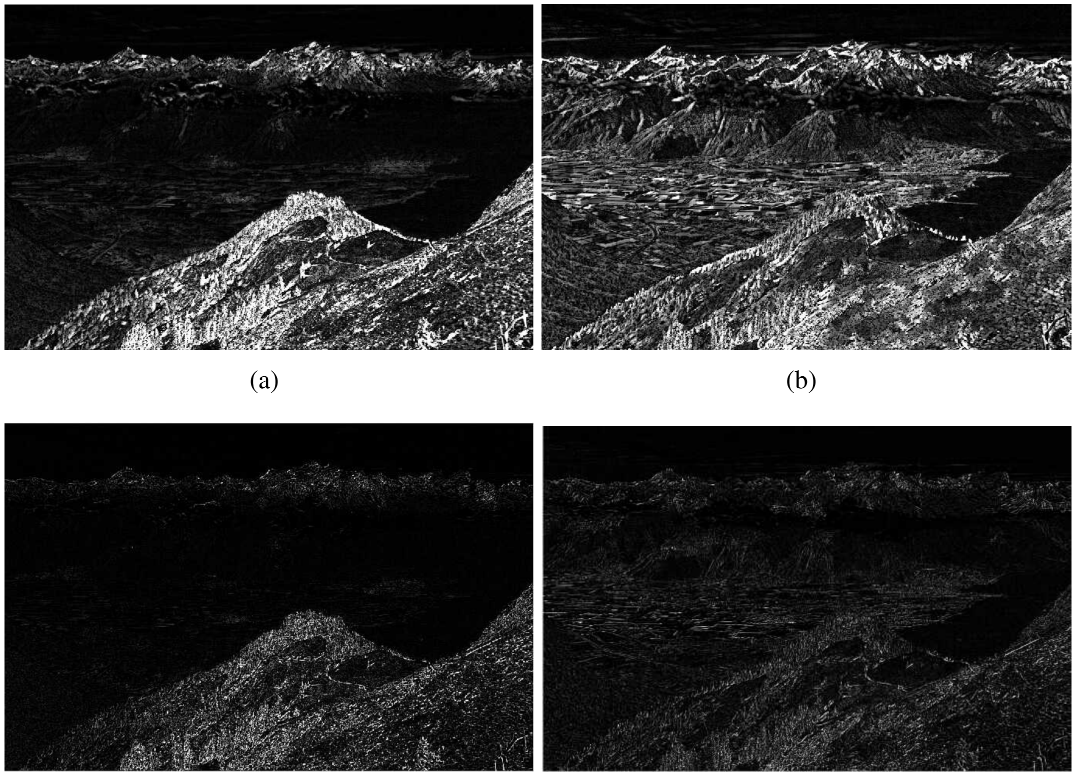
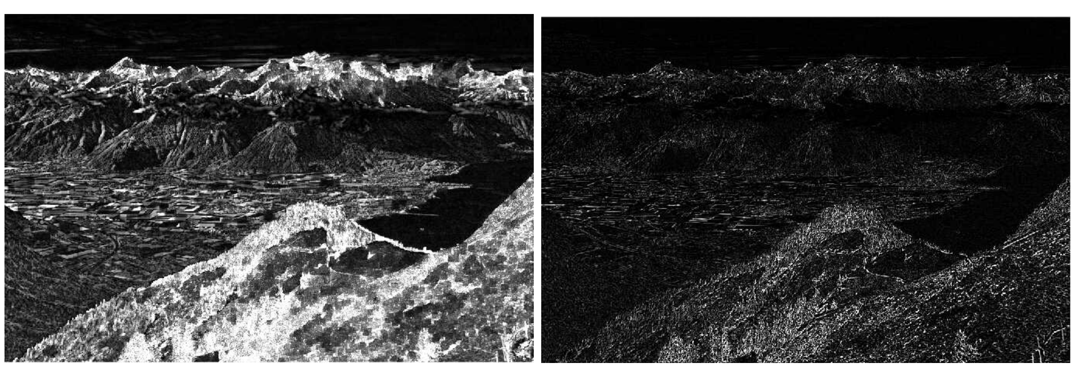


Figure 4.9 panels: (a), (b), (c), (d)

Figure 4.9: Results of: (a) top-hat applied to l_V, (b) top-hat applied to l_N, (c) bottom-hat applied to l_V, (d) bottom-hat applied to l_N



(a) (b)

Figure 4.10: The maximum value of: (a) top-hat (T_T) and (b) bottom-hat (B_T).

incremental value T_T by taking the maximum value between between T_V and T_N. The fusion criterion must consider that the analyzed images contain a well-contrasted region to keep and a low-contrast region that requires improvement. Thus, if the maximum value comes from T_V, the original luminance is preserved, but if it comes from T_N, the original luminance is increased to reach contrast improvement, as stated in (Eq. 4.1). However, a direct increment $l_V + T_T$ may cause a value overflow; to avoid it, the value of T_V is used as a limiter. By removing B_T, the dark areas are highlighted, specially non-hazy regions. In this way, fused luminance l_T and the channels α_V and β_V are used to return to the RGB color space, obtaining the final fused image F (Figure **??**).

In this section, two methods were proposed; the first one is based on assigning weights through a difference map whose part of fusing the different information. However, despite having more information than the original image, it has desaturated colors. So it was decided to find another way to extract the information. In the second method, the information is extracted using morphological transformations. The top-hat is used to extract information from light regions, while the bottom-hat is used to extract information from dark regions. The main contribution of the fusion by morphological transformations is the fusion of edges. With this method, information is fused selectively, avoiding the appearance of artifacts and color over-saturation.

Chapter 5

Experiments and results

This chapter shows the experimental results of the visible and NIR image fusion process using the methods described in methodology (Section 4). The experiments were done on the method based on mathematical morphology that, from here on, is called the proposed method.

First, it describes the characteristics of the database used. The second part shows the experiments carried out to set the parameters of the proposed method. Then the parameters of the methods with which the comparison was made. Next, the qualitative results and quantitative evaluations using the evaluation functions mentioned in the background (Section 2.4.2), following luminance analysis. Finally, the discussion is made, including over-saturation, blurred edges, and a possible application.

5.1 Experiments

The experiments run using Matlab 2020b. The images with which the experiments were carried out belong to the database V-NIR EPFL (École Polytechnique Fédérale de Lausanne). Images were captured with the Nikon D90 and a Canon T1i cameras, using the ”B+W 486” filter for visible and the ”093” filter for NIR images. The cutoff between the two filters is approximately 750nm.

Visible and NIR images are 1024×680 pixels [Brown and Süsstrunk, 2011]. This dataset consists of 477 pairs of images divided into nine categories: country, field, forest, indoor, mountain, old building, street, urban, water. But for this work, the indoor and old building images were discarded the same way the images that are not aligned. Therefore the total images used in the experiments were 305 pairs.

ANOVA tests were used to evaluate the results of the experiments where the null hypothesis (H0) proposes no difference between the means of evaluation functions values. In contrast, the alternative hypothesis (Ha) states that at least one mean is different. The H0 is rejected if the $P \leq \alpha$ value equality of means and concludes that at least one population means different from the others. The ANOVA table shows the between-groups variation and error. Where **SS** is the sum of squares, and **df** is the degrees of freedom. While **MS** is the mean squared error. Finally the **F** statistics is the ratio of the mean squared errors [Triola, 2018].

5.1.1 Shape of the structural element.

Even though it is not common to find a specific morphology in outdoor images, this experiment is carried on to know if there is a morphology that could be better adapted and improve the subsequent results. This experiment is done to know if the shape of the structural element affects the contrast results. The size that was already selected in the previous step is left fixed. The analyzed shapes are diamond, square, disk, cube, and octagon. In this experiment, the p-value ($Prob > F$) is 1. This p-value can be taken as evidence that all means were equal when the image dataset was processed with different shapes of structural elements. Usually, a square structuring element is used unless objects with specific morphologies are processed, which in the case of natural outdoor images does not occur because of the diversity of the regions. Moreover, a square shape may also facilitate further replicating and comparing the work since most software tools include it. Therefore, the usual square shape was used in the following processes. The ANOVA comparison is summarized in Table 5.1. The quality measure of contrast was taken as reference to perform the test, showing that there was not a significant difference in the results when the image dataset was processed by the common square shape or by some other shapes (Figure 5.1). Commonly, the difference generated by the shape of the structuring element is reflected in the edge of the objects

Table 5.1: ANOVA Table for different structural element shape

Source	**SS**	**df**	MS	F	P>F
Groups	0.03	4	0.00655	0	1
Error	5351.12	870	6.15072	-	-
Total	5351.15	874	-	-	-

when they contrast from their surroundings.

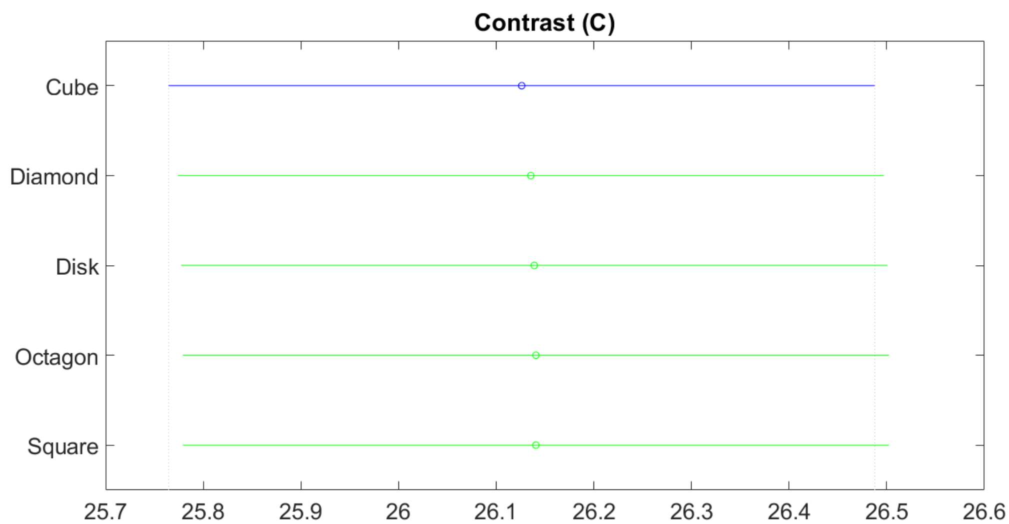


Figure 5.1: Comparison of the mean and standard deviation estimations of the contrast evaluation function by using different structural element shapes.

5.1.2 Color spaces

Different color spaces are used in the literature for V-NIR fusion and have not yet defined an optimal one for these applications. Moreover, each color space has characteristics that can favor or harm the fusion results. Therefore, the following experiment seeks to know how the results are affected by using the chosen color space $l\alpha\beta$ or a different one. So the next test is performed using some of the state-of-the-art color spaces that separate the luminance information from the chroma, such as YCbCr, HSV, CIE $L^*a^*b^*$, and $l\alpha\beta$. The ANOVA test of the C value for the color spaces (Table 5.2) showed that the difference was not significant ; therefore, similar results by the proposed method regardless of the color space were observed. Consequently, it can be inferred that the proposed method results are independent of the color space used.

