

Hasselt University
and transnational University of Limburg
School of Information Technology

Perceptual Decolorization and Dehazing of Images
and Videos

Dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science
Hasselt University, Belgium
to be defended by

Codruta Orniana Ancuti

on December 6, 2011

Supervisor: Prof. dr. Philippe Bekaert

Universiteit Hasselt
en transnationale Universiteit Limburg
School voor Informatietechnologie

Perceptual Decolorization and Dehazing of Images
and Videos

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen, richting Informatica
aan de Universiteit Hasselt
te verdedigen door

Codruta Orniana Ancuti

op December 6, 2011

Promotor: Prof. dr. Philippe Bekaert

To my beloved son Radu Philippe

Acknowledgments

I owe much to my supervisor Prof. Dr. Philippe Bekaert for giving me the opportunity and the support to work on this very interesting research project. His insightful remarks and suggestions were invaluable for me during this period. I have learned much that will last longer than this dissertation.

I would like to thank to the EDM staff, in particular prof. dr. Eddy Flerackers, managing director, prof. dr. Frank Van Reeth and Peter Vandore. During the undertaking of this project, the faculty staff have been very supportive and have provided a wonderful working environment. Roger and Ingrid deserve also special recognition for their assistance.

I want to warmly acknowledge the members of the group: prof. dr. Fabian Di Fiore, dr. Tom Mertens, dr. William Van Haevre, dr. Tom Van Laerhoven, dr. Cedric Vanaken, Tom Haber, Tom Cuypers, dr. Bert Dedecker, Karel Frederix, Maarten Dumont, Steven Maesen, Sammy Rogmans, Johannes Taelman, Nick Michiels but also to my former colleagues: dr. Erik Hubo, dr. Chris Hermans, dr. Yannick Francken, dr. Koen Beets, dr. Jan Fransens, Mark Gerrits and Johan Nulens. Particularly, for their opinions, collaboration and fun, I would like to mention Cedric, Erik, Mark, Chris and Tom. I want to thanks to everyone that participated into my user studies, for their patience and cooperation.

This section will not be complete if I did not acknowledge my family for their appreciation and encouragement throughout this long journey and my husband Cosmin for his love and support.

Abstract

The mastering craft of photography aspiration is to communicate *the equivalent of what we saw and felt* (Alfred Stieglitz). The key to obtain a satisfactory image representation lies in reproducing the expressive scene characteristics. Although photographic work is often regarded as a literal transcription of the reality, due to the depiction constrains, artists perform tedious work and employ many photographic controls to obtain a *realistic* representation.

In this work we address the problem of enhancing images by contrast manipulation. The importance of contrast holds not just for the perceived dimensionality of lightness but for colors as well. Therefore, proper techniques for contrast manipulation in the image will automatically yield desired adjustments over perceived image appearance. The presented work proposes new ways of transforming the image color, style and appearances. Our work is motivated by several photographic and artistic techniques being validated by perceptual studies.

In the first part, we introduce an algorithm that decolorize images and videos guided by the original saliency. The method is inspired by the Hering's opponent process theory and aims to increase the contrast of the regions of interest, rather than over the entire image. We have based our approach on the assumption that preserving these salient regions in the converted image will result in a better preservation of the visual contrast and overall perceptual appearance. After the monochromatic luminance channel is filtered and stored as a reference, the luminance values are computed pixel-wise by mixing both saturation and hue values, creating a new spatial distribution with an increased contrast of the interest regions. All the pre-computed values are normalized in order to fit the entire intensity range while the intensity is re-balanced in order to conserve the local contrast in the initial image. Since our decolorization is accurate and preserves finest details, we can exploit variations in chromacity as well as luminance for application such as video decolorization, segmentation under different illuminants, detail enhancement, wide-baseline image matching and auditory substitution systems.

In the subsequent part we present a novel single image dehazing strategy built on the well known dark object (Chavez 88) principle. Since images taken in bad weather conditions are characterized by poor contrast, the local values are less likely to change abruptly except depth discontinuities. By an extensive study it has been disclosed an important difference between hazy and non-hazy image regions, by performing a per pixel comparison of the hue values in the original image to their values in a 'semi-inversed' image. This 'semi-inversed' image version is obtained by replacing the RGB values of each pixel on a per channel basis by the maximum of the initial channel value (r , g or b) and its inverse ($1 - r$, $1 - g$ or $1 - b$), followed by an image-wide renormalization. This facilitates the estimation of the airlight constant color, and enables us to compute a good approximation of the haze-free image using a layer-based approach.

Finally, we present two novel fusion-based techniques that deal with image decolorization and image dehazing. We first demonstrate that by defining proper inputs and weight maps, a fusion-based strategy can yield accurate decolorized images, in which the original discriminability and appearance of the color images are well preserved. Besides the independent R, G, B channels, we employ as well an additional input channel that conserves color contrast, based on the Helmholtz-Kohlrausch effect. The algorithm employs three different weight maps in order to control saliency, exposure and saturation. In order to minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion, using a Laplacian pyramid representation of the inputs combined with Gaussian pyramids of normalized weights. The second fusion-based technique deals with single image dehazing (no additional information such as hardware and images are required). The method employs a fusion-based strategy that takes as inputs two adapted versions of the original image that are weighted by specific maps in order to yield accurate haze free results. The method computes in a per-pixel fashion and is straightforward to be implemented.

Comprehensive experiments and extensive comparison with the existing state-of-the-art related techniques demonstrate the accuracy and utility of all our novel methods.

Contents

Acknowledgments	iv
Abstract	vii
Contents	1
1 Introduction	3
1.1 Problem Statement and Motivation	3
1.2 Summary of Contributions	8
1.3 Overview of the Dissertation	10
2 Background Theory: Light, Color and Contrast	13
2.1 Lightness Perception	14
2.2 Global Contrast	15
2.3 Local Contrast	16
2.4 Multiscale Local Contrast	18
2.5 Helmholtz-Kohlrausch Effect	19
2.6 Contrast Sensitivity Measurement and Spatial Frequency	19
2.6.1 Image Quality Assessment (IQA) Metric	20
2.7 Visual Saliency	21
2.8 Contrast Visual Effects	23
2.8.1 Contour Contrast Illusions	23
2.8.2 Simultaneous Contrast Illusion	24
2.9 Color Perception	25
2.10 Theories of color vision	27
2.11 Color Spaces	30
2.11.1 <i>RGB/CMYK</i> Color Spaces	30
2.11.2 Perceptual Color Spaces	32
2.11.3 <i>HSL, HSV, HSI</i> Color Spaces	35

3	Perceptual Decolorization of Images and Videos	37
3.1	Introduction	38
3.2	Related Work	41
3.3	Saliency-guided Decolorization	43
3.3.1	Overview	43
3.3.2	Our Decolorization Approach	43
3.3.3	Offset Angle Selection	47
3.3.4	Results and Discussion	55
3.3.5	Validation	60
3.4	Video Decolorization	61
3.5	Other Applications	62
3.5.1	Detail Enhancing	63
3.5.2	Segmentation under Different Illuminants	66
3.5.3	Matching by Local Feature Points	66
3.5.4	Decolorizing Images for Auditory Substitution Systems	68
3.6	Conclusions	72
4	Single Image Dehazing	73
4.1	Introduction	74
4.2	Related Work	76
4.3	Optical Model	77
4.4	Single Image Dehazing by Per-Pixel Haze Detection	78
4.4.1	Haze Detection	79
4.4.2	Airlight Color (A_∞) Estimation	82
4.4.3	Layer-based Dehazing	83
4.4.4	Experimental Results and Discussion	85
4.4.5	IQA Assesment of Dehazing Methods	90
4.5	Summary	92
5	Enhancing by Fusion-based Techniques	93
5.1	Introduction	94
5.2	Image Decolorization by Fusion	96
5.2.1	Overview	96
5.2.2	Our Algorithm	97
5.2.3	$H - K$ Chromatic Adapted Lightness	98
5.2.4	Weight Maps Assignment	99
5.2.5	Multi-scale Fusion of the Inputs	101
5.2.6	Results and Discussion	104
5.2.7	Video Decolorization	105
5.2.8	Evaluation of Grayscale Operators	105

CONTENTS **1**

5.3	Fusion-based Single Image Dehazing	111
5.3.1	Overview	111
5.3.2	Definition of the Inputs	112
5.3.3	Weight Maps	113
5.3.4	Fusion Process	114
5.3.5	Results and Discussion	116
5.4	Summary	119
6	Conclusions	121
6.1	Summary of Contributions	122
6.2	Future Work	123
A	Scientific Contributions and Publications	127
B	Samenvatting (Dutch Summary)	131
	Appendices	133
	Bibliography	133

Chapter 1

Introduction

Contents

1.1	Problem Statement and Motivation	3
1.2	Summary of Contributions	8
1.3	Overview of the Dissertation	10

1.1 Problem Statement and Motivation

Photography, as the art of *drawing with light*, is the process of recording visually meaningful changes in the light leaving a scene. The philosophy of photography is that the lens may never lie, however in a certain degree it may bend the truth. The difference between camera and the human eye is mainly due to the differences in aptitude to adapt to the acquired light. The human eye is able to adapt locally, while the camera adjusts globally. The perception of lightness and color is in general influenced by the existing contrast among regions and their adjacent surroundings. Due to the light adaptation ability, the human visual system (HVS) is hypothetically more sensitive to local changes being quite insensitive to absolute global values under normal illumination conditions (Schubert 92). As suggested by Edwald Hering and later demonstrated by Hans Wallach (1948) (Palmer 99), the process of perceiving lightness is greatly influenced by relative contrast. Therefore, our way of seeing is not just a simple registration procedure of the luminance, but a complex balance and processing of the incoming light and the yielded contrast.

Nowadays, despite of the tremendous progress, the existing digital cameras still show some limitations. For example, when taking photos of a scene with high dynamic range of intensities the cameras will search to average

out the exposure for the entire photo. The compromise solution, performed by many photographers, is to select manually the proper exposure for the specific part of the scene that is interested to be emphasized. However, in both cases the dynamic range leads to under- or over- exposure for the regions of the scene that are not well exposed due to this lock exposure operation. Basically, when a high exposure is employed, visual details of dark regions are well depicted while bright regions are saturated. On the other hand, when low exposures are employed, visual details of bright regions are well imaged, but dark regions appears in general too dark. This pretty much trial process often may produce images that do not correspond to the real appearance of the scene. To deal with this problem, digital cameras manufacturers have proposed various schemes that aim to compute the appropriate exposure value or more complex combinations such as Auto Exposure Bracketing (AEB) that merges three different exposed frames into one image that searches to find the match of the closest exposure as possible (Raskar 06). Modern image acquisition employs High Dynamic Range(HDR) photos that are obtained by merging a series of images captured with different exposure times (Morimura 93; Burt 93; Tsai 93; Madden 93; Mertens 09) or more complex by estimating the radiometric response function of an imaging device before the fast merging operation (Mann 95; Debevec 97; Mitsunaga 99).

Although the progress of digital cameras persists, software post-processing strategies are still required in many cases. Due to the exposure, aperture, tone, color range and as well display medium there are still several long-standing limitations (since film-style photography age) such as constraints of dynamic range, depth of field, field of view, resolution and many others.

In the classical approach to obtain a pleasing *realistic* depiction of the scene, the artists are manually employing various techniques into the dark-room. On the same note, the newly emerging *Computational Photography* field aims to produce a richer visual experience beyond of just a simple set of pixels and to present the acquired information more machine readable. Contrast is a fundamental attribute that influences the visual communication. Contrast is often used by artists for content emphasis and to change the presented information into a more legible composition. For viewers, regions characterized by higher contrast will focus their attention. In images, contrast underlines the message being an indispensable compromise since the dynamic range of our visual system is several orders of magnitude greater than the ones available to image reproduction systems.

Therefore, it is unsurprising that contrast based manipulation techniques have been commonly involved in many applications that generally converts the raw data from the sensors into more elaborated outputs. For instance,

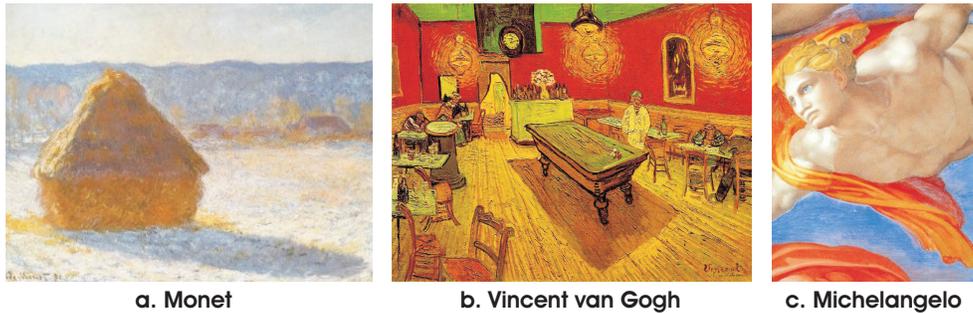


Figure 1.1: Opponent color contrast was used by famous painters.

appropriate processing might restore the image contrast from multiple images degraded by weather taken with different sensors (Schaul 09) or taken with different polarization angles (Treibitz 09; Shwartz 06). Images shot by different sensors are combined together to produce a better contrast rendering (Burt 93). Similarly, images taken in various illumination conditions are mixed in order to obtain surrealist images and videos (Raskar 04). As a well-known example, panoramic images are achieved by stitching a series of overlapping images that often involve contrast adjustment to obtain visually imperceptible transitions (Brown 07). Furthermore, tone mapped images are obtained by combining multiple exposures accounting for local contrast characteristics (Mertens 09).

In this dissertation we explore several novel techniques mainly founded on the contrast analysis and manipulation. Basically, we have investigated the impact of contrast manipulation for two important operators: color to grayscale (decolorization) transformation of the images/videos and the restoration of the hazy (foggy) images. Our aim is to overcome the existing constraints imposed by the image content and depiction. Since low image contrast limits the amount of information as they are conveyed to the viewers, images with higher contrast are in general preferred being considered perceptually more accurate. Moreover, a higher contrast has been demonstrated to simplify the comprehension of the visual stimulus and to increase the dynamic range of the images like in the real scenes (Yoshida 05).

A great deal of work on this topic has been inspired by the visual perception. This field investigates the way of our brain interprets the signals received from the eye. The knowledge of the human visual system facilitates scientists to understand the image content, to obtain images that convey realistically the visual sensation generated by the scene, or even more can surpass the visual

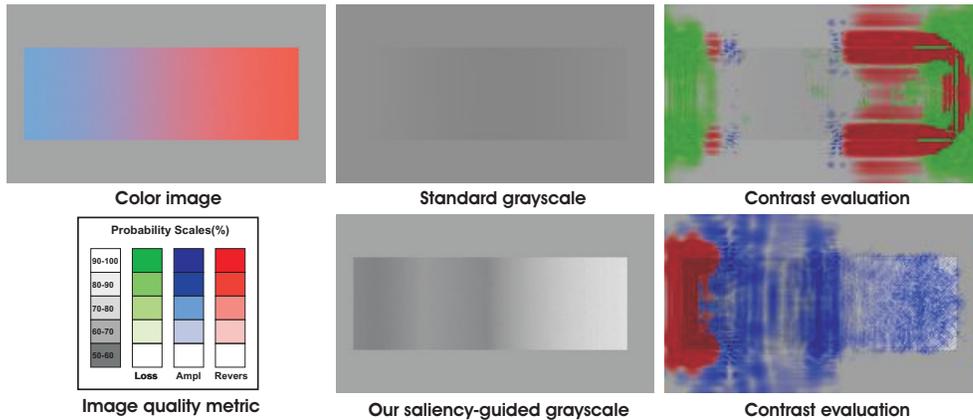


Figure 1.2: Standard grayscale conversion is unable to preserve the original contrast.

acuity performances obtained by our system (e.g. restoring images ruined by haze or fog). Similarly, as will be shown in the next chapters, parts of our work have been inspired as well from perceptual assumptions and observations.