Table 5.2: ANOVA table for color spaces

Source	SS	df	MS	F	Prob>F
Groups	17.06	3	5.686	0.97	0.4073
Error	39994.79	680	5.874	-	-
Total	4011.85	683	-	-	-

The proposed method results were compared with some V-NIR fusion state-of-the-state methods: [Vanmali et al., 2015], [Vanmali and Gadre, 2017], [Sharma et al., 2017], [Elliethy and Aly, 2017] and with a previous proposal of our authorship based on weighted luminance [Herrera-Arellano et al., 2019]. The codes of the fusion methods are all publicly available, and their parameters are all set in accordance with those in the original studies, Table 5.3 summarizes the parameters values. In [Vanmali et al., 2015] uses a constant λ_1 to calculated depth map values, while [Vanmali and Gadre, 2017], adds an β_1 value to control the color saturation of the fused image. Secondly [Sharma et al., 2017] used a n to defined the levels of decomposition of a WLS filter. The proposed top-hat and bottom-hat method uses a $\lambda_{top-hat}$ and $\lambda_{bottom-hat}$ associated with the size of the structural element of the transformations. On the other hand [Herrera-Arellano et al., 2019] does not need parameter settings.

Table 5.3: Parameters settings of V-NIR fusion methods.

Method	**Parameters**	**Color space**
[Vanmali et al., 2015]	$\lambda_1 = 1$	HSV
[Vanmali and Gadre, 2017]	$\beta_1 = 1.5$	HSV
[Sharma et al., 2017]	$\Lambda = 0.125$, $c = 1.2$, $n = 1$	YCbCr
[Elliethy and Aly, 2017]	$S = 5$, ω_{cut}, $k = 19$	YCbCr
[Herrera-Arellano et al., 2019]	-	$l\alpha\beta$
Proposed method	$\lambda_{top-hat} = 9$, $\lambda_{bottom-hat} = 3$, square	$l\alpha\beta$

5.2 Results

The first section of the results shows some of the images obtained from the proposed method against the other methods. Subsequently, the average values of the contrast, entropy, saturation, and color similarity evaluation functions are obtained. An ANOVA analysis was then carried out to determine if there are significant differences between the resulting images and the original image with respect to these same evaluation functions values. Then, the discussion is addressed over-saturation also, an analysis of luminance and edges after image fusion is needed. Finally, it is examined how the proposed image fusion results could improve a post-processing task such as segmentation.

With the proposed fusion method, the top-hat and bottom-hat transformations extract the information from the visible and the NIR. At the same time, the $l\alpha\beta$ color space is used to process only the luminance of the V images. Under this approach, it is possible to fuse the visible and NIR information to improve the visualization of low contrast structures, hidden by haze, resulting in a higher quality image. While using a selective approach for fusion (top-hat and bottom-hat), it is possible to avoid over-saturation. In the Figure 5.2 is shown a compendium of results of the comparison fusion methods against the proposed method. The first two rows show the input images V and N, while following rows show the results of the methods. In general, the compared methods manage to recover information that is behind the haze. But in the particular case of [Vanmali et al., 2015] and [Vanmali and Gadre, 2017] the results show that over-saturation is generated. Also, the amount of detail revealed is visually less noticeable compared to other methods. On the other hand, in the results of [Herrera-Arellano et al., 2019] the visibility of the structures behind the haze is improved compared to the methods [Vanmali et al., 2015], [Vanmali and Gadre, 2017] and [Sharma et al., 2017]. However, there is color desaturation. Finally, the proposed method highlighted and improved the structures providing a more detailed visual description than the other methods. This increase in detail is due to the contribution of NIR images. Furthermore, the proposed method provides additional and detailed structures and preserves those of the haze-free regions. A notable advantage is that there is no over-saturation or desaturation, which is reflected in the V image colors preserved after V-NIR image fusion.

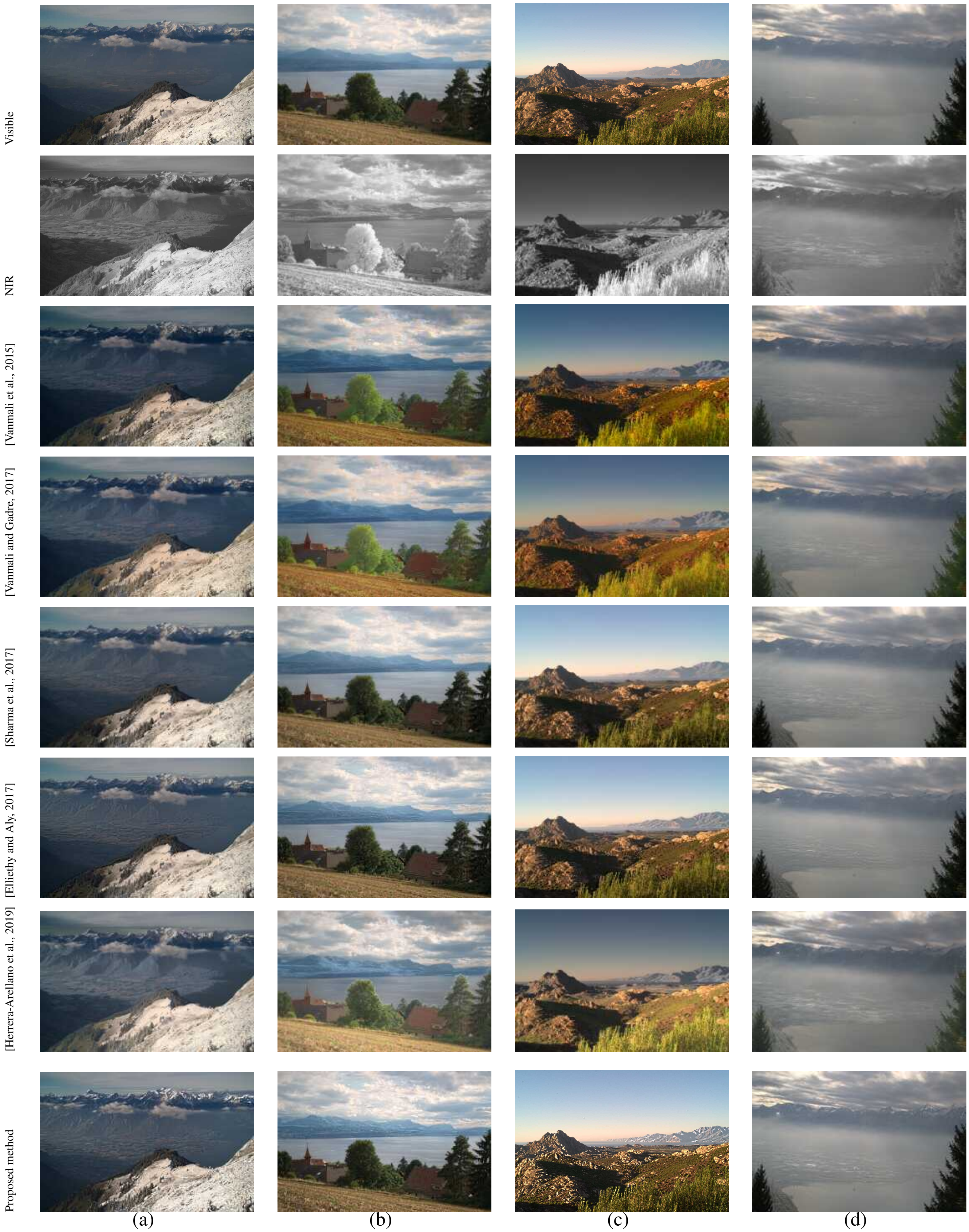


Figure 5.2: Comparison between images fused with reference V-NIR methods on four NIR and the proposed method. Quantitative evaluations associated with the results is provided in Table 5.4.

5.2.1 Quantitative evaluation.

All outdoor images included in the dataset were evaluated with contrast, entropy, saturation, and color similarity evaluation functions. As indicated above, the objective of the method is to recover as much information from the NIR to merge it with V (which is measured by contrast and entropy) while avoiding over-saturation (measured by saturation and color similarity). It is expected to have higher values of entropy and contrast to consider an improvement. Also, saturation values stay the closest to the input image V while the color similarity values are no greater than 5; this indicates that color is similar. Table 5.4 shows the results obtained in the evaluation functions. In addition, an ANOVA test was performed to know if the differences in the results were statistically significant comparing to original. It is complemented with multiple comparisons of the means using Tukey-kramer, which allows examining which means are different and estimating the degree of difference.

Table 5.4: Average of 175 images of fusion results based on the selected evaluation functions

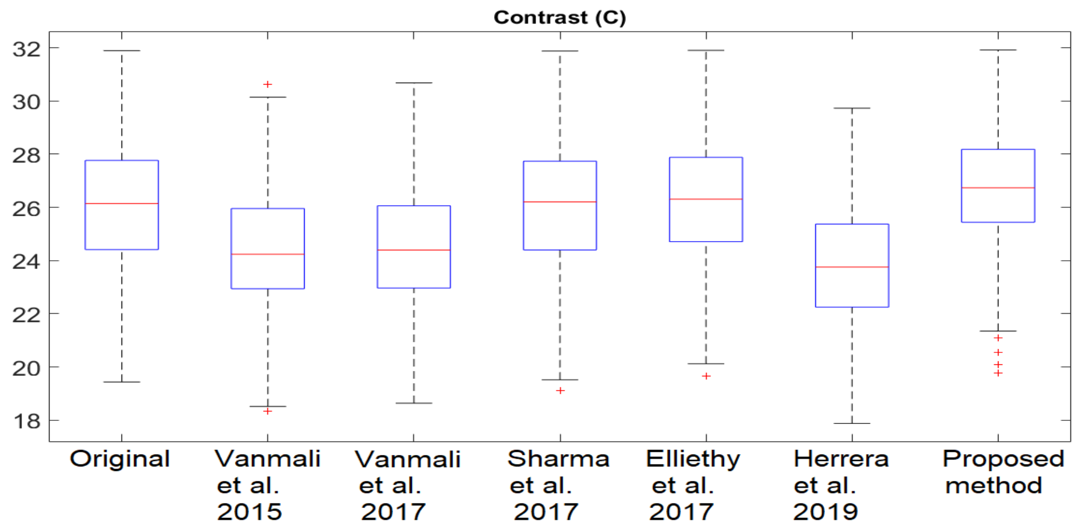
Method	**evaluation function**			
	$\Delta E_{00} \leq 5$	$\hat{S} =$	EN $\uparrow$	C $\uparrow$
Original	-	12.63 ± 5.62	13.98 ± 0.41	26.06 ± 2.45
[Vanmali et al., 2015]	10.47 ± 3.80	22.97 ± 13.67	13.82 ± 0.45	24.36 ± 2.10
[Vanmali and Gadre, 2017]	8.83 ± 3.37	18.18 ± 7.72	13.94 ± 0.40	24.42 ± 2.17
[Sharma et al., 2017]	1.75 ± 0.53	12.86 ± 5.77	13.99 ± 0.39	26.04 ± 2.45
[Elliethy and Aly, 2017]	0.87 ± 0.54	13.06 ± 5.83	14.02 ± 0.39	26.23 ± 2.35
[Herrera-Arellano et al., 2019]	11.19 ± 3.39	11.54 ± 4.96	13.91 ± 0.41	23.79 ± 2.10
Proposed	2.15 ± 0.89	12.54 ± 5.61	14.09 ± 0.39	26.70 ± 2.36

Added information

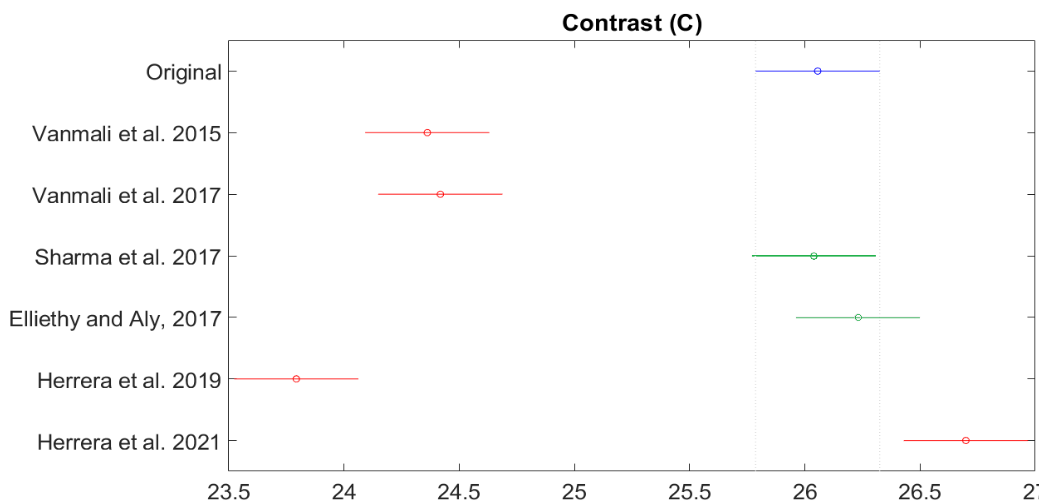
When a contrast method or entropy values are lower than the original image, it indicates a loss of information due to, for instance, a blurring in the edges of the original image. These are the cases of [Vanmali et al., 2015], [Vanmali and Gadre, 2017] and [Herrera-Arellano et al., 2019] where the contrast quality evaluation function are lower respect to the original.The proposed method has higher contrast and entropy values than the original image. Also they are significantly different as shown in Figures 5.3 (b) and 5.5 (b). In Figures 5.3 (a) and 5.5 (a) the differences also can be seen by comparing the median line of the box plot associated with the value of the tested set. If the median line of a box plot lies outside of a comparison box plot, then there is a difference between the two groups; otherwise, the highest values of contrast are obtained by [Sharma et al., 2017], [Elliethy and Aly, 2017] and the proposed method (Table 5.4). The slight difference may due to the contrast value evaluates the whole image, but regions, where information was added by fusion, cover a smaller area in most cases. However, another way to assess methods performance is through the outliers they generate (red crosses in Figures 5.3 and 5.5). An outlier is a datum point located outside the whiskers of the box plot. Two characteristics are essential for analysis, the box length and the number of outliers. On the one hand, the larger the box, the bigger the dispersion of the values. On the other hand, a higher amount of outliers represents those cases for which the method cannot achieve a result near to its mean improvement or indicates that there are conditions not considered by the method that affects its final result. A few outliers are associated with the original set, indicating that the image set has similar conditions, according to the obtained evaluation functions values. Therefore, if images in similar conditions are processed by the same method, it is expected that results behave similarly, e.g., a few outliers close to the whiskers of the box. Then, in fusion methods, it is important not only to reach a high contrast and entropy value but also a small number of outliers to ensure a reliable performance. Then the multiple comparison of group means plots (Figures 5.3 (b) and 5.5 (b)) illustrate whether the values are significantly different among methods. On the y-axis are the compared methods, while values in the x-axis are related to the mean values of each analyzed group. The group original mean and the comparison interval is in blue. If the comparison intervals for the other groups do not intersect with the intervals for the group original mean, they are highlighted in red. Otherwise, they are highlighted in green. The ANOVA table corresponding to these plots are shown in Table 5.5 and Table 5.6.

Table 5.5: ANOVA Contrast (C)

Source	**SS**	**df**	**MS**	**F**	**Prob>F**
Groups	1681.72	6	280.286	57.31	1.55843e-62
Error	5956.94	1218	4.891		
Total	7638.66	1224			

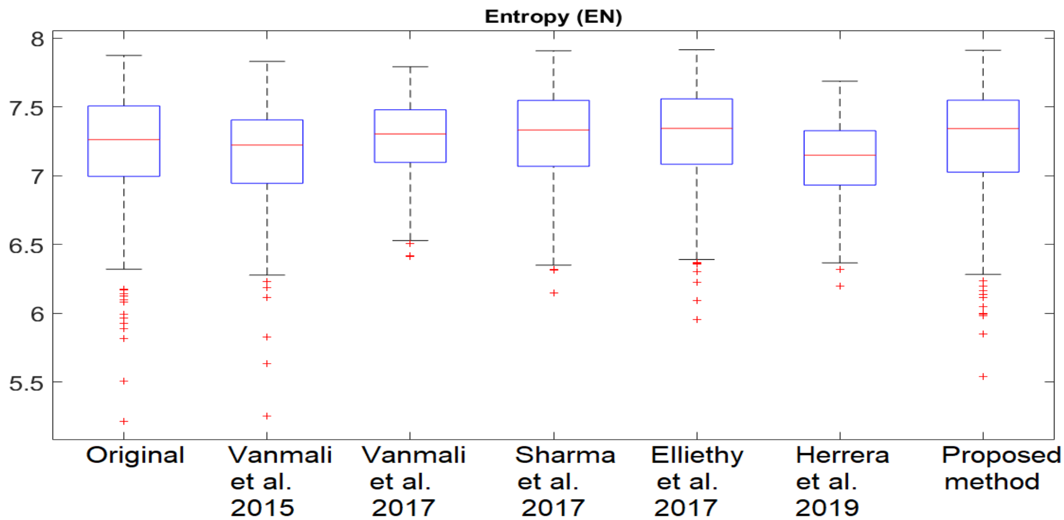


(a)

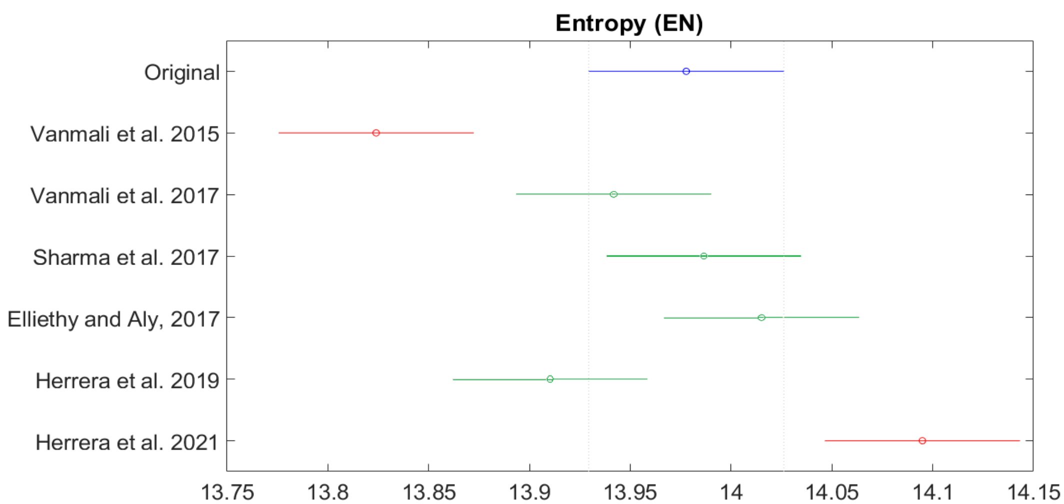


(b)

Figure 5.3: Contrast comparison of fusion methods to compare the new added information.



(a)



(b)

Figure 5.4: Entropy comparison of fusion methods to compare the new added information.

Table 5.6: ANOVA Entropy (EN)

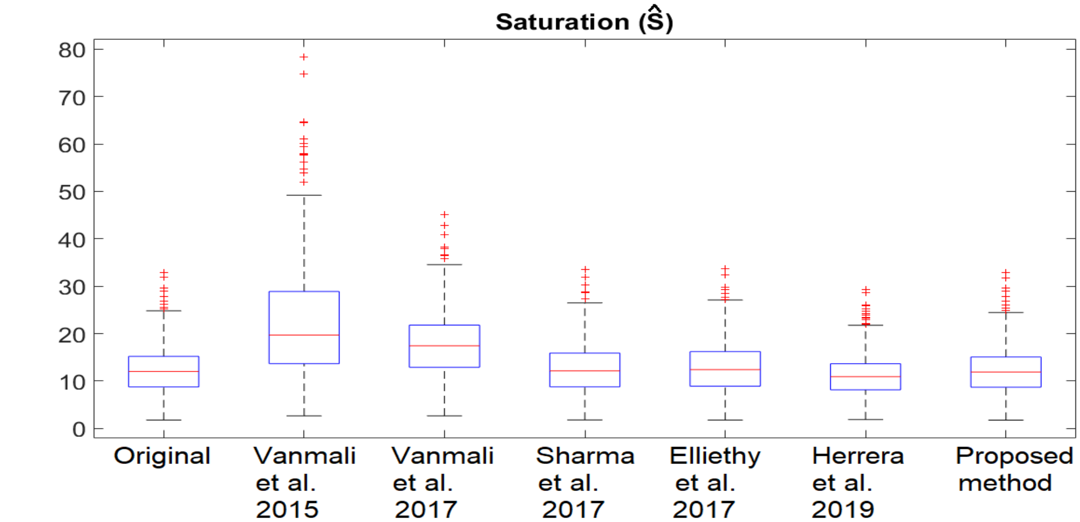
Source	SS	df	MS	F	Prob>F
Groups	4.317	6	0.71954	4.65	0.0001
Error	188.332	1218	0.15462		
Total	192.649	1224			