Contrast has been defined in various ways, however, the basic definition is that the contrast reflects the disparity in luminance or color between two regions (next chapter elaborates about this topic). Chromatic contrast is tightly connected with the *opponent process*. Introduced by Edward Hering, aiming to explain the phenomenology of color perception, the color *opponent process* has numerous implications in visualization and color technology (we detail the subject in section 2.10). For centuries, artists have explored the subtleties of color opponency effects to produce different psychological effects (Livingstone 08). Using blue for the shadows and juxtaposing yellow in the near field, Monet has increased the apparent contrast. Red and green accentuate each other dramatically as has been illustrated by Vincent van Gogh in *Night Cafe in Arles*. Michelangelo Buonarroti has also combined colors to create mid-tones and light tones for example, adding yellow highlights to an orange robe (note the produced visual effect in Figure 1.1). We as well employ the *opponent process* to perceptually decolorize images as will be demonstrated in chapter 3.

In addition, *local contrast* represents a valuable information that is associated with contours related effects that not surprisingly can influence in a certain degree apparent contrast of the image. As an important application is multi-scale (pyramid) image decomposition that we have exploit in our fusion-based techniques (see chapter 5).

As we will demonstrate, the way of preserving and adjusting these contrast



Figure 1.3: Images degraded by haze are characterized by poor contrast and lower saturation.

relations could have a great influence to the *apparent contrast* of the image. The main goal of our approaches is to detect the loss of contrast and to produce images with higher contrast.

Grayscale conversion. Although color is an important image descriptor, there are several applications such as printing, aesthetic stylization and also computer vision applications that employs grayscale images. When the image attributes such as contrast and saliency are depicted by the chrominance channels, standard conversions that discharge the color information will produce images that will not maintain the original visual appearance of the image and the discriminability between different regions. Furthermore, since the straightforward conversion averages the values of the three image channels the resulted image in many cases presents a flat appearance. This limitation is more prominent in general in the regions where the contrast is already degraded. This assumption is verified by applying the image quality assessment (IQA) metric of Aydin et al. (Aydin 08) (please refer to section 2.6.1 for details about this metric). In Figure 1.2 notice regarding the standard grayscale conversion the loss of visible contrast marked with green color and the reverse contrast marked with red color. To overcome the limitations of the standard conversions we introduced two perceptually accurate decolorizing methods presented in chapter 3 and chapter 5. As can be observed in Figure 1.2 our perceptual saliency-driven grayscale conversion increases the contrast (the IQA blue is more prominent).

Restoration of hazy images. Contrast manipulation has found utility as well for the problem of restoration of images degraded by atmospheric phenomena. Images taken in outdoor environments might be degraded by phenomena such as fog and haze. Scattering and absorption influence the reflected light of distant objects reaching the camera lens. As a result, images taken in bad weather conditions (or similarly, underwater and aerial photographs) are

characterized by poor contrast, lower saturation and additional noise. Notice in Figure 1.3 the contrast reversal (red) and amplification (blue) by applying the same image quality assessment (Aydin 08) metric between the hazy and restored images. We give an effective solution to this problem in chapter 4 that describes a fast approach to detect and remove the haze from a single image. Moreover, in chapter 5 we introduce an original fusion-based single image dehazing strategy.

1.2 Summary of Contributions

This dissertation presents several contrast-based algorithms that are able to perceptually enhance the original image appearance. We have focused mainly to color-to-grayscale conversion and single image dehazing, but several other applications are considered as well.

Color to Grayscale Guided by Saliency Firstly, we introduce an effective color-to-grayscale algorithm that preserves the image image appearance by primary searching to maintain the contrast in the salient regions. Guided by the original saliency, the method blends the luminance and the chrominance information in order to conserve the initial color disparity while enhancing the chromatic contrast. As a result, our straightforward fusing strategy generates a new spatial distribution that discriminates better the illuminated areas and color features. Since we do not employ quantization or a per-pixel optimization (computationally expensive), the algorithm has a linear runtime. Extensive experiments and a comprehensive evaluation against existing state-of-the-art methods demonstrate the potential of our grayscale operator. Furthermore, since the method accurately preserves the finest details while enhancing the chromatic contrast, the utility and versatility of our operator have been proved for several other challenging applications such as video decolorization, detail enhancement, single image dehazing, segmentation under different illuminants, image matching by local feature points and sound substitution systems.

A Fast Semi-Inverse Approach to Detect and Remove the Haze from a Single Image Restoration of image degraded by atmospheric phenomena such as fog or haze is solved in general by employing either additional im-

ages/hardware or expensive prior-based solutions. We present a novel single image dehazing algorithm that allows for fast identification of hazy regions of an image, without making use of expensive optimization and refinement procedures. By applying a single per pixel operation on the original image, we produce a semi-inverse of the image. Based on the hue disparity between the original image and its semi-inverse, we are then able to identify hazy regions on a per pixel basis. This enables for a simple estimation of the airlight constant and the transmission map. Our approach is based on an extensive study on a large data set of images, and validated based on a metric that measures the contrast but also the structural changes. The algorithm is straightforward and performs faster than existing strategies while yielding comparative and even better results. We also provide a comparative evaluation against other recent single image dehazing methods, demonstrating the efficiency and utility of our approach.

Fusion-based Image and Video Decolorization Fusion is a well-studied topic in computer graphics that has found many useful applications, such as interactive photomontage (Agarwala 04), image editing (Perez 03), image compositing (Brinkmann 99; Grundland 06) and HDR imaging (Burt 93), (Mertens 09). However, we are the first that demonstrate the utility of fusion to effectively decolorize images and videos. Defining proper inputs and weight maps, our fusion-based strategy proves to yield accurate decolorized images, in which the original discriminability and appearance of the color images are well preserved. The algorithm takes four inputs, beside the independent R, G, B channels, we have additionally considered an input channel that represents color contrast, based on the Helmholtz-Kohlrausch effect. The algorithm uses three different weight maps in order to control saliency, exposure and saturation. The concept of this approach is that guided by our defined quality measures, the algorithm select the best pixels from the inputs and combine them into the final result. The algorithm is designed as a multi-scale approach and therefore prevents potential artifacts that could be introduced by applying the weight maps in a per pixel fashion.

Single Image Dehazing by Fusion We also demonstrate that the image fusion technique is suitable to restore hazy images. The input of the fusion algorithm are two modified versions of the original image that represent the white balance version of the original hazy image and an adapted version that presents increase contrast into the initial hazy regions. In the fusion pro-

cess, the inputs are weighted by three specific computed maps for luminance, saliency and chromatic information, in order to generate the most prominent detected features into the final result. The method is effective since it computes in a per-pixel fashion and demonstrates to yield comparative and even better results than the more complex state-of-the-art techniques. It is suitable for real-time applications.

1.3 Overview of the Dissertation

The dissertation is structured as follows:

Chapter 2 provides a background for vision and lightness perception. We introduce a general description on perceived lightness. We review the various existing definitions for global contrast and local contrast. We present the color constancy mechanism and the three main color theories : the trichromatic theory, the color opponency and the dual process theory. In the final part we briefly review several important color spaces.

In *chapter 3* we first discuss the problem and the related work of color to grayscale techniques. Afterwards, we introduce a novel contrast preservation algorithm driven by the saliency information and built on the color opponency principle. To validate the perceptual accuracy of our operator we performed an extensive user evaluation of the existing state-of-the-art grayscale transformations where the participants had the entire control to analyze and compare the results of different methods. The utility of our operator has been proved for several other challenging applications (e.g. video decolorization, detail enhancement, single image dehazing, segmentation under different illuminants, image matching).

Chapter 4 deals with the restoration of the images degraded by atmospheric phenomena such as fog or haze. We describe an original single image dehazing technique that allows for fast identification of hazy regions of an image in a pixel-wise fashion. The accuracy of our dehazing technique has been demonstrated by a comprehensive comparison with the recent single image dehazing methods.

In *chapter 5* we analyze the fusion-based techniques that recently have been

employed in several applications. We first present a novel fusion-based image decolorization technique that is able to preserve accurately the original appearance and discriminability of the colors in the converted image version. As well, we present a fusion-based solution for single image dehazing. For both fusion-based techniques (decolorization and dehazing) we provide extensive evaluations and comparisons against the recent related techniques.

A brief summary and as well a discussion of several future directions are presented in *chapter 6*.

Chapter 2

Background Theory: Light, Color and Contrast

Contents

2.1	Lightness Perception	14
2.2	Global Contrast	15
2.3	Local Contrast	16
2.4	Multiscale Local Contrast	18
2.5	Helmholtz-Kohlrausch Effect	19
2.6	Contrast Sensitivity Measurement and Spatial Frequency	19
2.6.1	Image Quality Assessment (IQA) Metric	20
2.7	Visual Saliency	21
2.8	Contrast Visual Effects	23
2.8.1	Contour Contrast Illusions	23
2.8.2	Simultaneous Contrast Illusion	24
2.9	Color Perception	25
2.10	Theories of color vision	27
2.11	Color Spaces	30
2.11.1	<i>RGB/CMYK</i> Color Spaces	30
2.11.2	Perceptual Color Spaces	32
2.11.3	<i>HSL, HSV, HSI</i> Color Spaces	35

This chapter discusses the physiological impact of the contrast and as well general aspects of the luminosity and color perception. Besides defining and explaining fundamental terms that we are going to use later in this thesis, the content of this chapter argues that the contrast plays a crucial role in the humans visual perception (Palmer 99). Practically, this chapter represents the background theory to understand our techniques that will be described in the next chapters.

After discussing the perception of lightness, in the next sections we review various definitions of the contrast, introduce the contrast sensitivity measurement and as well discuss several known visual effects of the contrast on the perceive lightness, like contour contrast and simultaneous contrast illusions. In the last part is discussed the color perception and a brief theory of color spaces.

2.1 Lightness Perception

Color vision is the ability to distinguish changes in the wavelength composition of a visual stimulus, that ultimately let us to discriminate objects based on the wavelengths (or frequencies) of the light they reflect, emit or transmit. Vision is distinctive from any other sensory modality due to the unique experiences that arises from color and light perception. The way we experience the surface color is very different in structure than the physical light. A physicist requires a large number of components to describe the spectrum of a given light, while in comparison, a psychologist needs only three values. This reduction of information complexity appears due to the fact that surface color can be described by people with normal vision in three dimensions: lightness, hue and saturation. This involves that color vision loses a certain amount of information since a similar perception might arise from lights with different spectrums.

Luminosity or *perceived lightness* is a term that characterizes the way how human visual system (HVS) reacts to and interprets the light. Even though the scene light reflected by an object can be easily measured, it is relatively difficult to estimate its impact on the HVS. Understanding how HVS reacts to the light could help scientists to have a more clear perspective about several complex tasks such as depth estimation, spatial organization, motion and more others, which are carried out by different parts of our visual system.

Lightness constancy is the property of the HVS to perceive a surface consistent as having the same lightness despite lying in different illumination fields. Most of our perception of lightness arises from observations of surfaces reflec-

tions since only a few objects (bulbs, displays, sun) actually emit light. When judging the appearance of lightness we are in a certain degree influenced also by the contrast between regions and the surrounding background. A common example of contrast influence to the HVS is when two surfaces that reflect the equal amount of photons are perceived differently when surrounded by different background.

2.2 Global Contrast

Contrast perception is one of the most important aspects of vision. A significant research topic about contrast is whether the luminance that helps us to determine the surface appearance are computed globally across large distances or locally at luminance edges.

Contrast is the consequence of the luminance difference, created by the difference in the amount of the reflected light from two surfaces in their vicinities. Contrast phenomena can be generally defined as a disparity in visual properties at which the observer can distinguish an object from the background. The perception of contrast has mostly been defined based on trivial tests that evaluate the appearances of patches and various backgrounds. However, it is rather complex to measure contrast in natural images, since test images used in the study of pattern perception contains small contrast variations. Therefore, there are several slightly different ways to define the contrast, that depend mainly by the specific situations and scenarios.

Apparent contrast is a perceptual measure computed as the disparity of brightness changes, where brightness is the apparent luminance or the perceived amount of light that reaches the eye from any part of the visual field.

Global contrast can be defined simply as a ratio between the maximum and the minimum of the luminance value. Due to this property, this function can provide information about the luminance range of the scene and as well might help to identify light sources and to separate the foreground from the background. Considering that the luminance covers a wider range in the real world and also in the HDR images, a logarithmic ratio should be applied instead. The same expression is applicable for measuring the contrast of the perceived lightness:

$$\mathcal{C}_{Gen} = \frac{L_{Max}}{L_{Min}} \quad \mathcal{C}_{log} = \log_{10} \frac{L_{Max}}{L_{Min}} \quad (2.1)$$

Often employed in clinical work, the *Michelson contrast* is measured for maximum L_{Max} and minimum L_{Min} luminance. In a similar manner, the

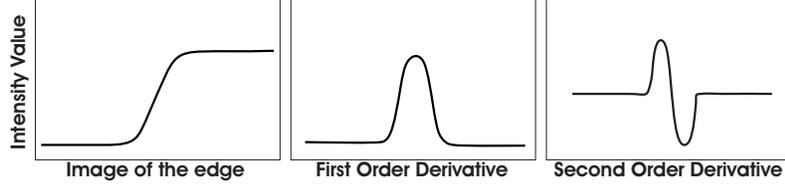


Figure 2.1: First order derivative and second order derivative.

contrast can be computed employing the mean luminance L_{Mean} :

$$\mathcal{C}_{Michelson} = \frac{L_{Max} - L_{Min}}{L_{Max} + L_{Min}} \quad \mathcal{C}_{Mean} = \frac{L_{Max} - L_{Mean}}{L_{Mean}} \quad (2.2)$$

The *root mean square (RMS)* contrast is expressed as the standard deviation of the pixel intensities from the average intensity of all pixels in the image \bar{I} . It does not account to spatial frequency content or the spatial distribution of contrast in the image (Peli 90):

$$\mathcal{C}_{RMS} = \sqrt{\frac{1}{MN} \sum_{i=0}^{N-1} \sum_{j=0}^{M-1} (I_{ij} - \bar{I})^2}, \quad (2.3)$$

where intensity I_{ij} points to the i -th j -th element from the two dimensional image of size M by N . The image I is normalized to fit the the $[0, 1]$ range.

The *Weber contrast* is measured for a background intensity I_b and for a patch that is ΔI brighter or darker than the reference background.

$$\mathcal{C}_{Weber} = \frac{\Delta I}{I_b} \quad (2.4)$$

2.3 Local Contrast

In comparison with the test images, natural images are more complex since the contrast may vary spatially over the depicted scene. Therefore, natural scene depictions can contain both areas characterized by high contrast (e.g. a forest wall, a field of flowers) and regions with low contrast (e.g. hazy/foggy regions, clear sky). The contrast expressions presented before may not capture sufficient information that accurately characterize the image, especially when the image contains both different types of contrast areas that will add their influence to the final result. Moreover, a single global measure can be

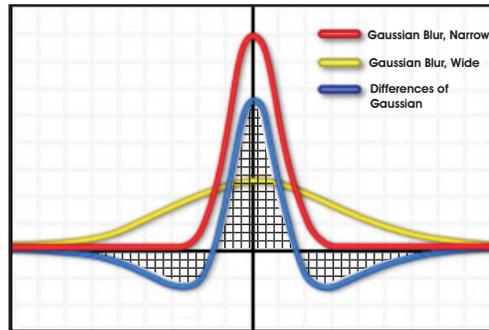


Figure 2.2: *Laplacian of Gaussian* is approximated by the difference between the wider and the narrower Gaussian.

influenced by highlights and shadows and thus, images perceived with low contrast may register high contrast rates. For local contrast depiction, the same presented measures could be applied locally to each pixel and its surrounding. Physiology studies have connected local contrast perception with the local gradient (Valois 90), (Magnussen 75). Indeed, local contrast measures a discontinuity in intensity, and the peak of the change is marking edges. Local contrast is therefore characterized by the magnitude and the orientation of edges. Practically, the rate of change of intensity or the intensity gradient measures the local contrast. In computer vision and image processing the procedure of detecting gradient variations is useful in for feature detection. In general, detecting sharp contrast variations has large applicability, since discontinuities in image brightness may be related with the depth discontinuities, scene illumination variation, changes in material properties, etc.

There are many ways to determine the contrast transitions. Probably, the most straightforward way is to compute the first order derivative (see Figure 2.1). Typically convolution kernels such as *Prewitt*, *Roberts* and *Sobel* (Rafael 08) are employed to approximate a 2D spatial gradient on an image. This operation finds the approximate absolute gradient magnitude at each point in an given grayscale image.

Since in most of the images the contrast transition is more gradual than abrupt (e.g. ramp-like transition), it is best characterized by a contour detection, where the magnitude first rises, peaks and then falls. The second order derivative yields positive magnitude on the negative side, negative magnitude on the bright side and peaks where the second derivative is zero (see Figure 2.1).

A formulation of the second order derivative may be expressed by using

isotropic filters such as the simple Laplacian (Rosenfeld 82). Since derivatives of any order are linear operations, the discrete formulation can be implemented using filter masks (Rafael 08). Given the fact that the Laplacian is a second-order derivative, it has the property to enhance the details but also to enhance the undesired noise of the image. An alternative approach is the *Marr-Hildreth edge detection* (Marr 80) method that convolves the image with the Laplacian of the Gaussian (LoG) function 2.6, or with Difference of Gaussians (DoG) 2.5 (DoG represents the fast approximation of the LoG).

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.5)$$

$$LoG(x, y, \sigma) = \frac{1}{\pi\sigma^4} \left(1 - \frac{x^2 + y^2}{2\pi\sigma^2}\right) \exp^{-\frac{x^2+y^2}{2\sigma^2}} \quad (2.6)$$

2.4 Multiscale Local Contrast

The Laplacian pyramid has been often described as a data structure that captures the spatially varying contrast that is well-suited for scaled-image analysis. The utility of this representation has been exploited for instance to isolate important components of the image pattern for analysis and direct transformations. Both Laplacian of Gaussian (LoG) and Difference of Gaussian (DoG) approaches have been employed for pyramid decomposition. This image representation is a close simulation of the local spatial processing and the band-limiting nature of human vision (Palmer 99) which involves that the image is split into several bandpass images and one lowpass image.

The well-known pyramid decomposition of Burt and Adelson (Burt 83) splits the image into several details levels employing the LoG operator. Bandpass information is obtained by extracting Gaussian (lowpass) pyramid level from the next lower level in the pyramid. A significant quality of the Laplacian pyramid is that it contains a complete image representation. To reconstruct the image the steps can be reversed. The pyramid composition involves convolution of the original image with a set of Gaussian-like weighting functions. The image I is decomposed into n variants b_i each representing the local contrast for the level i and the low pass image l .

$$I = \sum_{i=1}^n b_i + l \quad (2.7)$$

We employed as well a multi-scale Laplacian decomposition in our fusion-based techniques described in chapter 5.

In the work of Peli (Peli 00) is presented an alternative decomposition where the contrast is computed separately for each frequency band. The contrast at each point in the image is the ratio between the intensity at a pixel and the corresponding local luminance mean image:

$$\mathcal{C}(x, y) = \frac{b(x, y)}{\bar{i}(x, y)} \quad (2.8)$$

where $b(x, y)$ represents the information of the bandpass-filtered image for a certain decomposition level and $\bar{i}(x, y)$ (with $\bar{i}(x, y) > 0$) is the arithmetic local mean luminance image containing all frequencies below the band i .

In situations when is needed more accuracy, an alternative approach is to measure the contrast in the frequency domain instead of spatial domain by representing the image into the Fourier space. For HDR images Mantiuk et al. (Mantiuk 06) employ the perceptual *Just Noticeable Difference* that is obtained by transforming contrast magnitudes through a transducer function that mimics their threshold (super-threshold) behavior.

2.5 Helmholtz-Kohlrausch Effect

The *Helmholtz-Kohlrausch* effect is an entoptic phenomenon in which a higher saturation of hues spectral influences the brightness being perceived as part of the color's luminance. Therefore, *chromatic luminance* is induced proportionally with the saturation, while *white* represents the standard of comparison. The *Helmholtz-Kohlrausch* effect is defined as: *A chromatic stimulus with the same luminance as a white reference stimulus will appear brighter than the reference* (Nayatani 97; Nayatani 98). The *Helmholtz-Kohlrausch* effect inspired our grayscale operator described in chapter 3. We also employed *Helmholtz-Kohlrausch* as an additional input in our fusion-based decolorization scheme presented in chapter 5.

2.6 Contrast Sensitivity Measurement and Spatial Frequency

Since contrast perception plays a key role in vision, much effort has been done to define and measure the contrast sensitivity of the HVS. Contrast sensitivity refers to the visual ability to see subtle changes in stimulus contrast. For instance, contrast sensitivity let us to see a shade of gray on a white background or the facial features of a person standing in a poorly lit room.

Which are the conditions that must be respected in order to see the difference? A generally feasible formulation is the Weber equation that can be used in many areas as a baseline to compare performance and as a rule-of-thumb. During his experiments, E. H. Weber (1834) (Palmer 99) has gradually increased the weight hold by a blindfolded man while he asked him to indicate when he first felt the weight adjustment. Weber found that the smallest noticeable difference in weight was proportional in ratio with the initial weight. The law is suitable also for the auditory sensory, like for example in a noisy environment a person must speak loudly to be heard while a whisper is distinguished in a quiet room. The Weber's law expresses that the ratio of increment threshold to the reference is a constant. For visual contrast sensitivity, Weber law's indicates that the contrast is constant in proportion to mean local luminance in the bright range, then decreases as illumination descends to dark.

In assessing the *Contrast Sensitivity Function*(CSF) the common test is using sinusoid grating functions (see Figure 2.3) that help to examine the perceptual response to simple stimuli at a broad range of spatial frequencies and increasing contrast magnitudes. Typically, the best scores are recorded at low-frequencies, while at high-frequencies the visibility decreases since the grating (stripes) becomes very thin.

Moreover, the assessment of contrast sensitivity can be performed in a similar way as the vision standard test that uses a letter chart (the Snellen standard tests of visual acuity evaluates the grade we can see objects of different sizes at high contrast). The contrast sensitivity Pelli-Robson (Pelli 02) chart can verify the ability to detect objects of different sizes and decreasing contrast. The better sensitivity is obtained for medium-sized objects when their contrast is low, while the smallest objects can be detected only when their contrast is very high.

2.6.1 Image Quality Assessment (IQA) Metric

In general *image quality metrics* estimate the magnitude of visible distortions of different degraded versions relative to the ideal reference image. Even if the literature abounds of different strategies, there is no specialized metrics to evaluate the quality of grayscale transformations and dehazing strategies that we are going to discuss and validate in this dissertation. While most of the existing metrics (Lubin 95; Wang 02; Wang 04) are focusing mainly on the structural changes we believe that the metric of Aydin et al. (Aydin 08) would be more appropriate for our evaluations. This metric (with an online implementation¹) is built on a model of the human visual system (HVS), as

¹<http://drim.mpi-sb.mpg.de/>

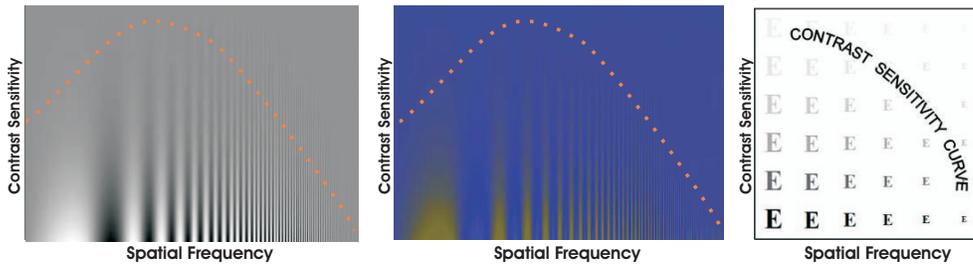


Figure 2.3: *Contrast sensitivity Function(CFS) diagram.* Notice that the contrast sensitivity is reduced at high frequencies. Better sensitivity is registered for lower-frequencies or when the grating is wider.

it effectively blends the contrast detection with the structural similarity measurement. Instead of detecting only the contrast changes, the metric of Aydin et al. (Aydin 08) is sensitive to three types of structural changes: (a)- *loss of visible contrast (green)* - a contrast that was visible in the reference image becomes invisible in the transformed version, (b)- *amplification of invisible contrast (blue)* - a contrast that was invisible in the reference image becomes visible in the transformed version and (c)- *reversal of visible contrast (red)* - a contrast is visible in both images, but has different polarity. As a general interpretation contrast loss (green) has been related with blurring, while contrast amplification (blue) and reversal (red) have been connected with sharpening (see Figure 2.4 for the response of this metric for a simple example where the contrast is increased gradually).

As will be presented in the next chapters, we employed this metric to evaluate the contrast changes for decolorizing techniques and as well for dehazing methods.

2.7 Visual Saliency

Visual saliency is a broadly defined concept that indicates that some of the scene parts pre-attentively stands-out more than other and produce within the early stages of HVS a some form of immediate visual arousal. That means, that when we briefly look at images, some structures will attract our immediate attention even without scanning the entire image.

The concept of saliency is associated with the potential influence of a stimulus onto our perception. Saliency detection is often researched in the context of the visual system, but similar mechanisms operate in other sensory systems.

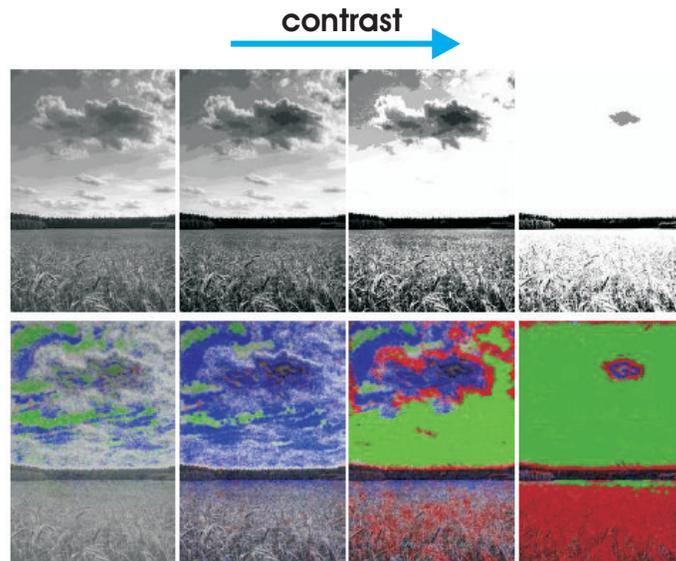


Figure 2.4: *Response of IQA metric to simple contrast stretching.* Contrast is increased from left to right, which results in more clipping and generates stronger visible contrast loss and reversal responses (images courtesy to Aydin et al. [2008]).

A number of experiments (Treisman 85) has reported that certain visual features *pop-out* or are immediately perceivable within the HVS (a review of these can be found in (Milanese 93)).

Saliency information measures the degree of conspicuousness with respect to the neighborhood regions. Itti et al. (Itti 98) have developed a biologically-inspired model of bottom-up, task-independent, saliency-based selective visual attention. They have extended the original idea of Koch and Ullman (Koch 85) of the existence, in the brain, of a specific visual map that is responsible encoding for local visual conspicuity. The model estimates a set of multi-scale neural *feature maps* such as color, intensity and orientation. Each of the interest feature map is modulated by the activity in neighboring neurons, inspired from recent neurobiological findings. To estimate the most salient feature of the map, this approach uses a winner-take-all neural network. In addition, to direct the focus of attention to the next most salient region, an inhibition-of-return mechanism (inspired from human psychophysics) is employed. This two steps guarantee that the saliency map reflects and is scanned in decreasing order.

A more effective approach of Achanta et al. (Achanta 09) has been inspired

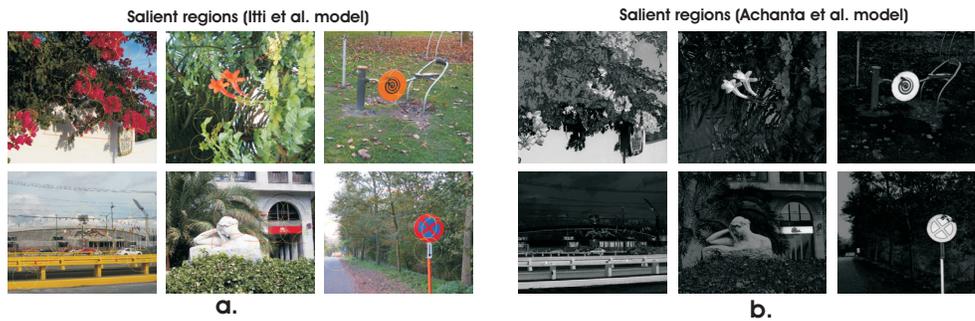


Figure 2.5: From left to right: a) salient regions detected by the algorithm of Itti et al. (Itti 98) when both color and intensity are considered. b) saliency map estimated by the effective method of Achanta et al. (Achanta 09)

by the biological concept of center-surround contrast. They have proposed a frequency tuned method that straightforwardly estimates pixel saliency using a pixel's color difference from the average image color (in the perceptual color space $CIEL^*a^*b^*$). This approach is computationally effective, however can be insufficient to analyze complex variations common in natural images.

2.8 Contrast Visual Effects

2.8.1 Contour Contrast Illusions

Detection of edges may be related with several illusions. A direct transition from low to high intensity produces the edge appearance, that is a thin connecting border between the two adjacent areas. Most of the transitions in the real world are produced gradually and seem to blend the regions. Illusion may appear in images that contain both gradual and sharp transitions.

The *Mach bands effect* (named after the physicist Ernst Mach) is perceived when two regions with uniform intensity are connected by a narrower strip with uniform gradient region. Notice that the light band 2.6 is seen to the left side and the black band is seen to the right. This effect is explained by the high-boosting filtering process that involves lateral inhibition (Palmer 99).

Sharp contrast transitions have also the effect to increase the apparent contrast. The *cornsweet illusion* also known as Craik O'Brian illusion (Ware 83; Palmer 99; Livingstone 08) demonstrates that a sharp discontinuity can influence the perception of a uniform patch. In comparison with the *Mach bands* where the effect appears only on areas that are close to the gradient,

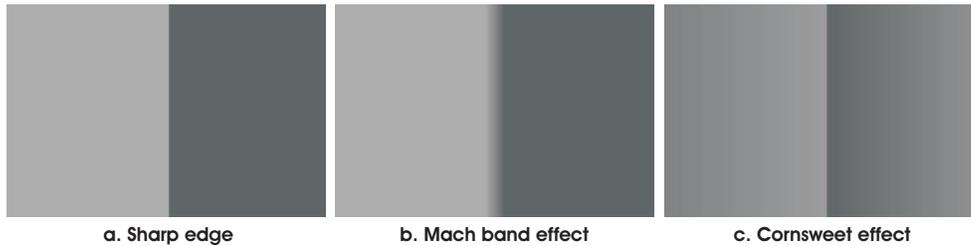


Figure 2.6: *The left image contains two uniform regions and a narrower strip (with uniform lightness gradient) in between. In the Mach band illusion two bands, a white one and a black one appear immediately to the left and to the right of the middle strip. In mornsweet illusion is noticeable that the right part of the image seems to be darker than the left one even though both have exactly the same gray level.*

the *cornsweet illusion* seems to influence the perception of the entire areas 2.6. Notice that the adjacent region to the light part appears lighter and the region adjacent to the dark part appears darker.

2.8.2 Simultaneous Contrast Illusion

Introduced by Michel Eugne Chevreul, *simultaneous contrast* involves both achromatic perception and color perception. *Simultaneous contrast* may be demonstrated by displaying regions with identical spectra surrounded by different backgrounds. The central regions (notice Figure 2.7) have different perceived lightness, depending on their backgrounds. The patch within the black square appears lightest than the patch within the white square. If the background is cut out and the patches are examined against uniform background, it can be seen that they all have the same shade of gray. The simultaneous color contrast is valid as well for combining color patches against different color backgrounds. Notice in Figure 2.7 (b,c) that when the color patch is surrounded by a strongly colored background, the perception of the center patch is visibly affected by the background color. The perceived color is shifted toward the complementary hue to a degree that is proportional with the saturation and the viewing duration. This effect is due to the chromatic adaptation process that is depicted in the next part 2.9.