Color over-saturation

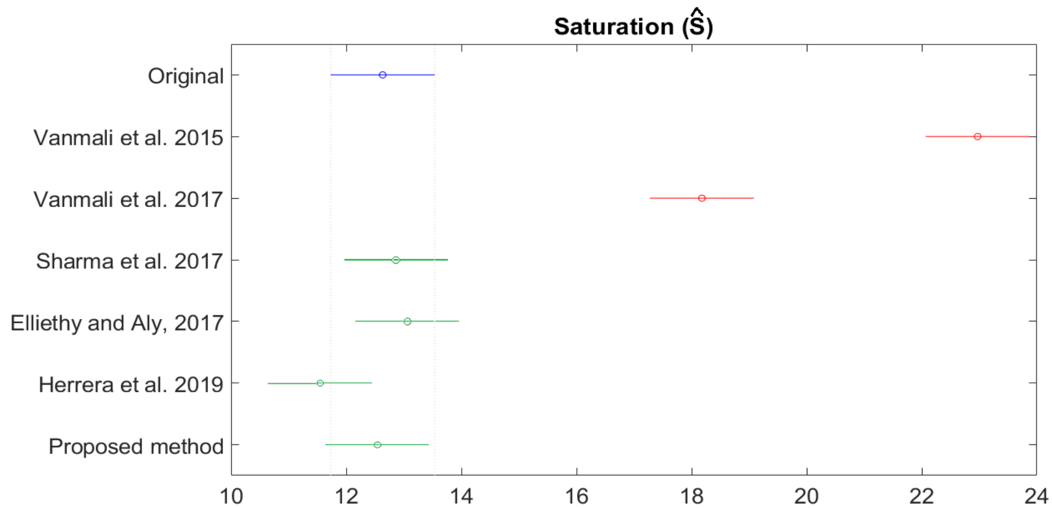
As it has been handled throughout the investigation, less significant changes are desirable regarding the color change. Thus is desirable the $\hat{S}$ values keep similar as possible from the original images and the color similarity value under five. The Figures 5.5 (a) and 5.6 show the results of evaluate the color changes; with respect to $\hat{S}$, [Vanmali et al., 2015] and [Vanmali and Gadre, 2017] methods obtained significant differences. While [Herrera-Arellano et al., 2019] method presents a slightly lower value than the original one (Figure 5.5 (b)). Finally, [Sharma et al., 2017], [Elliethy and Aly, 2017], and the proposed method are not different from the original. Concerning the color similarity Figure 5.6, the [Sharma et al., 2017], [Elliethy and Aly, 2017], and the proposed method are the only ones that keep above five. In addition, these methods have a lower dispersion of values Figure 5.5 (a), which means that the proposed image fusion can provide consistent results and is less affected by variations in the input images. Therefore, the results demonstrated that the proposed method could add new information while avoiding significant color changes. The ANOVA table corresponding to this plot are shown in Table 5.7.

Table 5.7: ANOVA Saturation ($\hat{S}$)

Source	SS	df	MS	F	Prob>F
Groups	26854.3	6	4475.71	73.63	1.88309e-78
Error	74040	1218	60.79		
Total	100894.2	1224			



(a)

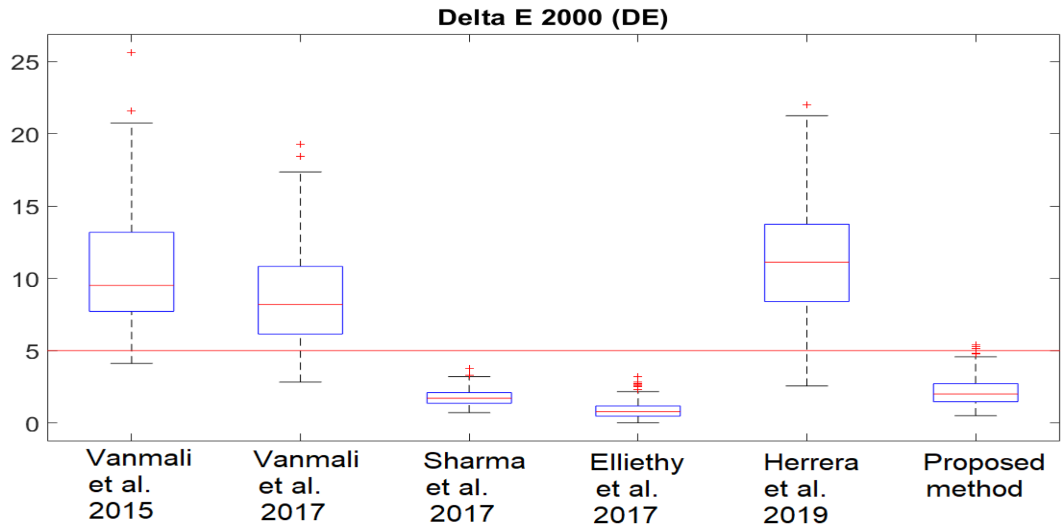


(b)

Figure 5.5: Evaluation functions comparison of fusion methods for assessing color over-saturation.

5.2.2 Luminance

This thesis mainly works with luminance, so we analyze how this process affects luminance (l_T). Considering a row segment of the images in Figure 5.7 is taken to analyze the fusion based on a top hat and a bottom hat. The Figure 5.8 (a) shows the original luminances l_V (solid black line) and l_N (solid red line) of the segment and its corresponding top hat T_V (dotted black line) and T_N

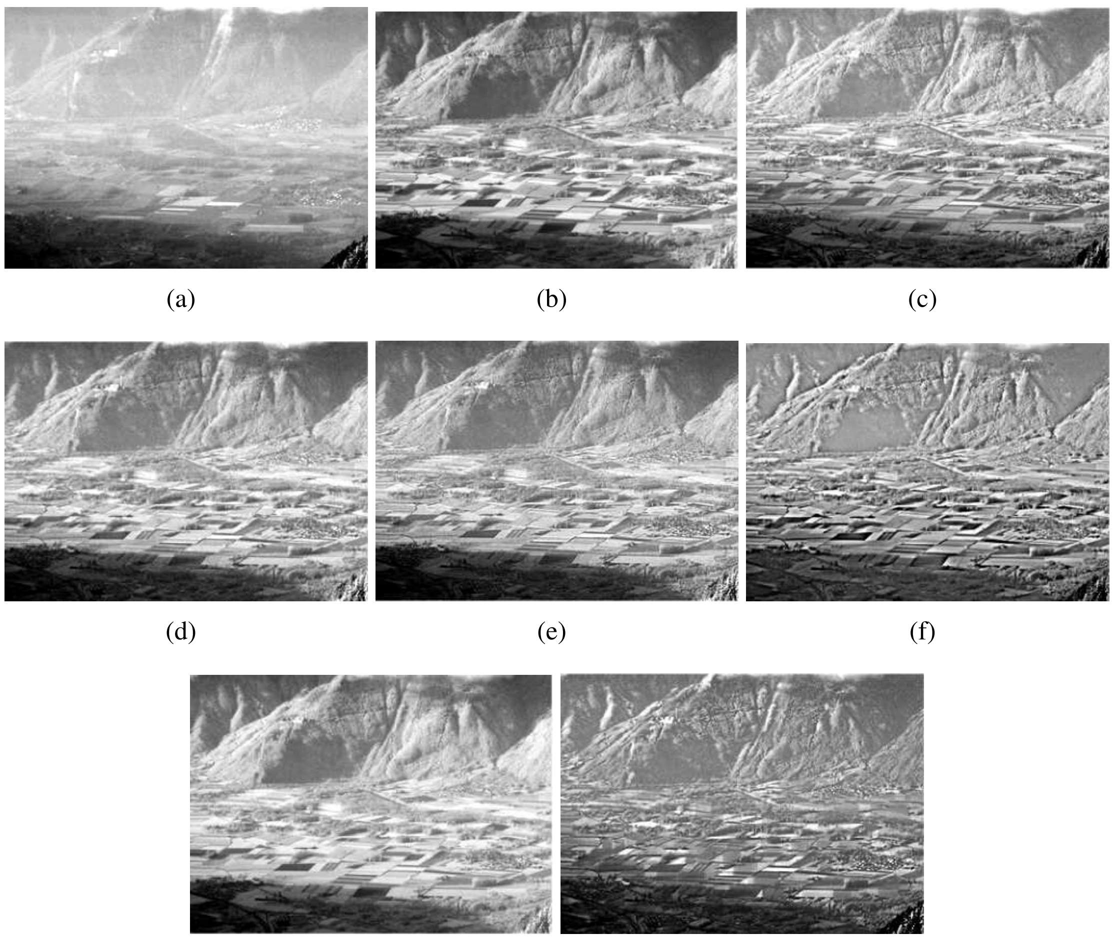


(a)

Figure 5.6: Comparison of fusion methods for assessing color difference.

(dotted red line). It can be seen that when the top hat T_V is greater than or equal to T_N (columns 5 to 15), the luminance result (l_T) remains the same as the luminance of the visible. On the other hand, for columns 40 to 45, T_N is greater than T_V so l_T is greater than l_v. This means that l_n provides more information on the fusion process according to Eq. 4.1. Figure 5.8 (b) shows a comparison between the fusion results obtained by the proposed method and the state-of-the-art methods. When we see the difference in the shapes of the lines (column 10-20), it can see that the luminance l_V appears to be constant, while in the fusion methods, it is varying either with higher or lower values. In this case, we can say that a structure or shape that was not in visible was fused. It cannot be said that if the value of l_T is greater than l_V it is better because that depends on the image. For example, we could find an image with fog whose value of l_V is high, and when merging, the result of l_T will be less than the original. What can also be said is that when performing a direct weighted combination of information from the visible and NIR bands, some regions are affected, especially when there are conditions where there is low contrast mixed with well-contrasted regions. As is the case with [Herrera-Arellano et al., 2019], higher values are always obtained in l_T than in l_V and the resulting values of the evaluation functions are not the best. This is because if both source images are combined directly, they could still cancel each other out. The proposed approach considers that image fusion can be improved if the information is first extracted from both sources, and only the most relevant information is part of the fusion (Eq.

4.1). As a result, contrast enhancement is ensured or maintained when it comes to well-contrasted regions.