Some isoluminant color combinations (e.g. those presented in Figure 2.8) cause illusions when placed together. The theoretic explanation is that the human visual system treats separately luminance and color. The information

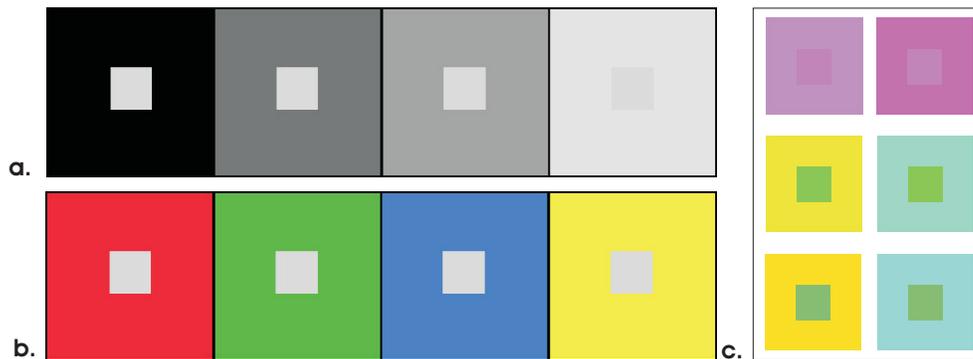


Figure 2.7: Simultaneous contrast produced in achromatic (a) images. The uniform patch appears darkest against white background, lightest against a black background and intermediate lightnesses against intermediate backgrounds. Similarly, this effect is also noticeable for color images. Notice in (b) and (c) the color shift produced by the simultaneous hue contrast.

of *WHERE* (Livingstone 08) such as shapes, lighting or depth is obtained from luminance, while the information of *WHAT* (Livingstone 08) like texture and material is extracted from color. The final scene interpretation is obtained by combining at the cognitive level this information.

2.9 Color Perception

Chromatic color constancy is the mechanism of the visual system that contributes to the perception of invariant properties of a surface's spectral reflectance for various changes in illumination or viewing conditions. Our perception about objects that surround us seems unaltered by the illumination changes (as examples can be considered morning versus afternoon light and fluorescent versus tungsten). The physical phenomena is explained by the fact that the changes in spectral of different illumination sources can cause the modifications in the perception of the surfaces. The observed color of different regions in a chromatic image is in theory a mathematical multiplication between the surface reflectance spectrum and the light spectrum. It is not fully understood yet, how can our visual system separates the wavelengths in the reflectance from the wavelengths in the illuminating light. In theory this is an inverse problem with no unique solution, since there are six unknown parameters for each surface (three from the reflectance and three from the il-

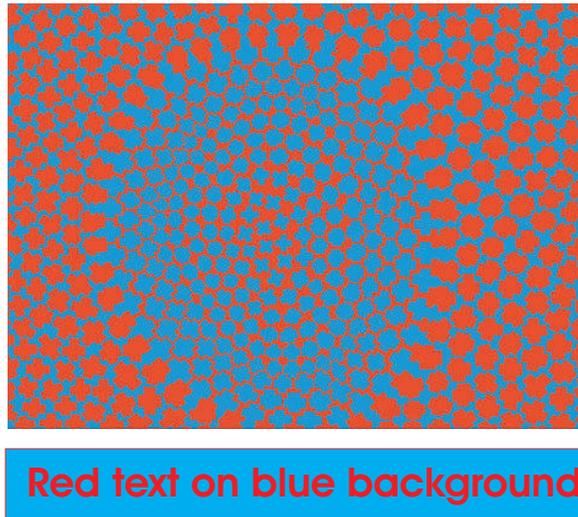


Figure 2.8: Some color combinations, such as the isoluminant blue and red in this synthetic image or the red text on a blue background cause illusions - shapes may appear jittery, text may vibrate and cast a shadow.

luminating light) and only three observation parameters (the information from the cones output). Thus, it is remarkable the fact that our HVS is capable to achieve, apparently with no effort, color constancy in normal conditions.

Chromatic adaptation (Palmer 99) plays a key role in color constancy. Neurophysiological basis assumed that when a component, for example green is introduced by blue-tinted glasses, the M cones (Palmer 99) will fire less due to the fatigue induced by the constant presence. This adaptation mechanism discharges the additional hue-shifting effect. This process of adaptation is not enough to explain the entire effect. It is believed that an observer who faces these conditions can make several assumptions to get the desirable solution. Identifying these assumptions, a study case can be modeled properly and correct solutions can be achieved. In general three assumptions are made: *1. consistency of illumination* - this assumption regards the spectrum properties of the light source that can be considered to be similar for the entire image, with small variation of the intensity due to the distance from light, surface orientation or shadows; *2. restricted range of light sources*- in most of the cases and especially in natural images, a person will meet only a small range of possible light sources. We rarely meet light sources that differ considerably chromatically. This assumption constrains the theoretically full range of possible color constancy to a significant reduced searching space (Judd 64); *3. restricted reflectances* - in natural image only a small amount of possible

surface reflectances are found and even more, these tend to be smooth rather than sharp (Cohen 64).

These constraints have been incorporated in the color constancy approach of Maloney and Wandell (Maloney 86). The color constancy phenomena and the evolutionary adaptation of the visual systems have been considered by Shepard (Shepard 92) to be the explanation for the tridimensional basis of color vision. His main idea states that *variations in daylight have three degrees of freedom*. This can be interpreted by the fact that any spectral variation can be approximated by adjusting three of the considered parameters: light/dark level controls the quantity of considered light, that can vary on the entire possible range from the deep shade of moonlight to the midday sunlight; red/green balance supervises the amount of long (red) versus other wavelengths with spectral centered on green; blue/yellow level controls similarly the amount of short (blue) versus other wavelengths with spectral centered on yellow.

Shepard's theory (Shepard 92) can be related up to a certain point with the Herring's color opponency (discussed in section 2.10). A significant argument about the evolutionary role of color vision is that it enables the approximately constant perception of the surfaces under various illumination conditions. To accomplish this task the theory proposed a compensatory process for the changes in the light spectrum by adapting the black/white, red/green and blue/yellow balance.

An important question is whether color constancy is an inborn ability. The answer may have important influence to fields such as understanding the color perception or visual particularities like color blindness. Although there are a few experiments on infants, they all agree so far that both chromatic and achromatic color constancy are absent at birth and are actually learned in the first few months of life. Dannemiller experiments(1989) (Dannemiller 89) has demonstrated by involving infants of two and four months of age using the adaptation paradigm, that four months old infants have developed chromatic color constancy while two months old infants do not.

2.10 Theories of color vision

Trichromatic Theory or **Young-Helmholtz three component theory** states that there are three types of color receptors in the human eye. These elements that responds differently to the wavelength of the photons that falls on them, were considered to produce the main psychologically primary color sensation of red, green and blue. The trichromatic theory was proposed several times by Mikhail Lomonosov in 1757, George Palmer in 1777 and by the physicist Thomas Young in 1802, but since all have lacked the essential opti-



Figure 2.9: *Hue cancellation experiment*. Notice that the intermediate color can not be described as yellowish-blue or reddish-green.

cal tools the hypothesis was not demonstrated at that moment. The theory has been demonstrated by the German physicist and physiologist Hermann von Helmholtz (1821-1894) (*On the Theory of Compound Colors (1850)* and *Handbook of Physiological Optics (1856)*). He stated that just three types of receptors or *nervous fibers* and the overlapping sensitivity functions are adequate to cause the color sensation.

Color Opponency Theory introduced by Ewald Hering (1878) is based on the analysis of sensation of color. This theory assumes that there are six independent unitary colors, although the color identification is not separated in six independent systems, but is based on three counterbalancing processes: black-white, yellow-blue, and red-green. Hering was intrigued by the impossibility of experimenting or even imagining any colors that can be described as yellowish-blue or reddish-green 2.9 and by the apparent mutual canceling of blue and yellow or of red and green when they are combined in the right proportions (Wyszecki 00; Hubel 88). Work on the physiology of isolated fish retinas confirmed such *double-opponent* cells. Stephen Kuffler (1950) and later Nigel Daw (1967) (Conway 02) have shown that the receptive field of such cells were organized such as center-surrounding that were activated by small spots of a different set of wavelengths and suppressed by light of a different set of wavelengths. Similar spatial and chromatic opponencies have been demonstrated to exist in the human primary visual cortex (Engel 97).

The theory was considered only in the last few decades, notably by psychophysicists such as Leo Hurvich and Dorothea Jameson, Deane Judd, and Edwin Land (Hubel 88). Edwin Land color constancy demonstration on the *Mondrian* like painting in the 1950s and its Retinex (Land 71) algorithm have demonstrated the importance of this concept.

Dual Process Theory (Wyszecki 00) or *Stage Theory* solves the controversy about which of the two previous theories is correct or not. The theory states

that both of the previous concepts correspond to the two-stage theory of color vision. The first stage, that is the receptor stage involves three photopigments (blue, green and red cones). This is the stage that validates the trichromatic theory. The second stage is the post-receptor level and occurs as early as the horizontal cell validates the opponency theory.

Lateral Inhibition is a pervasive architecture in our visual system that is responsible for the color processing. It is considered to be responsible for many contrast related effects such as *Mach bands* and *simultaneous contrast* (Ratlif 65). This neural organization occurs in a two-layers network when each neuron in the first layer excites a corresponding neuron in the second layer and in the same time inhibits the second-layer neurons lateral to the excitatory connection. The output for each neuron can be obtained by multiplying all its connection weights with their associated inputs and summing all these products over the input. Notice the Figure 2.10 in which the output of the second layer shows a *Mach band* effect.

Since the later-inhibition is only one-dimensional, a more realistic model for retinal processing it is necessary. An accepted solution is to implement lateral inhibition in two-dimensional structure. That involves rotating the one-dimensional section through a second dimension. The result is that the given receptor spreads out in all directions from the excitatory center. This two-dimensional structure is called *center-surround* organization. Therefore, the cells in the second layer respond strongly to a bright spot that activate the excitatory center and a dark ring surrounding the spot. By this organization the uniform light will stimulate homogeneously the entire receptive field and will reduce the response from the second-layer (Palmer 99; Livingstone 08).

Double Opponent Cells is generally thought responsible for the simultaneous color contrast effects. Identified in the visual cortex (Michael 78) this presents a similar spatially opponent center-surround organization like the one that considers the brightness. The main difference is that double opponent cells have a chromatically opponent structure. That means that the center and the surround have color-opponent coding. Therefore, a red/green double opponent cell can be either (R^+G^-), excitatory to red and inhibitory to green or (R^-G^+) that is excitatory to green and inhibitory to red. A blue/yellow double opponent cell can be either (B^+Y^-) that is excitatory in the center to blue and inhibitory to yellow in the surround or vice versa.

This organization is also considered to play a key role in *chromatic adaptation*. The process is much like the *light adaptation*. When a person walks from a dark room into the sunlight it is experienced the blinding brightness. This is due to the fact that the long dark period adaptation has activated most of the photosensitive pigment molecules in both rods and cones into their maxi-

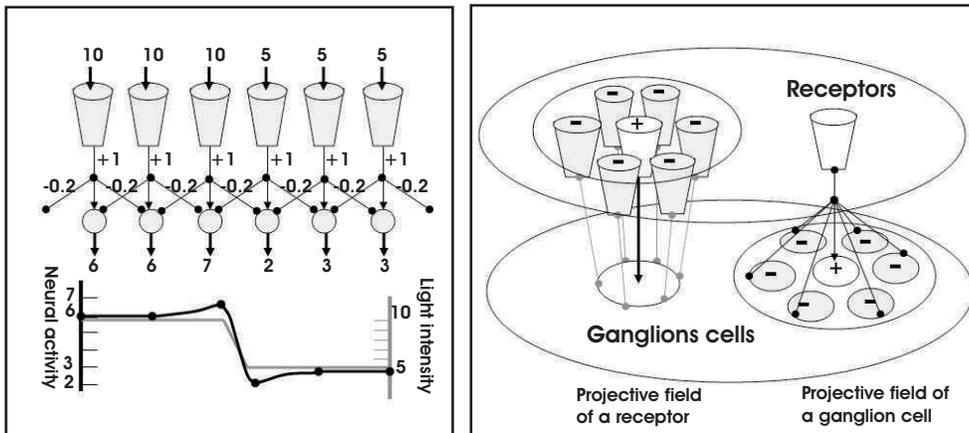


Figure 2.10: *Lateral Inhibition*. In the left part is shown one-dimensional representation of lateral inhibition mechanism. In the right part is presented the two-dimensional structure also known as *center-surround* organization.

mally sensitive state. Therefore, their response is much stronger than usually. Continuing the exposure to sunlight triggers *light adaptation* by which a large number of photo-pigment molecules are bleached and put into a less sensitive state that reduces their response to light. In case of *chromatic adaptation* the chromatic afterimages are the residual effect. Changing the view from a highly saturated color that has been fixed in the gaze for a relative long time, will produce the perception of the opposite hue in the corresponding retinal region as a response to a neutral (white) text field (Palmer 99; Livingstone 08).

2.11 Color Spaces

This section presents a brief description of color spaces models. For the sake of completeness more exact and more complex description of these models can be found in the specialized work of (Fairchild 05; Wyszecki 00; Hunt 98; Col 86).

2.11.1 RGB/CMYK Color Spaces

The additive *RGB* color model is built on three additive primary colors: red, green and blue. By mixing two primary colors can be obtained a secondary color of light, while mixing all three primaries with the right intensity can be produced white light. Most of our devices such as digital cameras, image scanners, displays although with different technologies use the same fundamental

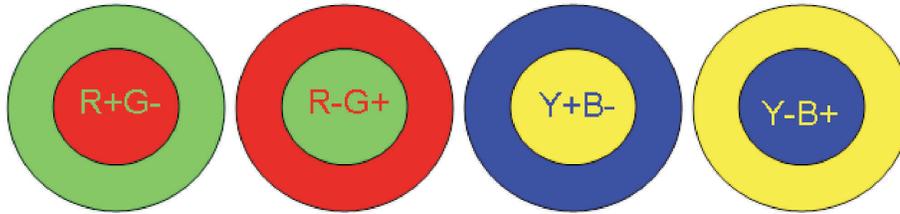


Figure 2.11: *Opponent cells organization.* R^+G^- is excitatory to red and inhibitory to green, R^-G^+ is excitatory to green and inhibitory to red, B^+Y^- is excitatory in the center to blue and inhibitory to yellow in the surround and vice versa for B^-Y^+ .

RGB color model principle. Since different devices have various responses levels of R , G , and B it implicates that this color model is device dependent. Not less important, mixing all the primary colors can be reproduced a broad range of colors, however since the primaries can variate only the intensities and not also the wavelength, it is not possible to reproduce all visible colors. In consequence, the transfer across devices implicates a certain standard of color management. In the subtractive *CMY* (cyan, magenta, yellow)/*CMYK* (cyan, magenta, yellow, black) model, the color is determined by the absorption of light. The subtractive principle that is used for color printing, implicates that black is obtained by combining the cyan, magenta and yellow colors. The primary color of the subtractive model is the one that absorbs the primary color of light, and for example when printing ink on paper the impurities in the pigment produce a subtractive operation between colors.

With only a few exceptions, most of the displays employ the three primaries (R , G and B). This implies that most of the video standards considers color primaries in terms of the chromaticities of R , G and B and the white point (represented by an achromatic color with the highest luminance). Digital control of the displays is practically unintuitive, for most common users, to be performed by simply employing the *RGB* color model. Typically, a color is obtained by combining the three R , G , B values in a certain amount. However, by simply modifying the value of one channel, this operation does not match the way of humans perceive and expect to model the color appearance. Many times the color processing implicates first transforming the *RGB* image to a new more perceptual color space, and converting it back to *RGB* once the

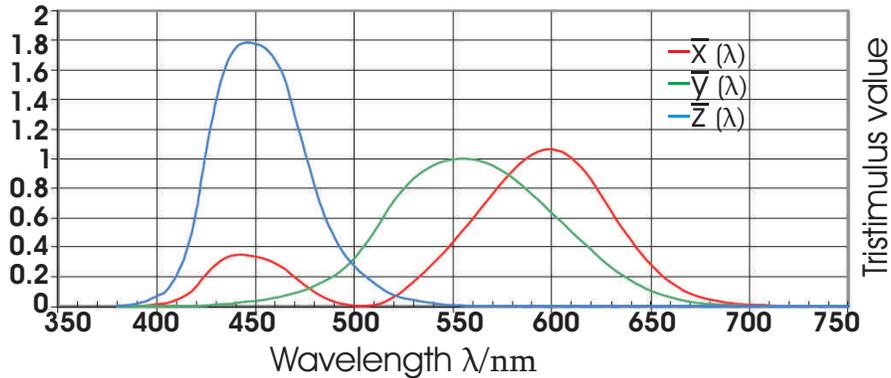


Figure 2.12: *The CIE standard observer color matching functions*

adjustments are complete.

2.11.2 Perceptual Color Spaces

Our tristimulus vision system gathers information from about 6-7 millions of cones. The cones are found in three different categories that present sensitivity to red(65%), green(33%) and blue(2%).

Based on the physiological perception of light, in 1931 the Commission Internationale de l'Éclairage (*CIE*) have introduced one of the first mathematically defined spaces. The *CIEXYZ* color space (Guild 31b) has the foundation on a set of three color-matching functions called *Standard Observer* functions. These *CIE* standard observer color matching functions depicted in Figure 2.12 have been experimented with independently on human sight by W. David Wright (Wright 28) and John Guild (Guild 31a). During the experiments subjects have been shown patches of color and have been requested to adjust the output of three primary colors (blue=435.8nm, green=546.1nm, and red=700nm) in order to obtain an equivalent sensation with the test color. The three absorption sensitivity curves are overlapped for these three varieties mostly between red and green cones.

The *CIEXYZ* color space contains the human gamut and since it was designed to be device independent, it represents the basis for most of the color management systems. The tristimulus values X, Y, Z for a color with a spectral power distribution $I(\lambda)$ is obtained by employing the color-matching functions 2.12, called $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$:

$$X = \int_0^{\infty} I(\lambda) \bar{x}(\lambda) d\lambda \quad (2.9)$$

$$Y = \int_0^{\infty} I(\lambda) \bar{y}(\lambda) d\lambda \quad (2.10)$$

$$Z = \int_0^{\infty} I(\lambda) \bar{z}(\lambda) d\lambda \quad (2.11)$$

where λ represents the wavelength of the equivalent monochromatic light (measured in nanometers).

Soon, after the *CIE* diagram has been presented, it has been observed that *CIEXYZ* color space does not cover uniformly the color gradation. Measurements have disclosed that distances between two distinguishable colors are unequal over the entire space and the trace of the minimal discernible colors around a point form an ellipse, called the *MacAdam ellipse*. Another drawback that has been noticed for *CIE1931* was a consequence of the *Field of View (FOV)* for the test patches. The initial measurements have been done by using small patches and the angle subtended by the eye's fovea. Since the cones density falls rapidly from the center of fovea, only color patches that fall in the *FOV* under four degrees are validated by the measurements of *CIE1931*. For larger patches it has been introduced a new set of measurements called the *CIE1964*. However based on the assumption that most displays and print materials present combinations of small color patches the *CIE1931* is still maintained as a reference for analysis. Additional several transformations have been proposed (in 1948 the *CIELab* and in 1960 the *CIEUCS* (Uniform Chromatic Space)) aiming to obtain a uniform chromatic space.

In 1976 two more intuitive 3D color spaces derived from the tristimulus *CIEXYZ* have been introduced. The *CIEL*a*b** (1976) and the *CIEL*u*v** present the advantage of being better perceptual uniform considering that the differences in the color space are more correlated with the perceived difference between colors. Besides being more intuitive, the color spaces of both *CIEL*a*b** and *CIEL*u*v** estimate the perceived lightness L^* , chroma C^* and hue H^* (where C^* and H^* are computed from a^* , b^* respectively u^* and v^*).

As defined by Fairchild (Fairchild 05) **lightness** is the brightness of a region judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting.

The lightness is computed from the luminance Y and the reference white of the luminance Y_n :

$$L^* = 116 f(Y/Y_n) - 16 \quad (2.12)$$

In both color spaces L^* is the lightness axis, and the two channels are representing the decorrelated chromatic channels. In *CIEL*a*b** the two axes represent the red-green and yellow-blue opponent color pairs. The *CIEL*a*b**

color axes are obtained as a derivation from *CIEXYZ* using the reference white:

$$a^* = 500 [f(X/X_n) - f(Y/Y_n)] \quad (2.13)$$

$$b^* = 200 [f(Y/Y_n) - f(Z/Z_n)] \quad (2.14)$$

$$f(t) = \begin{cases} t^{1/3} & t > (6/29)^3 \\ \frac{1}{3} \left(\frac{841}{108}\right) t + \frac{4}{29} & \text{otherwise} \end{cases} \quad (2.15)$$

where X_n , Y_n and Z_n represent the normalized X, Y and Z tristimulus values for a specified reference white point. For the common uses, the standard *CIE* illuminant *D65* is chosen because it roughly corresponds to the daylight average. The exponent $1/3$ of function $f(t)$ corresponds to non-linear perception of the eye to the increased luminance (e.g. dynamic range compression). The function $f(t)$ is assumed linear below value $t = t_0$ and match the $t^{1/3}$ in value and in slope. To match in value : $t_0^{1/3} = at_0 + b$ and in slope $\frac{1}{(3t_0^{2/3})} = a$. For a chosen value of $b = 16/116$, the two equations can be solved for $\delta = 629$: $a = \frac{1}{(3\delta^2)} = 7.787037 \dots$ and $t_0 = \delta^3 = 0.008856 \dots$. The slope at the joint is $b = \frac{16}{116} = \frac{2 \cdot \delta}{3}$.

The derivation of the *CIEL* u^*v^** color axes from *CIEXYZ* is obtained by applying a subtractive shift of a *specified white object* (Col 86) reference:

$$L^* = \begin{cases} \frac{29^3}{3} \frac{Y}{Y_n} & \frac{Y}{Y_n} > \frac{6^3}{29} \\ 116 \frac{Y}{Y_n}^{1/3} - 16 & \text{otherwise} \end{cases} \quad (2.16)$$

$$u^* = 13L^* \cdot (u' - u'_n) \quad (2.17)$$

$$v^* = 13L^* \cdot (v' - v'_n) \quad (2.18)$$

$$u' = \frac{4X}{X + 15Y + 3Z} \quad (2.19)$$

$$v' = \frac{9Y}{X + 15Y + 3Z} \quad (2.20)$$

It has been shown that is more intuitively to employ the cylindrical version of *CIEL* C^*H_{ab}* and *CIEL* C^*H_{uv}* . The main reason is that the chromatic dimensions are H^* that represents the hue angle over the color disk and C^* that represents the perceived colorfulness.

$$C_{ab}^* = \sqrt{(a^*)^2 + (b^*)^2} \quad C_{uv}^* = \sqrt{(u^*)^2 + (v^*)^2} \quad (2.21)$$

$$H_{ab} = \tan^{-1} \frac{b^*}{a^*} \quad H_{uv} = \tan^{-1} \frac{v^*}{u^*} \quad (2.22)$$

Although there are several modes to define saturation s , the most common way to define it is the perceived colorfulness of a stimulus with respect to its lightness.

$$s_{ab} = \frac{C^*}{L^*} = \frac{\sqrt{(a^*)^2 + (b^*)^2}}{L^*} \quad s_{uv} = \frac{C^*}{L^*} = 13\sqrt{(u' - u'_n)^2 + (v' - v'_n)^2} \quad (2.23)$$

The chromatic contrast or the color difference is in general computed by employing the Euclidean distance of the (L^*, u^*, v^*) or the (L^*, a^*, b^*) 2.24.

$$\Delta E_{a^*b^*} = \sqrt{\Delta(a^*)^2 + \Delta(b^*)^2 + \Delta(L^*)^2} \quad (2.24)$$

However, since the color difference is illuminant dependent more complex measures such as the one described in (Fairchild 05) can be employed as well.

2.11.3 HSL, HSV, HSI Color Spaces

As already discussed, since *RGB* color space does not correspond to human perception of color more perceptual and intuitive alternative color spaces have been proposed in literature (Smith 78). Compared with *CIE* color spaces, the *HSV* and *HSL* color spaces are straightforward transformations of the non-perceptually-based *RGB* model. In comparison with the related models as defined by *CIE*, these models are not strictly connected by the photometric color-making attributes. Nevertheless, they are widely accepted and involved in image editing and graphic software. The hexagonal *HSL/HSV* models are obtained by projecting the *RGB* color cube along its main diagonal (the gray axis) onto a plane perpendicular to the diagonal. Colors seen along the gray axis on the cube become the interior points of the transformed color space.

The definition for value/brightness, lightness and intensity although at the first glance seems to represent the same thing, they are all different. The value/brightness is obtain from the maximum of the R, G, B component, lightness is the midpoint value between the maximum and the minimum of the R, G, B values and finally, the intensity is the average of the R, G, B value:

$$I = \frac{R + G + B}{3} \quad (2.25)$$

$$V = \max(R, G, B) \quad (2.26)$$

$$L = \frac{(\max(R, G, B) + \min(R, G, B))}{2} \quad (2.27)$$

In the hexagonal projection, hue represents the rate distance around the edge of the hexagon which passes through the projected point. For image analysis the hue is usually defined by the cartesian-to-polar coordinate transformations that is expressed as:

$$H = \text{atan2} \frac{\frac{1}{2}(2R - G - B)}{\sqrt{(R - G)^2 - (R - B)(G - B)}} \quad (2.28)$$

Finally, the saturation attribute is defined slightly different in these color spaces:

$$\begin{aligned} S_{HSV} &= \begin{cases} 0, & \text{if } \max(R, G, B) - \min(R, G, B) = 0 \\ \frac{\max(R, G, B) - \min(R, G, B)}{V}, & \text{otherwise} \end{cases} \\ S_{HSL} &= \begin{cases} 0, & \text{if } \max(R, G, B) - \min(R, G, B) = 0 \\ \frac{\max(R, G, B) - \min(R, G, B)}{2L}, & \text{if } L \leq \frac{1}{2} \\ \frac{\max(R, G, B) - \min(R, G, B)}{2 - 2L}, & \text{if } L > \frac{1}{2} \end{cases} \\ S_{HSI} &= \begin{cases} 0, & \text{if } \max(R, G, B) - \min(R, G, B) = 0 \\ 1 - \frac{\min(R, G, B)}{I}, & \text{otherwise} \end{cases} \end{aligned} \quad (2.29)$$

Chapter 3

Perceptual Decolorization of Images and Videos

Colors are only symbols; reality is to be found in lightness alone

Pablo Picasso

Contents

3.1	Introduction	38
3.2	Related Work	41
3.3	Saliency-guided Decolorization	43
3.3.1	Overview	43
3.3.2	Our Decolorization Approach	43
3.3.3	Offset Angle Selection	47
3.3.4	Results and Discussion	55
3.3.5	Validation	60
3.4	Video Decolorization	61
3.5	Other Applications	62
3.5.1	Detail Enhancing	63
3.5.2	Segmentation under Different Illuminants	66
3.5.3	Matching by Local Feature Points	66
3.5.4	Decolorizing Images for Auditory Substitution Systems	68
3.6	Conclusions	72

This chapter takes a more in-depth look at the perception of *apparent brightness*. We discuss the influence of the color contrast on it, mainly of the *color opponency* process. Combining two opponent colors increases the apparent contrast of the image, effect that has been exploited by many artists (see chapters 1.1, 2.10). Often in photography or publishing, images that have higher contrast can have a stronger visual impact. Due to the preference for higher contrast, color-to-grayscale techniques search to increase the contrast to enhance or otherwise to improve the standard luminance of the original image. We present the limitations of the standard transformation and then we overview the work that has been done in this area. The goal of our novel approach is to restore the impression of color contrast that we relate with overall appearance and especially of the most salient features. Preserving the salient area, the conversion increases the discriminability of the scene objects ensuring that color local contrast in that areas is well maintained. In section 3.5.1, 3.5.2, 3.5.3 we explore the potential of decolorizing images for several applications such as detail enhancing, segmentation under different illuminants and image matching. Finally we discuss the potential of employing the proposed novel straightforward method for auditory substitution systems.

3.1 Introduction

Recent efforts to accurately integrate properties of human visual perception have fueled an impressive improvement not only in digital photography but also in color manipulation tools. Color is one of the main attributes that influences human visual system, being ultimately the result of our experience and perception. However, several important applications such as substitution systems for visually impaired people, the display of medical imaging, printed textbooks and aesthetical stylization still require a reliable grayscale representation of images.

Mapping three dimensional color information onto a single dimension while still preserving the original appearance, contrast and finest details is not a trivial task. Standard monochromatic transformations, found in commercial image editing software, neglect the color distribution, and as a result they are commonly unable to conserve the discriminability of the original chromatic contrast (see Figure 3.1). Consequently, isoluminant colors are mapped to similar grayscale levels which increases the ambiguity in the decolorized image version. Although luminance is a definite value, our subjective interpretation that measures *apparent* brightness is very important in image decolorization process. Mastering the craft of photography is to be able to communicate “the

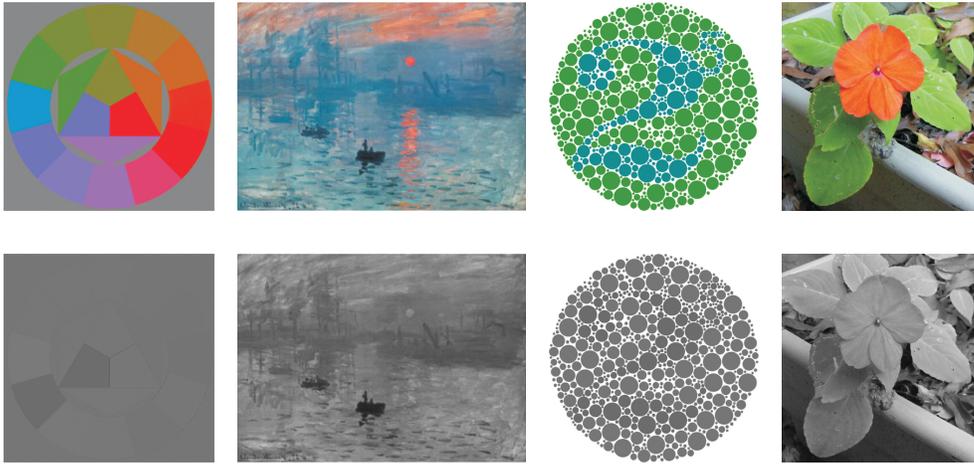
Color Images**Standard Grayscale Conversion (Luminance)**

Figure 3.1: *Limitation of the standard grayscale transformation.* Standard monochromatic transformation, found in commercial image editing software, neglects the color distribution, and as a result they are commonly unable to conserve the discriminability of the original chromatic contrast.

equivalent of what we saw and felt“ (*Alfred Stieglitz (1864-1946) - photographer*).

This problem has recently been addressed by several approaches. Apart from solving the discriminability of the compressed chromatic levels, other characteristics are also very important for a perceptually accurate and efficient conversion. In general, due to quantization strategies or prohibitive function optimization, many of the existing approaches fail to render the original image appearance, preserving the finest details and luminance consistency (shadows and highlights should not be reversed). Additionally, a crucial problem is the computational efficiency for real time applications. We argue that the concept of image decolorization is not to obtain a perfect optical match, but rather to obtain a plausible image that maintain the overall visual appearance and also the most salient features, which improves the discriminability of the scenes objects. However, due to the fact that black-and-white pictures are not able to render the entire luminance range, such images will always be to some degree an interpretation of their original.

In section 3.3 we introduce an alternative decolorization approach. In our scheme, the luminance level is progressively augmented by the chromatic vari-

ation of the salient information. Saliency, a well studied topic in computer vision, aims to emulate the attention mechanism of human visual system. We have based our approach on the assumption that preserving these salient regions in the converted image will result in a better preservation of the visual contrast and overall perceptual appearance. Guided by the prominent chromatic information, our technique introduces a mapping based on the *color opponency theory* (Judd 66), generally accepted as the foundation of many visual saliency models (Itti 98). After the monochromatic luminance channel is filtered and stored as a reference, the luminance values are computed pixel-wise by mixing both saturation and hue values, creating a new spatial distribution with an increased contrast of the interest regions. All the pre-computed values are normalized in order to fit the entire intensity range. The intensity is re-balanced in order to conserve the amount of glare in the initial image. For extreme lighting conditions, we apply several constraints in order to avoid clipping and fading of the apparent details.