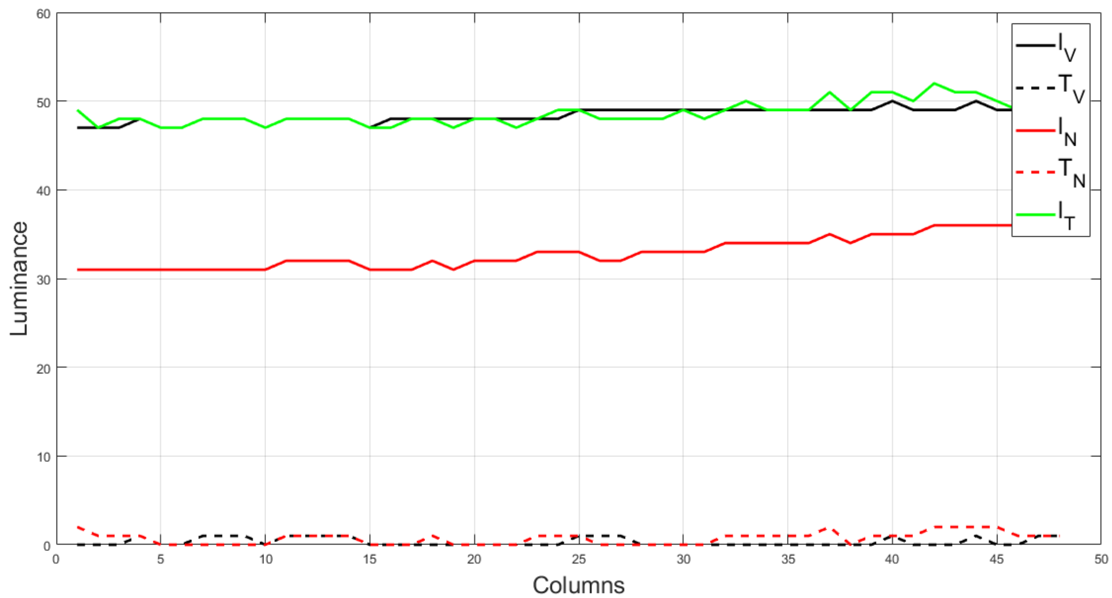
(a) (b) (c)

(d) (e) (f)

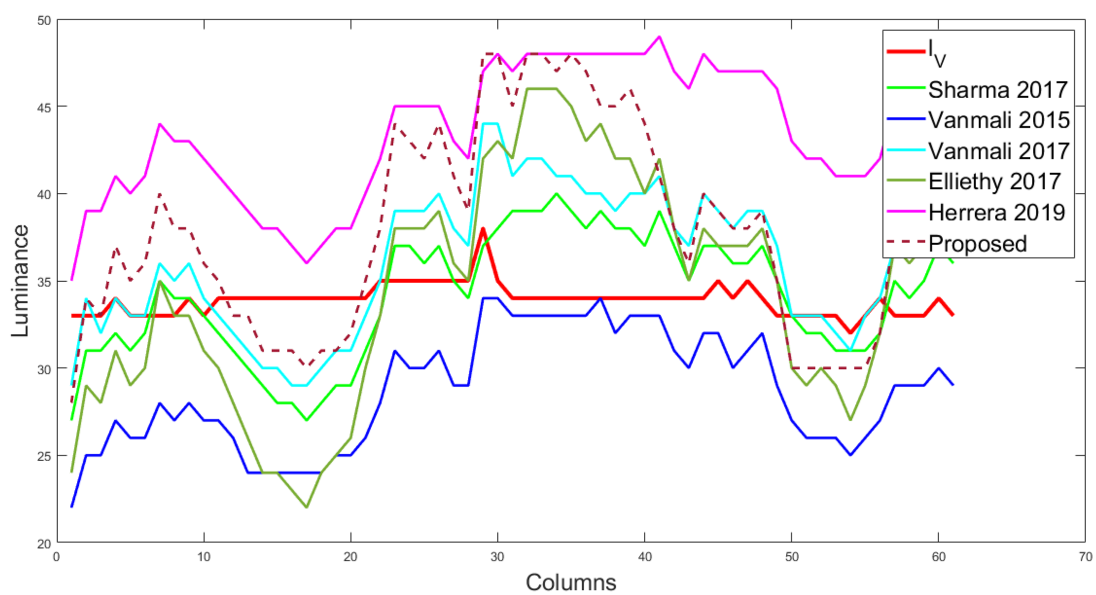
(g) (h)

Figure 5.7: Luminance comparison (a) Visiblel_V, (b) NIR l_N, and the comparison of the resulting fused luminance by (c) [Sharma et al., 2017], (d) [Vanmali et al., 2015], (e) [Vanmali and Gadre, 2017], (f) [Elliethy and Aly, 2017], (g) [Herrera-Arellano et al., 2019] and (h) proposed method.

In case of being required, the comparison with the results of the evaluation functions established in some articles the results when evaluating the state-of-the-art methods with the proposed method for the Mutual information, PSNR, RSMS, SSIM evaluation functions, are found in the Appendix C.



(a)



(b)

Figure 5.8: (a) Luminance enhancement through top-hat in l_V and l_N, and (b) the comparison with reference methods.

5.3 Discussion

5.3.1 Blurred edges

Once the images are fused, they are often used in further processing, for example, segmentation and corner detection. It is then that the quality of the image concerning the edges is important. For instance, if the borders between regions are well defined, the results of the segmentation process colud be more accurate. However, if the edges are poorly defined, the value between the regions and their edges is close. Thus two or more regions could be identified as those generating incorrect segmentation. Therefore, for example, although the entropy and contrast obtained by [Sharma et al., 2017] are generally higher than the originals, blurred edges persist in the results. Figure 5.9 shows an example of the differences in edges definition when the entropy and contrast are smaller than the original. In the case of the proposed method, this condition was found only one time in the 305 analyzed images, while for [Vanmali et al., 2015] 189 times, [Sharma et al., 2017] 36 times, and [Herrera-Arellano et al., 2019] 194 times.

5.3.2 Over-saturation

Although the proposed method does not generate over-saturation, it is necessary to consider what happens when it occurs. The over-saturation is generated in these methods mainly when the image does not contain a homogeneous haze distribution, mainly when it is only in a small region. For example, the original image in Figure 5.10 (a) has not area of haze. Therefore, the results of the fusion of the first image tend to be over-saturated by [Vanmali et al., 2015] method Figure (5.10 (b)). Considering that V-NIR fusion applications do not only focus on haze removal, it is important to consider the visible image saturation level. It is worth noting that when high over-saturation is generated, the resulting image would require color correction. Thus, an essential advantage of the proposed method is to avoid color over-saturation in the fusion results.

(a) $C = 20.3770$, $EN = 6.4281$

(b) $C = 19.6445$, $EN = 6.4035$

(c) $C = 20.4095$, $EN = 6.5185$

Figure 5.9: Difference in edges definition. A zoomed view of an example for (a) the original image, (b) [Sharma et al., 2017], and (c) the proposed method.

(a) $\hat{S} = 0.2207$

(b) $\hat{S} = 0.4420$

(c) $\hat{S} = 0.2203$

Figure 5.10: Saturation values for: (a) Original image and after image fusion through (b)[Vanmali et al., 2015], and (c) proposed method.

Chapter 6

Conclusions

This last chapter describes the conclusions of the research carried out and presented throughout this document. Likewise, this doctoral thesis offers different perspectives on future work in the coming years on the same subject under study. This research has addressed the fusion V-NIR images considering the contrast as a quality measure, including the color change problem. It was hypothesized that if relevant information is selected images, it could enhance low-contrast regions while avoiding oversaturation. The proposed method uses the morphological top-hat and bottom-hat transform to select the information, thus complementing the visible information with the NIR, increasing the contrast of the visible input image. Based on the comparison of the results obtained through the evaluation functions, the proposed method reaches a statistically significant difference concerning other methods. It was also tested to be robust regarding changes in the color space of images by providing consistent results. The main goal of this thesis was reached by accomplishing that:

The results show that it is possible to extract the information fused by using morphological transforms, avoiding unwanted artifacts. The performance of the proposed methods was compared against some state-of-the-art methods using the evaluation functions of contrast, entropy, color similarity, and saturation. Results have shown that the proposed method could fuse relevant information and details, keeping well-defined edges. Moreover, color oversaturation was avoided, regardless of the visible image's conditions, such as the area affected by the haze. The images resulting from the fusion method developed in this work can be used in other applications, such as segmentation.

6.1 Future work

Likewise, this doctoral thesis offers different perspectives on future work in the coming years on the same subject under study. Based on the results obtained, some future studies have been identified.

- The artifacts could stand out for the proposed methodology since it is based on a mathematical morphology transformation. For example, the fixed shape and size of the structuring element may generate artifacts. However, for the V-NIR EPFL database, only a few images have such conditions. To the best of our knowledge, ongoing research addresses structuring elements that dynamically change their shape depending on the processed region, which may significantly improve the results and reduce the artifacts generated in the images [Fang et al., 2020]. Thus, determining if changing the structuring element for an adaptive one can significantly improve the contrast in the images could be useful, also evaluating the computational cost versus the benefit of the adaptive structuring element.
- Another possible test could be to check what happens if a multiscale is used instead of a fixed scale. A test was carried out using the principles of the proposed methodology, where instead of defining a fixed scale, five scales obtained from granulometry are used. An example of the results is shown in the Figure 6.1.
- Use the proposed algorithm as a previous process for other applications, such as image segmentation, detection of objects, among others to verified the improvement. For this reason, a test was carried out using the original images and the images from the fused methods to determine the differences between the results of the segmentation process using the technique proposed in [Bampis et al., 2016]. This method uses segmentation based on the characteristics of the texture, and as a result, it merges the areas with similar textures labeling into five colored segments. The color scale for textured regions ranges from lower

(a)

(b)

(c)

Figure 6.1: (a) The original visible image, (b) Proposed method, and (c) multiscale method.

values (in blue) to higher values (in yellow), allowing visualize how the texture levels of the image change. As can be seen in Figure 6.2 (a) , there are lower texture values in the valley region of the original image, and in the mountains, the values are higher in texture. On the one hand, the proposed method presents more texture comparing to the other methods specially in the valley region, reflected in the segmented image (Figure 6.2 (b)). On the other hand [Vanmali et al., 2015, Vanmali and Gadre, 2017] present the lower values at the valley (Figure 6.2 (c-d)). While [Sharma et al., 2017] appears to have more textures in the mountain region, the regions with intermediate textures are reduced Figure 6.2 (f).

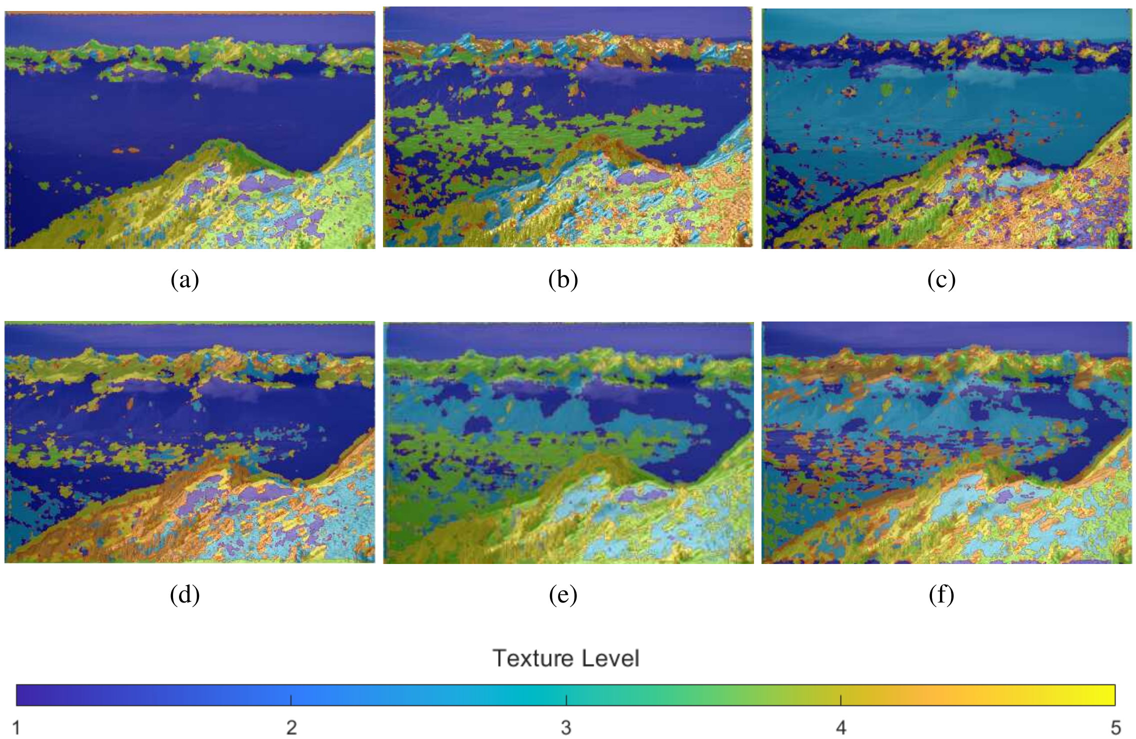


Figure 6.2: Segmentation by texture results: (a) Original, (b) Proposed method, (c) [Vanmali et al., 2015], (d) [Vanmali and Gadre, 2017], (e) [Herrera-Arellano et al., 2019], (f) [Sharma et al., 2017].

- Test the proposed fusion method on different bands of the electromagnetic spectrum, such as visible-infrared, due to infrared images can also avoid the influence of the external environment, such as sunlight, and smoke [Sun et al., 2020].And review what modifications or adaptations are necessary to obtain the desired results. Try also using visible-visible and making the necessary modifications or adaptations for obtaining the desired results.

- Explore another technique for information selection, e.g., generating a contrast map, salience map. Even proposed another fusion rule with different ratios.

Appendix A

Color space transformation arrays and algorithms

A.1 RGB to CIE L*a*b*

Convert RGB to CIE 1976 L*a*b*: L*,a*,b* The values corresponding to a stimulus with CIE XYZ [Sharma and Trussell, 1997].

$$L* = 116 f\left(\frac{Y}{Y_n}\right) - 16 \tag{A.1}$$

$$a* = 500\left(f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right) \tag{A.2}$$

$$b* = 200\left(f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right) \tag{A.3}$$

where: X_n, Y_n, Z_n are tristimuli of the white stimulus and

$$f(x) = \begin{cases} x^{\frac{1}{3}}; x > 0.008856 \\ 7.7787x + \frac{16}{116}; x \leq 0.008856 \end{cases} \tag{A.4}$$

A.2 RGB to YCbCr

Convert RGB to YC_rC_b where the components are derived directly from digital form [BT.601-5, 1995]

$$\begin{bmatrix} Er \\ Eg \\ Eb \end{bmatrix} = \begin{bmatrix} int(219R) + 16 \\ int(219G) + 16 \\ int(219B) + 16 \end{bmatrix} \tag{A.5}$$

$$Y = \frac{77}{256}E_r + \frac{150}{256}Eg + \frac{29}{256}Eb \tag{A.6}$$

$$C_r = \frac{131}{256}E_r + \frac{110}{256}Eg - \frac{21}{256}Eb + 128 \tag{A.7}$$

$$C_r = -\frac{44}{256}E_r - \frac{87}{256}Eg + \frac{131}{256}Eb + 128 \tag{A.8}$$

A.3 RGB to HSV

Convert the RGB values into HSV by using the algorithm proposed by [Smith, 1978].

$$R' = R/255 \tag{A.9}$$

$$G' = G/255 \tag{A.10}$$

$$B' = B/255 \tag{A.11}$$

$$V = max(R', G', B') \tag{A.12}$$

$$X = min(R', G', B') \tag{A.13}$$

Saturation calculation

$$S = \begin{cases} 0; V = 0 \\ \frac{V-X}{V}; V \neq 0 \end{cases} \tag{A.14}$$

Hue calculation

$$\begin{bmatrix} r \\ g \\ b \end{bmatrix} = \begin{bmatrix} \frac{V-R}{V-X} \\ \frac{V-G}{V-X} \\ \frac{V-B}{V-X} \end{bmatrix} \tag{A.15}$$

If $R = V$ then H=(if G=X then 5+b else 1-g);

If G=V then H=(if B=X then 1+r else 3-b);

else H=(if R=X then 3+g else 5-r);

$$H = \frac{H}{6} \tag{A.16}$$

HSV to RGB algorithm

$$H = 6 \times H \tag{A.17}$$

$$I = Floor(H) \tag{A.18}$$

where $Floor(x)$ is the integer just less than or equal x.

$$F = H - I \tag{A.19}$$

$$\begin{bmatrix} M \\ N \\ K \end{bmatrix} = \begin{bmatrix} V \times (1 - S) \\ V \times (1 - (S \times F)) \\ V \times (I - (S \times (I - F))) \end{bmatrix} \tag{A.20}$$

Switch on I into:

case 0: (R,G,B) = (V,K,M);

case 1: (R,G,B) = (N,V,M);

case 2: (R,G,B) = (M,V,K);

case 3: (R,G,B) = (M,N,V);

case 4: (R,G,B) = (K,M,V);

case 5: (R,G,B) = (V,M,N);

A.4 RGB to $l\alpha\beta$

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{A.21}$$

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{A.22}$$

The following transform from LMS to $l\alpha\beta$ by using Eq. A.23. These transformation matrices were proposed by [Ruderman et al., 1998].

$$\begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} logL \\ logM \\ logS \end{bmatrix} \tag{A.23}$$

To transfer $l\alpha\beta$ back to RGB

$$\begin{bmatrix} \mathbf{L} \\ \mathbf{M} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -2 & 0 \end{bmatrix} \begin{bmatrix} \frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & \frac{\sqrt{6}}{6} & 0 \\ 0 & 0 & \frac{\sqrt{2}}{2} \end{bmatrix} \begin{bmatrix} l \\ \alpha \\ \beta \end{bmatrix} \tag{A.24}$$

$$\begin{aligned} L &= 10^{\mathbf{L}} \\ S &= 10^{\mathbf{S}} \\ M &= 10^{\mathbf{M}} \end{aligned} \tag{A.25}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 4.4679 & -3.5873 & 0.1193 \\ -1.2186 & 2.3809 & -0.1624 \\ 0.0497 & -0.2439 & 1.2045 \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \tag{A.26}$$

Appendix B

Evaluations quality functions

B.1 Feature mutual information

A large Feature mutual information (FMI) value indicates that considerable feature information is transferred from source images to the fused image [Ma et al., 2019].

$$FMI = MI_{\acute{A},\acute{F}} + MI_{\acute{B},\acute{F}} \tag{B.1}$$

where $\acute{A}$,$\acute{B}$,and $\acute{F}$ denote the feature maps of infrared, visible, and fused images, respectively.

B.2 SSIM

Structural similarity index measure (SSIM) ranges into $[-1, 1]$, the best value is 1 and it occurs when $y_i = x_i$, for all $i = 1, 2, ..., N$ [Zhou Wang and Bovik, 2002].

$$Q = \frac{4\sigma_{xy}\bar{x}\bar{y}}{\sigma_x^2 + \sigma_y^2 \left[(\bar{x})^2 + (\bar{y})^2\right]} \tag{B.2}$$

- $X = \{x_i | i = 1, 2, ..., N\}$ and $y = \{y_i | i = 1, 2, ..., N\}$ are the original and the test image signals, respectively.

- $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$
- $\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})^2$, $\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$
- $\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})$

B.3 Anisotropic Quality Index

Anisotropic Quality Index method for determining the quality of digital images [Gabarda and Cristóbal, 2007].

$$\sigma(t) = \sqrt{\sum_{s=1}^{S} \left(\mu_t - \bar{R}(t,\theta_s)\right)^2 / S)} \tag{B.3}$$

- $\bar{R}[t,\theta_s] = \sum_n R_3[n,\theta_s]/M$ This equation provides a value of entropy $R_3[n,\theta_s]$ for each pixel, where $\theta_s \epsilon \, [\theta_1, \theta_2..., \theta_S]$ represents S different orientations taken to measure entropy.
- μ_t is the mean of the values $\bar{R}(t,\theta_s)$
- M is the image size.
- t count different images integrating the data set.
- The range in image t can be defined as $rg(t) = max\left\{\bar{R}(t,\theta_s)\right\} - min\left\{\bar{R}(t,\theta_s)\right\}$

B.4 MSE

Mean squared error (MSE) indicates a good fusion performance, which means that the fused image approximates to the source images and minimal error occurs in the fusion process [Ma et al., 2019].

$$MSE = \frac{MSE_{AF} + MSE_{BF}}{2} \tag{B.4}$$

- $MSE_{XF} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (X(i,j) - F(i,j))^2$
- M, N image weight and height respectively.

- MSE_{AF} denoted the dissimilarity between the fused and visible image.

- MSE_{BF} denoted the dissimilarity between the infrared and visible image.

B.5 PSNR

The larger Peak signal-to-noise ratio (PSNR), the closer the fused image is to the source images and the less distortion the fusion method produces [Ma et al., 2019].

$$PSNR = 10log_{10}\frac{r^2}{MSE} \tag{B.5}$$

Where r denoter the peak value of the fused image.

B.6 Colorfulness

$$CF = \sigma_{ab} + 0.94\mu_c \tag{B.6}$$

where:

- σ_a:Standard deviation along the a axis.

- σ_b:Standard deviation along the b axis.

- $\sigma_{ab} = \sqrt{\sigma_a^2 + \sigma_b^2}$: The trigonometric length of the standard deviation in lab space.

- $\mu_c = \frac{1}{N}\sum_{p=1}^{N} C_p$: The mean of chroma where: $p = 1$ to N. The image has N pixels, and $C_p = \sqrt{a^2 + b^2}$.