The method has been verified on a large database of color images, including the standard examples commonly used to test grayscale techniques. In order to validate our technique we have performed a comprehensive comparison to the existing state-of-the-art methods, including a perceptual evaluation. Our approach, although conceptually straightforward, is able to decolorize images competitively with the more elaborate techniques. In contrast, the runtime of our algorithm is linear, which means it becomes suitable for real-time applications. A small set of parameters allow the users to control the grayscale conversion. The method performs well with default parameters.

Since our decolorization is accurate and preserves finest details, we can exploit variations in chromacity as well as luminance to enhance color images simply, blending the processed luminance levels with the initial color. Observed also by Gooch et al. (Gooch 05), the human visual system has better spatial acuity in luminance than in chrominance; by fusing the chrominance with co-located luminance changes, weak chroma variations are rendered sharper and better-localized. In the last part of this chapter are demonstrated the utility of our operator for tasks like segmentation under different illuminants, detail enhancement and wide-baseline image matching. Furthermore, we show that the approach can also be applied robustly to video decolorization, which is more challenging due to the temporal coherence constraints that may introduce flickering artifacts.

3.2 Related Work

Traditional conversions disregard the chrominance information and consider the luminance level of different color spaces (e.g. $CIEL^*a^*b^*$, $YCbCr$, HSL/HSV) (Wyszecki 00). As a result, in many situations they fail to reproduce the original image appearance. This is a limitation that has been addressed particularly in the last years. Gooch et al. (Gooch 05) introduced a technique that iteratively searches the gray levels that best represent the color differences between all color pairs. Considering the spatial distribution, each color pixel value is compared with the average of its neighbor region in order to preserve the local contrast in the converted image. This approach demonstrates good results for images with reduced number of chrominance values but its main drawback, $O(N^4)$ complexity, determines to be computationally prohibitive.

Rasche et al. (Rasche 05) suggested a similar solution that seeks to optimize a quadratic objective function that incorporates both contrast preservation and luminance consistency. The initial color levels are quantized to a reduced set - *landmark points*, in order to express a solution in terms of linear programming. Still computationally expensive, this approach may introduce artifacts mainly due to the quantization strategy.

Bala and Eschbach (Bala 04) carried out a small experiment to evaluate the perceptual quality of their results. They exploit the *Helmholtz-Kohlrausch* effect (Judd 66) to preserve the local chrominance edges. Smith et al. (Smith 08) developed a two-step algorithm built also on the *Helmholtz-Kohlrausch* effect. First, the image is transformed in a pixel-wise manner while maintaining the apparent color lightness. Then, a multiscale chromatic filter is applied in order to enhance the discriminability over the salient color features. Despite classified in the Cadik's study (Cadik 08) as one of the most accurate, due to the unsharp mask-related strategy, the method may introduce strong discontinuities along edges even for fine chromatic contrast adjustments in their provided range of values (see Figure 3.2).

The approach of Neumann et al. (Neumann 07) computes the gradient field based on the *Coloroid* (Nemcsis 87) decomposition. The values of relative luminance indices have been assigned based on an extensive user-study. The *Decolorize* method (Grundland 07) performs the transformation in their YPQ color space, a direct derivation of the RGB space. The algorithm performs a dimensionality reduction using the predominant component analysis, a related technique of the principal component analysis (PCA). Although, more computationally efficient, this approach does not take into consideration chromatic differences that are spatially distant, mapping in some cases different colors

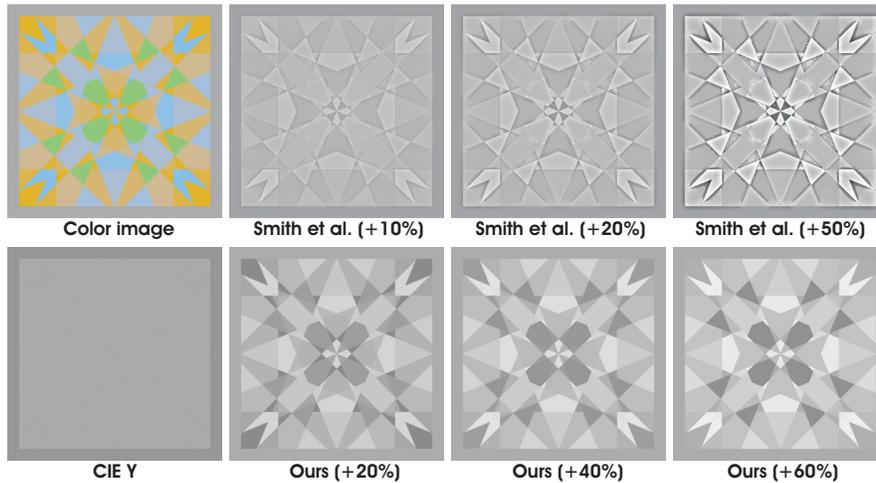


Figure 3.2: *Robustness of parameters to discontinuities.* Top line: Initial color image and the results of Smith et al. [2008] when varying contrast impact parameter with 10% , 20% and 50% of the default range (the maximum value is 4). Bottom line: CIE Y (luminance channel) fails to preserve the original appearance and our results when the equivalent parameter, γ , is varied with 20%, 40% and 60%. Note the robustness of our method that does not introduce artifacts even for large variations of the parameter γ .

into very similar gray levels. Moreover, because a single chromatic axis is not able to depict differently the existing chromatic changes the highlights can be mapped to obscure gray levels. As observed in our experiments this method is not suitable for enhancing operations presented in this work (see application section).

More recently, Kim et al. (Kim 09) have introduced a technique derived from Gooch et al. (Gooch 05) that aims to optimize a nonlinear global mapping function. It is more computationally effective, but like the previous optimization-based techniques (Rasche 05; Gooch 05) this strategy risks to destroy some of the fine details (see Figure 3.11). Moreover, the technique is not able to optimize reliably the considered parameters when images contain only a reduced number of colors (please observe their results of image sets 2,13,18,21,22 3.11 3.12 3.13).

3.3 Saliency-guided Decolorization

In this section we introduced the our decolorization algorithm. These model is inspired by the *color opponency* theory and requires knowledge about tonal values and saliency of the image.

3.3.1 Overview

Inspired of the art of photography (Adams 81), our algorithm is founded on similar principle as *color filters*. In the first step, the algorithm searches for information about saliency and tone structure. This step is motivated by the fact that using filters is tightly related with the existing tonal values of the scene but also with the object/part of interest of the scene. Next, a contrast enhancement scheme driven by the color filter calibrated on the fitted values (advantaging the regions that lost the saliency after decolorization) is employed. The result values of luminance are normalized and re-balanced to avoid an image with an over/under-exposure *look* and scene re-illumination. Finally, the luminance is mixed with the initial value in a proportion influenced directly by the output of the first step. This ensures consistent results, also for images with large chromatic range, where strategies based on a single dominant color axis may fail to preserve a consistent appearance since a single hue is highly advantaged (as will be discussed in results section).

Our approach is guided by the prominent chromatic information and introduces a mapping based on the *color opponency theory* (Judd 66), generally accepted as the foundation of many visual saliency models (Itti 98). *Color opponency theory* has also been successfully applied in related areas, such as color harmonization (Cohen-Or 06), detail enhancement (Benavente 03) and tone management (Pattanaik 98). As will be shown, the luminance values are computed pixel-wise by mixing both saturation and hue values, creating a new spatial distribution with an increased contrast of the interest regions. All the pre-computed values are normalized in order to fit the entire intensity range. The intensity is re-balanced in order to conserve the amount of glare in the initial image. For extreme lighting conditions, we apply several constraints in order to avoid clipping and fading of the apparent details.

3.3.2 Our Decolorization Approach

The algorithm has been implemented in the perceptually uniform $CIEL^*c^*h^*$ and HSV/HSL color spaces. In both cases, the strategy has shown robustness in preserving the initial color contrast discriminability and details (see Figure 3.4). However, we have determined experimentally that HSV/HSL ,

besides being faster, has the advantage to be more robust against adjusting parameters (e.g. the offset angle range can be highly compressed).

As we have mentioned earlier, the main goal of the algorithm is to increase the contrast of visually salient areas while maintaining the average gray shades that already exist in the image. Based on psychophysical experiments (Calabria 03), it seems that achromatic images are perceived to have a significantly higher level of contrast than images with only 20% of the original image chroma. The perceived contrast increases monotonically for images with chroma levels above this threshold. Therefore, enhancing the contrast of the luminance based on the chrominance information by a given degree γ can be expressed as:

$$\mathcal{L} = L(1 + \gamma S) \quad (3.1)$$

where L is the monochromatic normalized value of the luminance channel while S , the saturation, represents the chroma C normalized by lightness L .

In our approach, to integrate the chromatic polarity gain of the *opponent colors* (e.g. red pixels need to be mapped lighter when green pixels became darken), the output luminance value \mathcal{L} is computed based on the following chromatic filtering formula:

$$\mathcal{L} = L(1 + \gamma \cos(H\kappa + \phi)S) \quad (3.2)$$

where H represents the hue information of the chrominance channels and κ is the period. The parameter ϕ represents the offset angle of the color wheel ($0 - 360^\circ$) (see Figure 3.5). Its main effect is to set the chromatic enhancing filter on the offset position that best advantages the hue from the region that has lost its saliency due to the decolorization. By adjusting this coefficient, the shades are re-mapped to different gradients. The subsequent subsection goes into detail about how to optimally set this parameter.

The coefficient γ tempers the impact of the saturation and acts like a modulator that controls the amount of chromatic contrast. We found that for common desaturated images, the algorithm works well for γ set to small values (0.3) while for highly saturated images γ is assigned a higher value (≈ 1). This parameter ensures a linear dependency between saturation and global contrast that makes intense saturated images to be rendered with a higher global contrast (see the impact of varying γ in Figure 3.2). To generate the results in this thesis we used a default value $\gamma = 0.7$.

In our scheme, the multiplication by cosine plays an important role since it polarizes the LS gain value (equation 3.2) according to the period κ . In the absence of the cosine trigonometric function and for $\gamma = 1$ and if L and S have comparable levels then after the post-normalization process the new

value of the luminance remains almost constant. If L and S have different normalized values then the new luminance value increased/decreased according to saturation.

During our experiments, we found that in HSV/HSL color space a default value of $\kappa = 2$ results in a significant robustness of our operator. This robustness practically means that it is possible to considered a compressed range of offset angles ϕ (case that is often preferred due to target application requirements) and adjusting the γ parameter in a large range will not introduce significant degradation of the results. As our results demonstrate by setting a value of $\kappa = 2$ the *color opponency* is well preserved. The *color opponency* theory introduced by Ewald Hering (1878) assumes six independent unitary colors being based on three counterbalancing processes: black-white, yellow-blue, and red-green. Practically, our decolorized images display complementary gray levels for these color pairs. Moreover, as shown in Figure 3.6, this default value yields accurate preservation of the most salient regions extracted with the model of Itti et al. (Itti 98) that is characterized by a color map built on the Hering’s *color opponent* system.

For a value of $\kappa = 1$ (that theoretically ensures the color opponency) in both $CIEL^*c^*h^*$ and HSV/HSL color spaces our conversion performs decently, but in some cases the global contrast may be decreased (see Figure 3.4). The way ϕ is chosen makes the choice of $\kappa = 1$ less attractive since we have experimentally observed that our operator is less robust due to the fact that the offset angle range cannot be compressed to a reduced number of values as for $\kappa = 2$, and therefore a larger dataset of images would be required.

To prove the robustness of our operator the results in this work and the supplementary material have been generated automatically in HSV/HSL color space using the default parameter of $\kappa = 2$ and the compressed range of offset angles that contains only 3 main values (see subsection 3.3.3).

In some particular cases, the general equation 3.2 yields discontinuities on highlight areas. This is due to the fact that matte surfaces have a higher proportion of diffuse reflections while glossy surfaces exhibit a greater ratio of specular reflections. In order to solve this problem, for the pixels of the regions with the detected deviations, the LS gain is replaced by the average value \overline{LS} :

$$\overline{LS} = \left(\sum_{\Omega} L_{x,y} S_{x,y} \right) / N \quad S_{x,y} \geq \mu \text{ and } L_{x,y} \geq \nu \quad (3.3)$$

In this formula μ and ν are thresholds that filter highlighted regions without discontinuities, while N represents the total number of image pixels in the detected region Ω . The values of the thresholds that compensate for the saturation and intensity, μ and ν , are dependent on the selected color space.

In this work we generate the results using default values $\mu = 0.1$ and $\nu = 0.6$.

Adding previous constraints to equation 3.2, the luminance \mathcal{L}_{constr} can now be expressed as:

$$\mathcal{L}_{constr} = \begin{cases} L + \gamma \overline{\mathcal{L}S} \cos(H\kappa + \phi) & , \quad S \leq \mu \text{ and } L \geq \nu \\ L [1 + \gamma \cos(H\kappa + \phi)S] & , \quad otherwise \end{cases} \quad (3.4)$$

The *next step* is to normalize the luminance values to match the initial range:

$$\mathcal{L}_{res} = \chi \frac{\mathcal{L}_{constr} - \mathcal{L}_{min}}{\mathcal{L}_{max} - \mathcal{L}_{min}} \quad (3.5)$$

where \mathcal{L}_{min} , \mathcal{L}_{max} are the maximum and the minimum values after equation 3.4 has been computed on the entire image. The parameter χ (default value, $\chi = 0.9$) adjusts the global intensity being important to avoid an image over/under-exposure *look* or scene re-illumination. For a perceptually uniform output the luminance levels have to be related to the initial color values. It is well known that simultaneous combinations between *middle gray* and whiter or darker areas yield subjectively distinctive appearance due to simultaneous contrast (previously presented in section 2.8.2. Similar effects may be observed when combining colors (Adams 81; Fairchild 05) (e.g. red-green and red-blue). We state that the luminance has to remain in a certain range without exceeding the original levels of the input color channels R, G, B . Therefore, the intensity is restricted to remain in the range $[Min(R, G, B), Max(R, G, B)]$. This slicing technique is similar to the process found in electrical amplifier circuits, where a transducer can not amplify over the supplied power.

This step ensures that the original grayer, whiter or highlighted and darker or shadowed regions of the initial color image are preserved while enhancing the contrast. As a real example, for a scene in which the lightest object is a white paper, in comparison, a gleaming blue surface has far greater intensity than the diffuse white surface. This is inspired from the art of photography where the aim is to generate printouts in which specular reflections are brighter than diffuse areas since they lend a sense of brilliance (Adams 81).

A comparable postprocessing step has been applied also by Grundland and Dodgson (Grundland 07) but their conversion may fail due to the dimensionality compression when the color varies along dissimilar directions (please refer to the additional material) and therefore, in such cases the initial highlights appearance is not conserved.

Finally, the decolorized image \mathcal{I}_{dec} is the result of blending the initial luminance intensity L with the amount of chromatic enhanced luminance \mathcal{L}_{res}

computed with equation 3.5:

$$\mathcal{I}_{dec} = (\mathcal{L}_{res} + \eta L) / (\eta + 1) \quad (3.6)$$

where η is the parameter that controls the mixture ratio of the initial L and the enhanced luminance \mathcal{L}_{res} values. We believe that the way η is chosen may create a bridge between the *color opponency* and the well-known *Helmholtz-Kohlrausch* effect used in other grayscale conversions (Smith 08). We observed that selecting higher values (in the default range $\eta \in [0, 2]$) yields comparable results as with *Helmholtz-Kohlrausch* predictors. Please note in Figure 3.3 our result for $\eta = 2$ that has a comparable appearance with the grayscale palette of Smith et al. (Smith 08). To decolorize images and for image enhancing applications we set this parameter to $\eta = 0.2$. For videos, to yield consistent results, an average value of $\eta = 1.1$ is more beneficial. By using this value of η , the transition artifacts between adjacent frames are minimized while the saliency is well preserved.