B.7 Color similarity ΔE_{00} definitions

$$\Delta L' = L_2^* - L_1^* \tag{B.7}$$

$$\Delta C' = C_2' - C_1' \tag{B.8}$$

$$\Delta H' = 2\sqrt{C_1' C_2'} sin\left(\frac{\Delta h'}{2}\right) \tag{B.9}$$

$$h' = \begin{cases} 0 & b_i^* = a_i' = 0 \\ tan^{-1}(b_i^*, a_i') & ; i = 1, 2 \end{cases} \tag{B.10}$$

$$C_i' = \sqrt{(a_i')^2 + (b_i^*)^2}; i = 1, 2 \tag{B.11}$$

$$a_i' = (1 + G)\, a_i^*; i = 1, 2 \tag{B.12}$$

$$G = 0.5\left(1 - \sqrt{\frac{\bar{C}_{ab}^{*7}}{\bar{C}_{ab}^{*7} + 25^7}}\right) \tag{B.13}$$

$$\bar{C}_{ab}^* = C_{1,ab}^* + C_{2,ab}^* \tag{B.14}$$

$$C_{i,ab}^* = \sqrt{a_i^* + b_i^*} \quad ; i = 1, 2 \tag{B.15}$$

$$S_H = 1 + 0.015\bar{C}'T \tag{B.16}$$

$$S_c = 1 + 0.045\bar{C}' \tag{B.17}$$

$$S_L = 1 + \frac{0.015(\bar{L}' - 50)^2}{\sqrt{20 + (\bar{L}' - 50)^2}} \tag{B.18}$$

Where:

$$R_T = -sin(2\Delta\theta)R_c \tag{B.19}$$

$$R_C = 2\sqrt{\frac{\bar{C}'^7}{\bar{C}'^7 + 25^7}} \tag{B.20}$$

$$\Delta\theta = 30exp\left\{-\left[\frac{\bar{h}' - 275°}{25}\right]^2\right\} \tag{B.21}$$

$$T = 1 - 0.7cos(\bar{h}' - 30°) + 0.24cos(2\bar{h}') + 0.32cos(3\bar{h}' + 6°) - 0.20cos(4\bar{h}' + 63°) \tag{B.22}$$

$$\bar{h}' = \begin{cases} \frac{h_1' + h_2'}{2} & |h_1' - h_2'| \leq 180°; C_1'C_2' \neq 0 \\ \frac{h_1' + h_2' + 360°}{2} & |h_1' - h_2'| > 180°; (h_1' + h_2') < 360°; C_1'C_2' \neq 0 \\ \frac{h_1' + h_2' - 360°}{2} & |h_1' - h_2'| > 180°; (h_1' + h_2') \geq 360°; C_1'C_2' \neq 0 \\ (h_1' + h_2') & C_1'C_2' = 0 \end{cases} \tag{B.23}$$

Appendix C

Results of: Mutual information, PSNR, RMSE, and SSIM

As mentioned in the Section 2.4, there are several metrics used in the literature. This section adds some of the most used in state of the art. The values shown in bold are the best results; according to these metrics, the results obtained by the first of the proposed methods ,[Herrera-Arellano et al., 2019], have the best results in most of the parameters.

Method	**Metric**			
	Mutual information ↑	PSNR ↑	RMSE ↓	SSIM↑
[Sharma et al., 2017]	2.846±0.724	66.735±1.759	0.015±0.005	1.755±0.116
[Vanmali et al., 2015]	**5.756±0.334**	65.645±1.276	0.015±0.005	1.715±0.163
[Vanmali and Gadre, 2017]	2.411±0.732	66.775±1.568	0.015±0.005	1.757±0.104
[Elliethy and Aly, 2017]	3.807±0.801	66.749±1.795	0.015±0.005	1.717±0.125
[Herrera-Arellano et al., 2019]	3.126±0.820	**66.992±1.854**	**0.014±0.005**	**1.771±0.118**
[Herrera-Arellano et al., 2021]	2.493±0.608	65.838±1.384	0.018±0.005	1.587±0.141

*Bold text indicated the best scores.

Bibliography

[Ahn et al., 2011] Ahn, B., Bae, T., and Kweon, I. (2011). Haze removal using visible and infrared image fusion. In *2011 8th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 813–814.

[Ancuti and Ancuti, 2013] Ancuti, C. O. and Ancuti, C. (2013). Single Image Dehazing by Multi-Scale Fusion. *IEEE Transactions on Image Processing*, 22(8):3271–3282.

[Bai, 2012] Bai, X. (2012). Image fusion through feature extraction by using sequentially combined toggle and top-hat based contrast operator. *Appl. Opt.*, 51(31):7566–7575.

[Bampis et al., 2016] Bampis, C. G., Maragos, P., and Bovik, A. C. (2016). Projective non-negative matrix factorization for unsupervised graph clustering. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 1255–1258, Phoenix, AZ, USA. IEEE.

[Barreto and Villalobos, 2011] Barreto, H. P. and Villalobos, I. R. T. (2011). Contrast Enhancement Based on a Morphological Rational Multiscale Algorithm. 14(3):15.

[Barsi et al., 2014] Barsi, J., Lee, K., Kvaran, G., Markham, B., and Pedelty, J. (2014). The Spectral Response of the Landsat-8 Operational Land Imager. *Remote Sensing*, 6(10):10232–10251.

[Brown and Süsstrunk, 2011] Brown, M. and Süsstrunk, S. (2011). Multispectral SIFT for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR11)*, pages 177–184, Colorado Springs.

[BT.601-5, 1995] BT.601-5, I.-R. (1982-1986-1990-1992-1994-1995). *BT.601-5 - Studio encoding parameters of digital television for standard 4:3 and wide-screen 16:9 aspect ratios*.

[Burger and Burge, 2016] Burger, W. and Burge, M. J. (2016). *Digital Image Processing: An Algorithmic Introduction Using Java*. Texts in Computer Science. Springer London, London.

[Burt and Adelson, 1987] Burt, P. J. and Adelson, E. H. (1987). The laplacian pyramid as a compact image code. In *Readings in computer vision*, pages 671–679. Elsevier.

[Cai et al., 2016] Cai, B., Xu, X., Jia, K., Qing, C., and Tao, D. (2016). Dehazenet: An end-to-end system for single image haze removal. *IEEE Transactions on Image Processing*, 25(11):5187–5198.

[Choe et al., 2018] Choe, G., Kim, S.-H., Im, S., Lee, J.-Y., Narasimhan, S. G., and Kweon, I. S. (2018). RANUS: RGB and NIR Urban Scene Dataset for Deep Scene Parsing. *IEEE Robotics and Automation Letters*, 3(3):1808–1815.

[Choi et al., 2005] Choi, M., Kim, R., Nam, M.-R., and Kim, H. (2005). Fusion of Multispectral and Panchromatic Satellite Images Using the Curvelet Transform. *IEEE Geoscience and Remote Sensing Letters*, 2(2):136–140.

[Choi et al., 2017] Choi, S., Kwon, O.-J., and Lee, J. (2017). A Method for Fast Multi-Exposure Image Fusion. *IEEE Access*, 5:7371–7380.

[Dougherty and Lotufo, 2003] Dougherty, E. R. and Lotufo, R. A. (2003). *Hands-on morphological image processing*, volume 59. SPIE press.

[Elliethy and Aly, 2017] Elliethy, A. and Aly, H. A. (2017). Fast near infrared fusion-based adaptive enhancement of visible images. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 156–160, Montreal, QC. IEEE.

[Fairchild, 2005] Fairchild, M. D. (2005). *Color appearance models*. John Wiley & Sons.

[Fang et al., 2020] Fang, C., Wang, X.-P., Wang, Q.-S., and Liang, J.-C. (2020). Construction Algorithm for Adaptive Morphological Structuring Elements Based on the Neighborhood Gray Difference Changing Vector Field and Relative Density. *IEEE Access*, 8:54928–54939.

[Fattal, 2008] Fattal, R. (2008). Single image dehazing. *ACM transactions on graphics (TOG)*, 27(3):72.

[Feng et al., 2013] Feng, C., Zhuo, S., Zhang, X., Shen, L., and Susstrunk, S. (2013). Near-infrared guided color image dehazing. In *2013 IEEE International Conference on Image Processing*, pages 2363–2367, Melbourne, Australia. IEEE.

[Firmenichy et al., 2011] Firmenichy, D., Brown, M., and Susstrunk, S. (2011). Multispectral interest points for RGB-NIR image registration. In *2011 18th IEEE International Conference on Image Processing*, pages 181–184, Brussels, Belgium. IEEE.

[Fredembach and Süsstrunk, 2008] Fredembach, C. and Süsstrunk, S. (2008). Colouring the near-infrared. In *Color and Imaging Conference*, volume 2008, pages 176–182. Society for Imaging Science and Technology.

[Gabarda and Cristóbal, 2007] Gabarda, S. and Cristóbal, G. (2007). Blind image quality assessment through anisotropy. *Journal of the Optical Society of America A*, 24(12):B42.

[Ghassemian, 2016] Ghassemian, H. (2016). A review of remote sensing image fusion methods. *Information Fusion*, 32:75–89.

[Gonzalez Rafael C. et al., 2009] Gonzalez Rafael C., Woods Richard E., and Eddins Steven L. (2009). *Digital Image Processing, 3a ed.* Prentice Hall, 3ra edition.

[Haralick et al., 1987] Haralick, R. M., Sternberg, S. R., and Zhuang, X. (1987). Image Analysis Using Mathematical Morphology. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-9(4):532–550.

[Hasler and Susstrunk, 2003] Hasler, D. and Susstrunk, S. (2003). Measuring colourfulness in natural images. 5007:87–95.

[Hassanpour et al., 2015] Hassanpour, H., Samadiani, N., and Mahdi Salehi, S. (2015). Using morphological transforms to enhance the contrast of medical images. *The Egyptian Journal of Radiology and Nuclear Medicine*, 46(2):481–489.

[Hawkes, 2004] Hawkes, P. W. (2004). *Advances in imaging and electron physics*. Elsevier.

[He et al., 2009] He, K., Sun, J., and Tang, X. (2009). Single Image Haze Removal Using Dark Channel Prior. page 8.

[He et al., 2011] He, K., Sun, J., and Tang, X. (2011). Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353.

[Herrera-Arellano et al., 2021] Herrera-Arellano, M., Peregrina-Barreto, H., and Terol-Villalobos, I. (2021). Visible-NIR Image Fusion Based on Top-Hat Transform. *IEEE Transactions on Image Processing*, 30:4962–4972.

[Herrera-Arellano et al., 2019] Herrera-Arellano, M. A., Peregrina-Barreto, H., and Terol-Villalobos, I. (2019). Color outdoor image enhancement by V-NIR fusion and weighted luminance. In *2019 IEEE Autumn Meeting on Power, Electronics and Computing, 2019. ROPEC 2019. IEEE, isbn = 978-1-7281-2898-6*, Ixtapa, Mexico.

[Ibraheem et al., 2012] Ibraheem, N. A., Hasan, M. M., Khan, R. Z., and Mishra, P. K. (2012). Understanding Color Models: A Review. 2(3):11.

[Jagalingam and Hegde, 2015] Jagalingam, P. and Hegde, A. V. (2015). A Review of Quality Metrics for Fused Image. *Aquatic Procedia*, 4:133–142.

[Jang and Park, 2017] Jang, D.-W. and Park, R.-H. (2017). Colour image dehazing using near-infrared fusion. *IET Image Processing*, 11(8):587–594.

[Jobson, 2004] Jobson, D. J. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13(1):100.

[Jung et al., 2020] Jung, C., Zhou, K., and Feng, J. (2020). Fusionnet: Multispectral Fusion of RGB and NIR Images Using Two Stage Convolutional Neural Networks. *IEEE Access*, 8:23912–23919.

[Kahu et al., 2019] Kahu, S. Y., Raut, R. B., and Bhurchandi, K. M. (2019). Review and evaluation of color spaces for image/video compression. *Color Research & Application*, 44(1):8–33.

[Kaur et al., 2021] Kaur, H., Koundal, D., and Kadyan, V. (2021). Image Fusion Techniques: A Survey. *Archives of Computational Methods in Engineering*.

[Kostakoglu et al., 2004] Kostakoglu, L., Hardoff, R., Mirtcheva, R., and Goldsmith, S. J. (2004). PET-CT Fusion Imaging in Differentiating Physiologic from Pathologic FDG Uptake. *RadioGraphics*, 24(5):1411–1431.

[Kudo and Kubota, 2018] Kudo, Y. and Kubota, A. (2018). Image dehazing method by fusing weighted near-infrared image. In *2018 International Workshop on Advanced Image Technology (IWAIT)*, pages 1–2, Chiang Mai. IEEE.

[Kumar et al., 2019] Kumar, W. K., Nongmeikapam, K., and Singh, A. D. (2019). Enhancing scene perception using a multispectral fusion of visible–near-infrared image pair. *IET Image Processing*, 13(13):2467–2479.

[Li et al., 2017] 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*, 33:100–112.

[Ma et al., 2019] Ma, J., Ma, Y., and Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. *Information Fusion*, 45:153–178.

[McKendry, 2002] McKendry, P. (2002). Energy production from biomass (part 1): overview of biomass. *Bioresource technology*, 83(1):37–46.

[Meyer, 1978] Meyer, F. (1978). *Contrast feature extraction*, volume 2. Riederer Verlag.

[Mokrzycki and Tatol, 2011] Mokrzycki, W. and Tatol, M. (2011). Colour difference δ e-a survey. *Mach. Graph. Vis*, 20(4):383–411.

[Pal et al., 2015] Pal, C., Chakrabarti, A., and Ghosh, R. (2015). A Brief Survey of Recent Edge-Preserving Smoothing Algorithms on Digital Images. *arXiv:1503.07297 [cs]*. arXiv: 1503.07297.

[Park, 2020] Park, H. (2020). An rgb-nir image fusion method for improving feature matching. *International Journal of Engineering and Technology*, 10(3):225–234.

[Perona and Malik, 1990] Perona, P. and Malik, J. (1990). Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639.

[Qu et al., 2002] Qu, G., Zhang, D., and Yan, P. (2002). Information measure for performance of image fusion. *Electronics Letters*, 38(7):313.

[Rad et al., 2007] Rad, A. A., Meylan, L., Vandewalle, P., and Süsstrunk, S. (2007). Multidimensional image enhancement from a set of unregistered and differently exposed images. page 649808, San Jose, CA, USA.

[Reinhard et al., 2001] Reinhard, E., Ashikhmin, M., Gooch, B., and Shirley, P. (2001). Color Transfer between Images. *IEEE Computer Graphics and Applications*, page 8.

[Roberts et al., 2008] Roberts, J. W., Van Aardt, J. A., and Ahmed, F. B. (2008). Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(1):023522.

[Ruderman et al., 1998] Ruderman, D. L., Cronin, T. W., and Chiao, C.-C. (1998). Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15(8):2036.

[Russ and Neal, 2015] Russ, J. and Neal, F. (2015). *The Image Processing Handbook, Seventh Edition*. CRC Press.

[Sadeghipoor et al., 2015] Sadeghipoor, Z., Lu, Y. M., Mendez, E., and Susstrunk, S. (2015). Multiscale guided deblurring: Chromatic aberration correction in color and near-infrared imaging. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, pages 2336–2340, Nice. IEEE.

[Salamati et al., 2011] Salamati, N., Sadeghipoor, Z., and Süsstrunk, S. (2011). Analyzing near-infrared images for utility assessment. page 786508, San Francisco Airport, California, USA.

[Santibañez, 2007] Santibañez, J. D. M. (2007). Segmentación y análisis granulométrico de sustancia blanca y gris en IRM para el estudio del estrabismo usando transformaciones morfológicas. *Revista Mexicana de Ingeniería Biomédica*, page 13.

[Schaul et al., 2009] Schaul, L., Fredembach, C., and Susstrunk, S. (2009). Color image dehazing using the near-infrared. In *2009 16th IEEE International Conference on Image Processing (ICIP)*, pages 1629–1632, Cairo. IEEE.

[Serra, 1986] Serra, J. (1986). Introduction to Mathematical Morphology. *Computer vision, graphics, and image processing*, page 23.

[Sharma and Trussell, 1997] Sharma, G. and Trussell, H. (1997). Digital color imaging. *IEEE Transactions on Image Processing*, 6(7):901–932.

[Sharma et al., 2005] Sharma, G., Wu, W., and Dalal, E. N. (2005). The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, 30(1):21–30.

[Sharma and Gool, 2016] Sharma, V. and Gool, L. (2016). Does v-nir based image enhancement come with better features? *ArXiv*, abs/1608.06521.

[Sharma et al., 2017] Sharma, V., Hardeberg, J. Y., and George, S. (2017). RGB–NIR Image Enhancement by Fusing Bilateral and Weighted Least Squares Filters. *Color and Imaging Conference*, 2017(25):330–338.

[Shi et al., 2017] Shi, H., Kwok, N., Fang, G., Lin, S. C.-F., Lee, A., Li, H., and Yu, Y.-H. (2017). Gradient-guided color image contrast and saturation enhancement. *International Journal of Advanced Robotic Systems*, 14(3):172988141771168.

[Smith, 1978] Smith, A. R., editor (1978). *Color Gamut transform pairs*. SIGGRAPH 78 Conference Proceed.

[Smith and Heather, 2005] Smith, M. I. and Heather, J. P. (2005). A review of image fusion technology in 2005. In Peacock, G. R., Burleigh, D. D., and Miles, J. J., editors, *Thermosense XXVII*, volume 5782, pages 29 – 45. International Society for Optics and Photonics, SPIE.

[Soille, 2013] Soille, P. (2013). *Morphological image analysis: principles and applications*. Springer Science & Business Media.

[Son et al., 2015] Son, C., Zhang, X., and Lee, K. (2015). Near-infrared coloring via a contrast-preserving mapping model. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 677–681.

[Son and Zhang, 2017] Son, C.-H. and Zhang, X.-P. (2017). Near-Infrared Coloring via a Contrast-Preserving Mapping Model. *IEEE Transactions on Image Processing*, 26(11):5381–5394.

[Son and Zhang, 2018] Son, C.-H. and Zhang, X.-P. (2018). Near-Infrared Fusion via Color Regularization for Haze and Color Distortion Removals. *IEEE Transactions on Circuits and Systems for Video Technology*, 28(11):3111–3126.

[Sun et al., 2020] Sun, C., Zhang, C., and Xiong, N. (2020). Infrared and Visible Image Fusion Techniques Based on Deep Learning: A Review. *Electronics*, 9(12):2162.

[Swoger et al., 2007] Swoger, J., Verveer, P., Greger, K., Huisken, J., and Stelzer, E. H. (2007). Multi-view image fusion improves resolution in three-dimensional microscopy. *Optics Express*, 15(13):8029.

[Tan, 2008] Tan, R. T. (2008). Visibility in Bad Weather from A Single Image. page 8.

[Tang and Mat Isa, 2014] Tang, J. R. and Mat Isa, N. A. (2014). Adaptive Image Enhancement based on Bi-Histogram Equalization with a clipping limit. *Computers & Electrical Engineering*, 40(8):86–103.

[Tomasi and Manduchi, 1998] Tomasi, C. and Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*, pages 839–846, Bombay, India. Narosa Publishing House.

[Trinidad et al., 2019] Trinidad, M. C., Martin-Brualla, R., Kainz, F., and Kontkanen, J. (2019). Multi-View Image Fusion. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4100–4109, Seoul, Korea (South). IEEE.

[Triola, 2018] Triola, M. (2018). *Elementary Statistics*. What's New in Statistics. Pearson Education.

[Uzal, 2019] Uzal, J. (2019). *Diagnóstico por imagen en bandas no visibles sobre patrimonio cultural: Una aproximación a la imagen infrarroja, ultravioleta, fluorescencias y análisis de imagen*. Books on Demand.

[Vanmali and Gadre, 2017] Vanmali, A. V. and Gadre, V. M. (2017). Visible and nir image fusion using weight-map-guided laplacian–gaussian pyramid for improving scene visibility. *Sādhanā*, 42:1063–1082.

[Vanmali et al., 2015] Vanmali, A. V., Kelkar, S. G., and Gadre, V. M. (2015). A novel approach for image dehazing combining visible-NIR images. In *2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pages 1–4, Patna, India. IEEE.

[Vincent, 1994] Vincent, L. M. (1994). Fast opening functions and morphological granulometries. In *Image Algebra and Morphological Image Processing V*, volume 2300, pages 253–267. International Society for Optics and Photonics.

[Whetsel, 1968] Whetsel, K. B. (1968). Near-infrared spectrophotometry. *Applied Spectroscopy Reviews*, 2(1):1–67.

[Young et al., 1998] Young, I. T., Gerbrands, J. J., Vliet, L. J. v., and TU Delft, F. d. T. N. (1998). *Fundamentals of image processing: [dictaat behorende bij college et2720in*. Delft University of Technology, Delft. OCLC: 840370757.

[ZHANG Chang Jiang, 2004] ZHANG Chang Jiang, FU Meng Yin, J. M. Z. Q. H. (2004). Approach to enhance contrast of infrared image based on wavelet transform. *Journal of Infrared and Millimeter Waves*, 23[2]:119–124.

[Zhang Jingyun et al., 2016] Zhang Jingyun, Ding Yifan, Yang Yi, and Sun Jiasong (2016). Realtime defog model based on visible and near-infrared information. In *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6, Seattle, WA, USA. IEEE.

[Zhengguo Li et al., 2013] Zhengguo Li, Jinghong Zheng, Zijian Zhu, Wei Yao, Shiqian Wu, and Rahardja, S. (2013). Content adaptive bilateral filtering. In *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, pages 1–6, San Jose, CA, USA. IEEE.

[Zhou Wang and Bovik, 2002] Zhou Wang and Bovik, A. (2002). A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84.

[Zhu et al., 2015] Zhu, Q., Mai, J., and Shao, L. (2015). A fast single image haze removal algorithm using color attenuation prior. *IEEE Transactions on Image Processing*, 24(11):3522–3533.

[Zhuo et al., 2010] Zhuo, S., Zhang, X., Miao, X., and Sim, T. (2010). Enhancing low light images using near infrared flash images. In *2010 IEEE International Conference on Image Processing*, pages 2537–2540, Hong Kong, Hong Kong. IEEE.

[Zitová and Flusser, 2003] Zitová, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000.

[Zwinkels, 2014] Zwinkels, J. (2014). *Light, Electromagnetic Spectrum*, pages 1–8. Springer Berlin Heidelberg, Berlin, Heidelberg.