3.3.3 Offset Angle Selection

In this section we elaborate on how the offset angle parameter ϕ is determined, based on the image’s color distribution. In order to find the optimal offset angle, our algorithm requires a dataset of reference images with an offset angle previously assigned to each image.

The reference images may be real images (see Figure 3.8 for examples) or synthetically generated using patterns of different color combinations (e.g. red-green, yellow-blue, red-green-blue, red-green-yellow etc.). The allocated offset angle of a reference image is the one associated with the most preferred decolorized version among all possible decolorized variants obtained in the entire range 0° - 360° . Therefore, in order to simplify the procedure, it is desirable that this range is compressed as much as possible. Theoretically, a various dataset of reference images would ensure a more accurate selection of the offset ϕ since, this would better reflect the user’s preferences but it would require a great deal of work to create such a large data set. Using this additional information, the offset angle of a given image is determined based on a histogram matching with a selection of images from the reference image database.

After extensive testing (on +4000 images) we found out that in *HSV/HSI* color space and with $\kappa = 2$ the offset angle mainly needs to cover the 180° - 360° . As previously discussed in *HSV/HSI* color space for $\kappa = 2$ this range can be extremely compressed to the discrete range of 200° , 250° , 300° with little loss of accuracy. This feature of our operator can be exploited in real

time application and for video decolorization. In other color spaces (such as $CIEL^*c^*h^*$ but also in HSV/HSL for $\kappa = 1$) the offset angle selection might be more problematic since such a large discretization of the optimal angle range is not possible.

For more accurate results, we developed a simple yet effective procedure that selects the optimal offset angle by only taking the most salient regions of the input image into account, rather than using the entire image. This strategy aims to preserve the main salient regions (in both color and decolorized images). In order to identify the salient parts of the image, we have opted for the well-known method of Itti et al. (Itti 98)¹ that consists of several biologically motivated steps. The model extracts three different feature maps that are blended into a final saliency map: intensity, orientation and color. The color map of this model is built with the intent of preserving the *chromatic opponency*. Other more recent models (Liu 07; Marchesotti 09; Wang 09) that search for the saliency can be applied as well. Even though they are slightly more precise in terms of segmenting the salient regions, these recent methods are more computationally expensive.

Based on this model we first identify the most salient regions in both the color and the standard grayscale (luminance channel) images (see Figure 3.7). For grayscale images we suppressed the color map of the model (Itti 98). A circular region is considered to preserve the saliency only if its position remains relatively constant in both versions (color/grayscale) of the image. We use the default parameters of the model and restrict to a number of maximum five regions (*green circles*). After comparison, the regions with diminished saliency, seen as regions which lack chromatic contrast, are easily identified (marked as *green circular regions* in Figure 3.7). Practically, the distribution of the hue H in these regions determines the selection of the offset angle ϕ .

The results presented in the thesis used also in the validation, were generated based on this fast strategy using the HSV/HSL color space ($\kappa = 2$) and only the discrete range of 3 values (200° , 250° , 300°) employed for the selected salient regions.

As shown in Figure 3.7, our operator is robust: it is still able to perform generally plausible even though the entire image is considered in the matching process. If the offset angle were randomly selected, the discriminability and the fine details may be lost.

Figure 3.6 shows the most salient regions extracted using the model of Itti et al. (Itti 98). For all grayscale images the color map was suppressed and therefore, only the intensity and orientation maps influence the final result.

¹<http://ilab.usc.edu/bu/theory/index.html>

As can be observed, our method compared to the standard grayscale and the method of Smith et al. (Smith 08), is able to preserve similar salient regions as in the original color image.

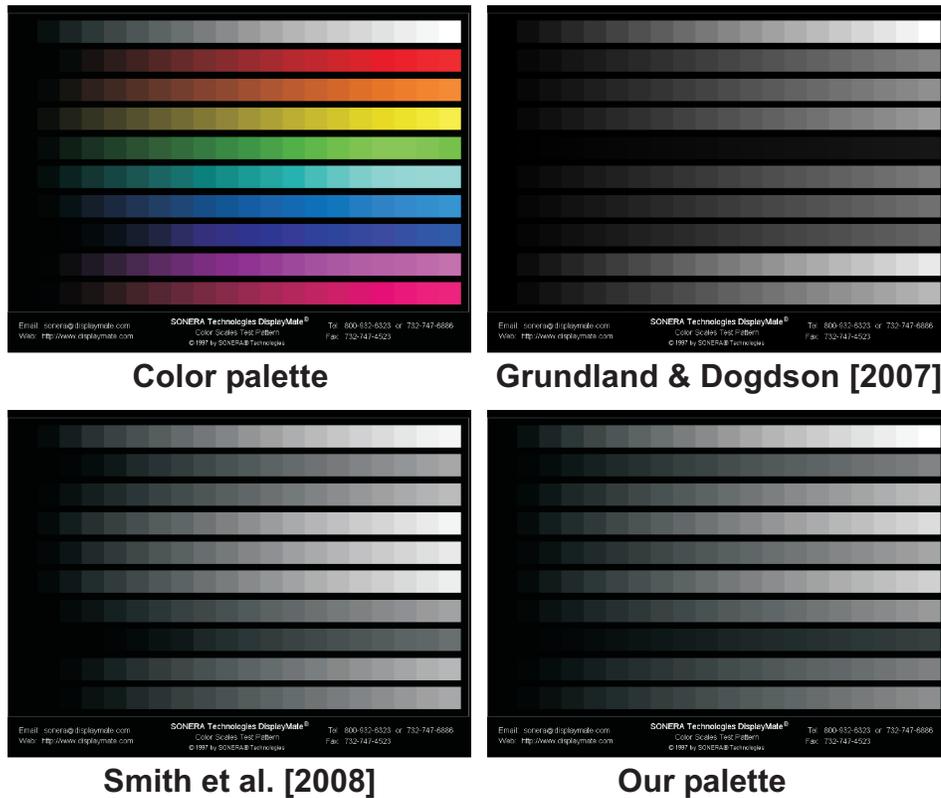


Figure 3.3: *Grayscale representation of the color palette. Strategies based on a single dominant color axis may fail to preserve a consistent appearance since a single hue is highly advantaged (notice Grundland and Dodgson [2007]). Our result generated using default parameters in HSL color space: $\kappa = 2$, $\gamma = 0.7$ while the offset was set to $\phi = 250^\circ$.*

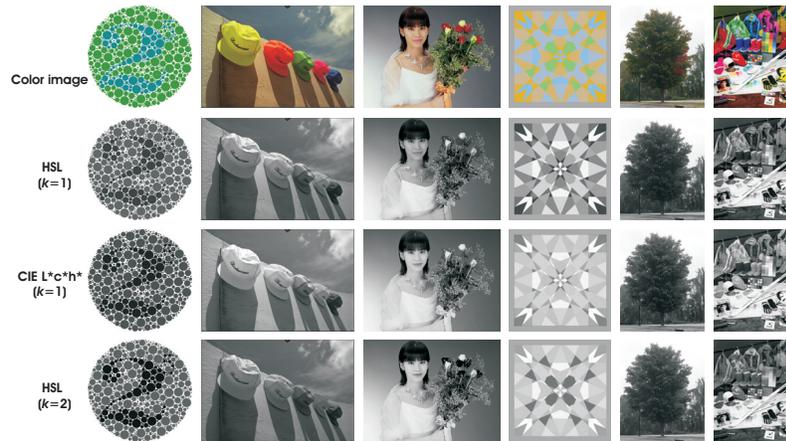


Figure 3.4: *Implementation in different color spaces (HSL and CIE $L^*c^*h^*$). The original color images are displayed on the top line. In the middle line are shown the results obtained by our operator in HSL and CIE $L^*c^*h^*$ for $\kappa = 1$ and the other parameters set on default values ($\gamma = 0.7$, $\chi = 0.9$); the selected offset angles ϕ from left to right: $300^\circ, 290^\circ, 240^\circ, 160^\circ, 320^\circ, 280^\circ$. The bottom line displays the results obtained by our fast saliency-based strategy in HSL by using default parameters ($\kappa = 2$, $\gamma = 0.7$, $\chi = 0.9$) and only the compressed offset range ($200^\circ, 250^\circ, 250^\circ, 300^\circ, 300^\circ, 300^\circ$).*

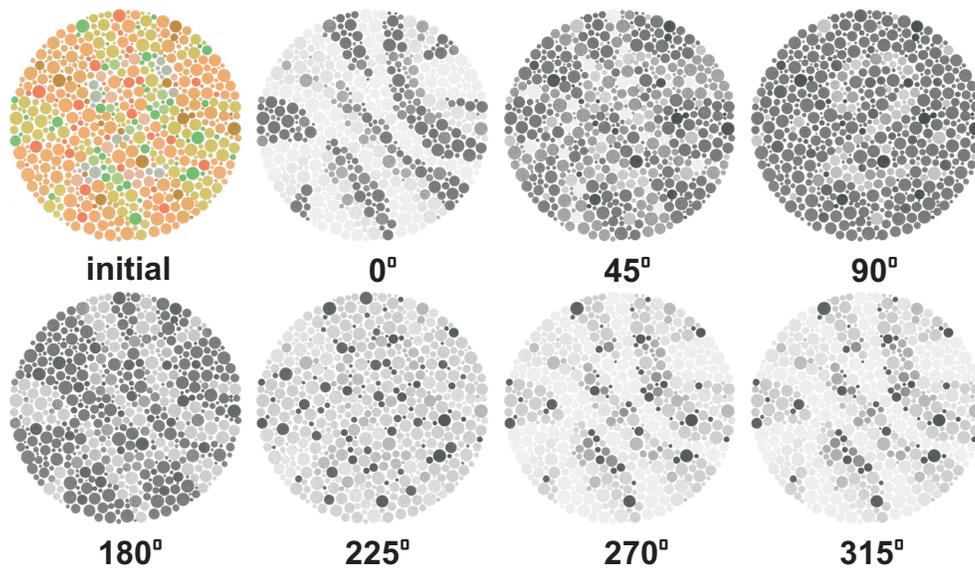


Figure 3.5: *Direct influence of the offset angle ϕ to the grayscale representation of the original color image. Note that our approach for entire range of the hue offset angle ϕ is able to preserve the initial white color of the background margins. The rest of the parameters are set to their default values.*

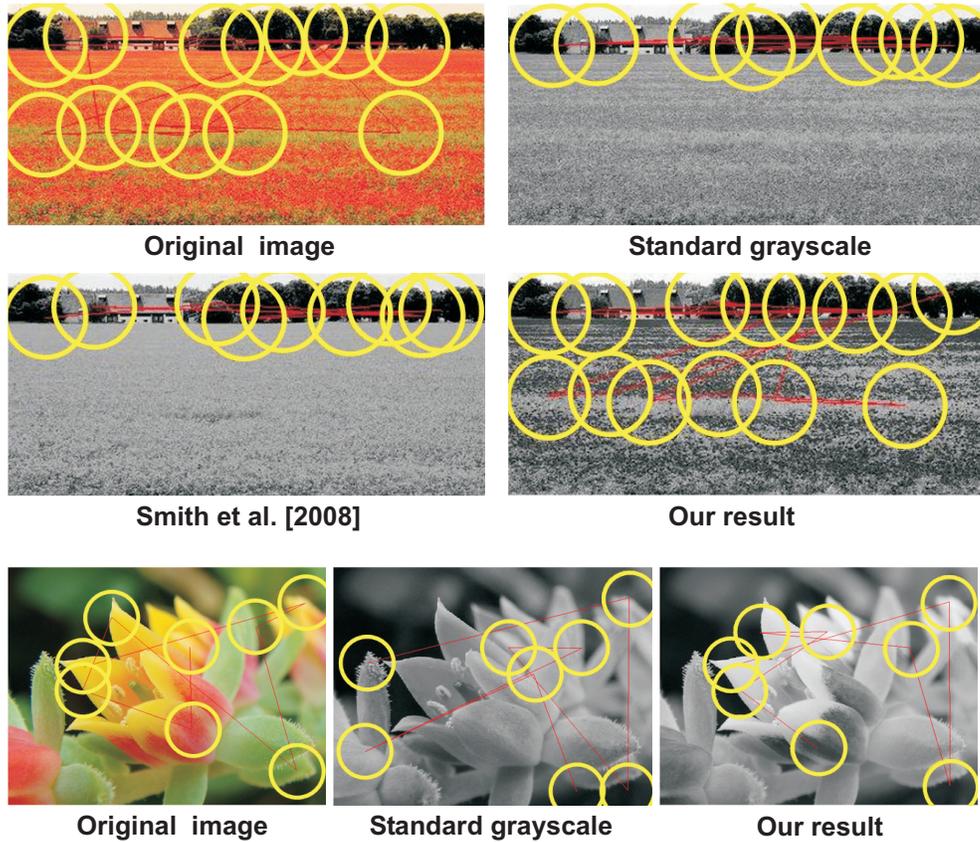


Figure 3.6: *Saliency preservations.* Giving a color image, our approach, compared with standard grayscale conversion and Smith et al. [2008], is able to conserve the initial color salient regions (yellow circular regions). Our results have been generated in HSL using default parameters $\kappa = 2$, $\gamma = 0.7$ while offset angle was set to $\phi = 200^\circ$ (top image) and $\phi = 250^\circ$ (bottom image).

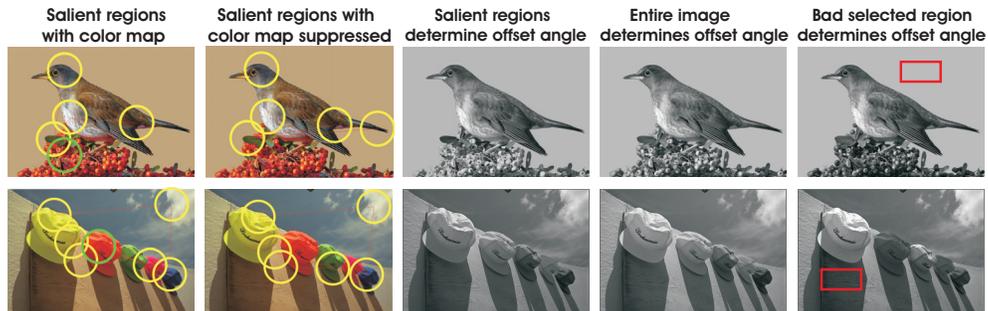


Figure 3.7: *The influence of salient regions on the offset angle selection.* From left to right: the color image with the salient regions extracted when the original Itti et al. model is applied (including the color map). In the second column are shown the salient regions extracted by Itti et al. model when color map was suppressed. The green circular salient regions (shown in the first images) are those that were not preserved after suppressing the color map. These regions determine the optimal offset angle - $\phi = 250^\circ$ (top) and $\phi = 250^\circ$ (bottom) (the decolorized images presented in the next column). Next are shown the results when the offset angle ($\phi = 300^\circ$ (top) and $\phi = 300^\circ$ (bottom)) is selected by using the entire image in the process of matching (note the details of the dragon, hats and sky). The last column show the results when a bad region (red rectangle) determine the selection of the offset angle $\phi = 200^\circ$ (top) and $\phi = 200^\circ$ (bottom).



Figure 3.8: *Example of reference images.* From left to right the assigned offset angles (HSL color space) are 200° , 250° and 300° .

3.3.4 Results and Discussion

A common problem of existing color to grayscale conversions is the parameters tweaking. Despite seeming complicated on the first glance, the results shown in this thesis and supplemental material (e.g. the entire testing set of 24 images 3.11 3.12 3.13) have been generated using the default values of the parameters discussed and motivated in previous section. The reliability of this automatic function of our operator has been obtained due to an extensive analysis based on a large dataset of images. However psychophysical experiments (Calabria 03) indicate that (especially) naive observers have large response variability in chroma-contrast experiments. For this reason, decolorization methods need to provide a set of intuitive parameters to the users. Besides personalized perception, this alternative allows creative interpretations. For example in a natural scene, the shades of gray that portray a water surface can vary over a considerable range. However, for such cases we sought for parameters that are robust against artifacts. The scheme proposed by Smith et al (Smith 08) offers several free parameters to the users, but tuning them risks to introduce artifacts close to edges (see supplementary material).

The runtime of our operator is linearly dependent on the image resolution and therefore it is computationally effective being suitable for real-time applications. The saliency employed is unfortunately highly computational expensive. We plan in the future to employ the more recent technique of Achanta (Achanta 09). However we have noticed several limitations of this techniques of this technique that we have employed in the decolorization by fusion algorithm (present in section 5.2.2). This saliency algorithm can fail to produce a consistent regional contrast since only the global contrast difference are accounted. We would like to extend this technique in order to consider simultaneously the global contrast and the spatial coherence.

During tests of the fusion algorithm, we have noticed that the employed saliency can fail to produce a consistent regional contrast since only the global contrast difference are accounted.

A limitation of our scheme is the amplification of already existing artifacts in the images or video frames due to compression. Furthermore, manually adjusting the offset angle parameter can yield reversed chromatic contrast (e.g. red may appear darker than green). Our saliency-based strategy may fail in cases when the employed regions (marked with green circle) do not reflect the true salient information of the image. This mainly occurs when specular reflections are identified as salient regions (see Figure 3.9), but also when the object/region of interest is not clearly distinct from the background.

The algorithms is able to generate robust results even for high variations of



Figure 3.9: *Conversion failure.* The saliency-based strategy may fail in cases when the employed region (marked with green circle) does not reflect substantially the true salient information of the image. In this case the discriminability between red and green regions is not reflected in the grayscale version.

the parameters. This fact relies on the well-defined constraints. For visually-substitutions systems or for automatic systems that use decolorized images, this is an important requirement, otherwise the image process can be affected by input errors. Most of the available algorithms are highly dependent by the parameters, and the result is not validated in any way. We show in figure 3.10 that our straightforward constraints can be applied successfully for other operators. The main effect of the constraints is that the shades of gray, white and black appearance are recovered and the results presents less artifacts.

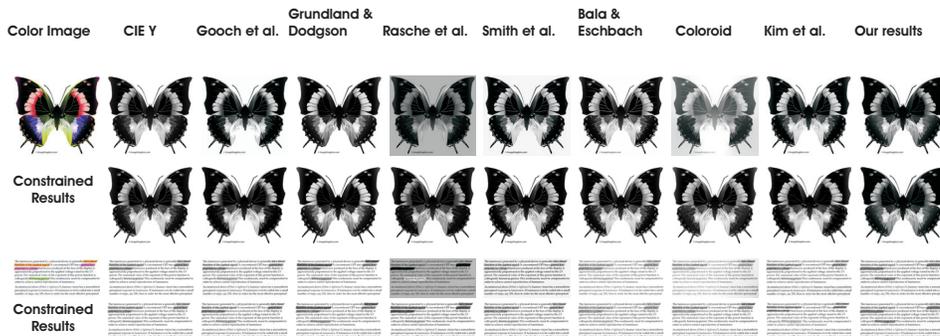


Figure 3.10: *Constraints effect* To demonstrate the effect of the final constraint we have applied it on several decolorized images obtained by different techniques.

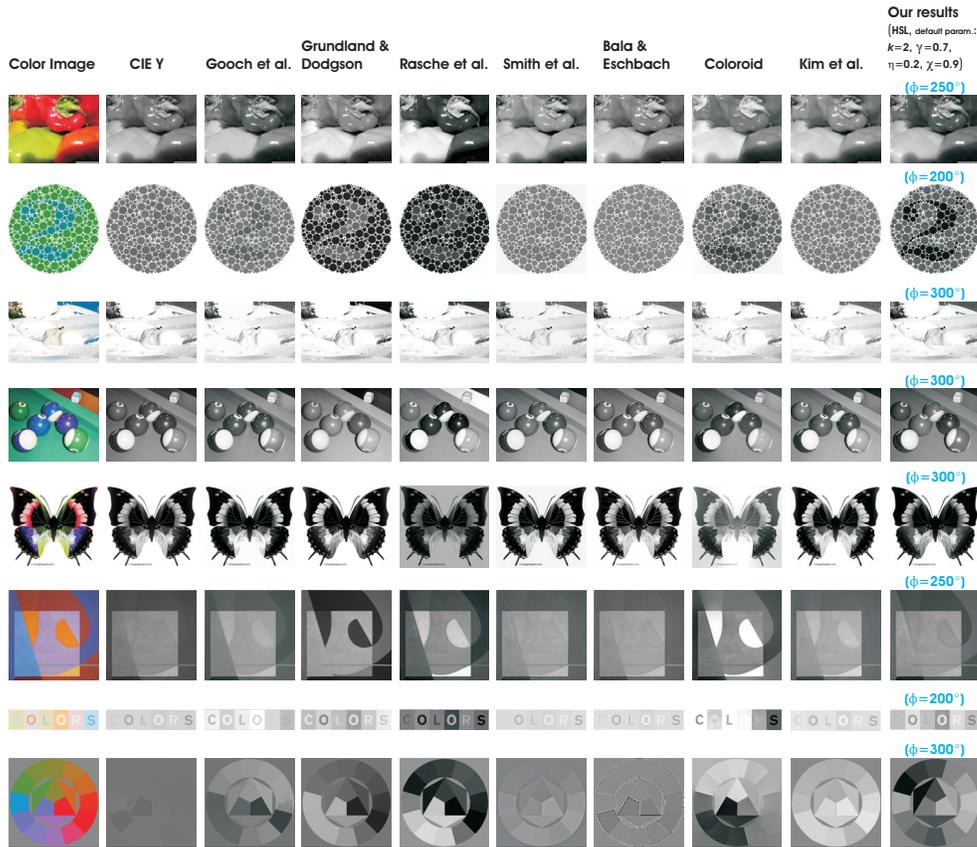


Figure 3.11: *Comparative results.* From left to right: color image, CIEY (standard luminance), Gooch et al. (Gooch 05), Grundland and Dodgson (Grundland 07), Rasche et al. (Rasche 05), Smith et al. (Smith 08), Bala and Eschbach (Bala 04), Coloroid (Neumann 07), Kim et al. (Kim 09) and our results with the employed parameters.

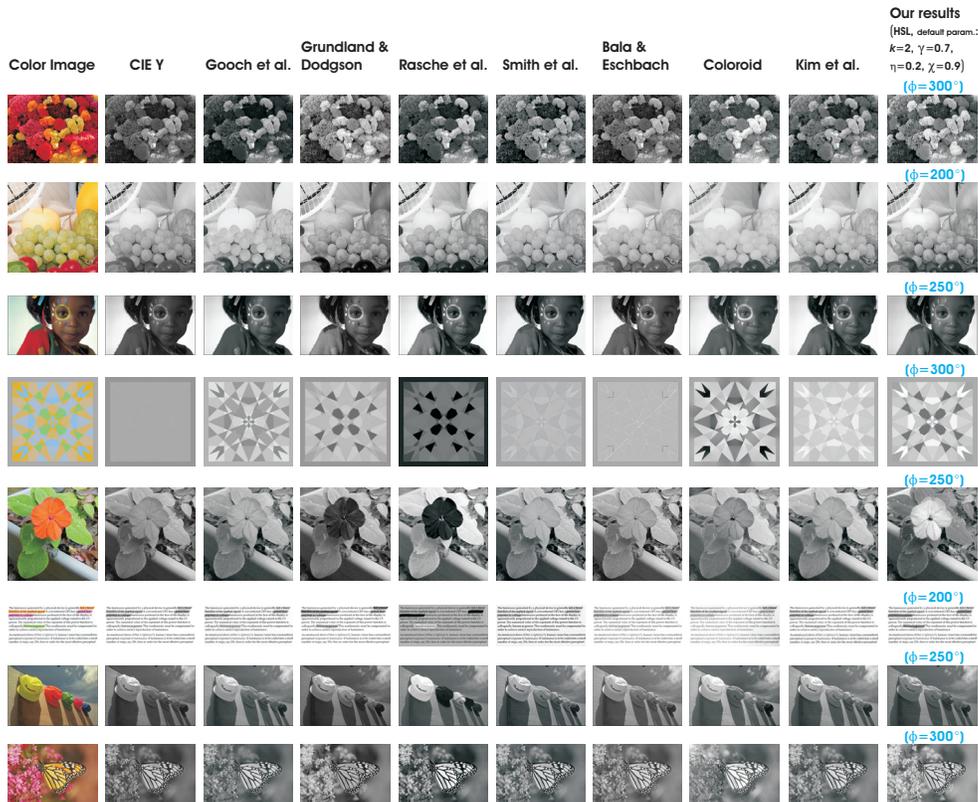


Figure 3.12: *Comparative results.* From left to right: color image, CIEY (standard luminance), Gooch et al. (Gooch 05), Grundland and Dodgson (Grundland 07), Rasche et al. (Rasche 05), Smith et al. (Smith 08), Bala and Eschbach (Bala 04), Coloroid (Neumann 07), Kim et al. (Kim 09) and our results with the employed parameters.

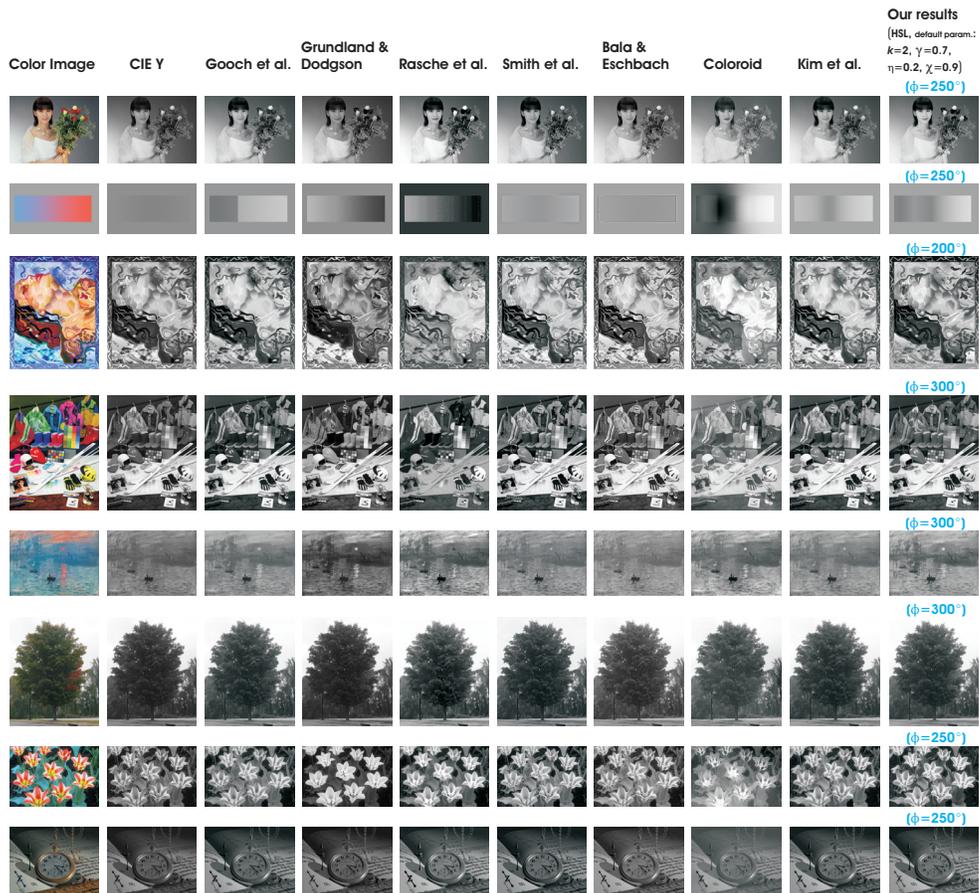


Figure 3.13: *Comparative results.* From left to right: color image, CIEY (standard luminance), Gooch et al. (Gooch 05), Grundland and Dodgson (Grundland 07), Rasche et al. (Rasche 05), Smith et al. (Smith 08), Bala and Eschbach (Bala 04), Coloroid (Neumann 07), Kim et al. (Kim 09) and our results with the employed parameters.

3.3.5 Validation

Similar to Cadik’s study (Cadik 08), for completeness we also managed a perceptually validation of the grayscale operator. The evaluation considers in addition to our technique and *CIEY* (luminance channel) several recently introduced decolorization strategies (Bala 04; Gooch 05; Rasche 05; Grundland 07; Neumann 07; Smith 08; Kim 09). For the test 14 volunteers were been involved (9 computer graphics students and researchers and 5 participants with no particular knowledge of computer graphics techniques) in the group of ages (21-45). Before starting the evaluation the grayscale problem has been briefly introduced to the volunteers (written instructions were also available during the test). All volunteers had normal or corrected-to-normal visual acuity and no issues of the color vision. The participants have been asked to rank the grayscale images focusing mainly to the overall appearance, details and contrast preservation. All the grayscale transformations together with the initial color image could be visualized simultaneously by *Cooliris*², a transparent and very intuitive interface to use.

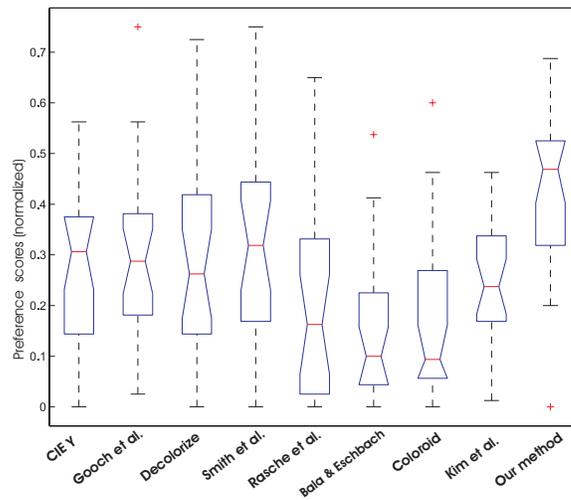


Figure 3.14: *Statistical interpretation of the perceptual evaluation results based on ANOVA ($p=1.6242E-06$, $F=5.67$).*

The images have been visualized on a calibrated 24-inch LCD monitor (Samsung SyncMaster 245B+) at native resolution in a segregated room where the

²<http://www.cooliris.com>

level of darkness was maintained approximately constant for all tests. Different than Cadik's study (Cadik 08) (where the test took about 20 min per observer), the volunteers had the entire control resulting in an average time of 90 min per test. The following ranking strategy has been used: for every set of images the best four results received high scores (80%, 60%, 40% and 20%) while the rest of them have been classified as acceptable (scored with 10%) or not acceptable (not scored). The results have been interpreted statistically using analysis of variance (ANOVA) (Tabachnick 05). As a general remark, analyzing the graphic of Figure 3.14, the method of Smith et al. (Smith 08) and our approach have been ranked as the most perceptually accurate methods. Additionally, as in Cadik's study it can be observed that the methods (Rasche 05; Gooch 05; Kim 09) that optimize an objective function have been classified as less perceptually accurate.

3.4 Video Decolorization

An important utility of this task is in substitution systems for visually impaired people (Durette 08) that provide assistance for visually disabled persons by exploiting other available senses. For this task, in order to compress the amount of information that needs to be translated, processed grayscale frames are employed. Since the goal is to guide the subjects's focus to prominent regions in order to better localize objects in the scene, in such cases rendering the most salient features is more critical than for people with normal visual acuity.

Seen as an extension of the image grayscale conversion, decolorizing videos is more complex due to the temporal coherence that needs to be preserved between adjacent frames. Firstly, in our algorithm we search in the entire sequence for the color palette that appears in each image (mostly identified with the static background). For sequences in which the color palette remains relatively constant, a single offset angle (ϕ) value computed for the middle frame, yields pleasant results. For more challenging cases (e.g. scenes that alternate very distinctive color palette schemes), the offset ϕ is precomputed as an average value of several frames, while constraining its variation in a certain range. To minimize the details loss, for the video decolorization task a value of $\eta=1.1$ (the other parameters are set to their default values) yields consistent outputs. Since this value influences the mean value of the image, the transition artifacts are minimized.

Compared with Smith et al. (Smith 08) approach (see Figure 3.15) our decolorized frames have a higher discriminative chromatic contrast, thus better preserving the initial saliency. For the entire movie (one of its frames is shown

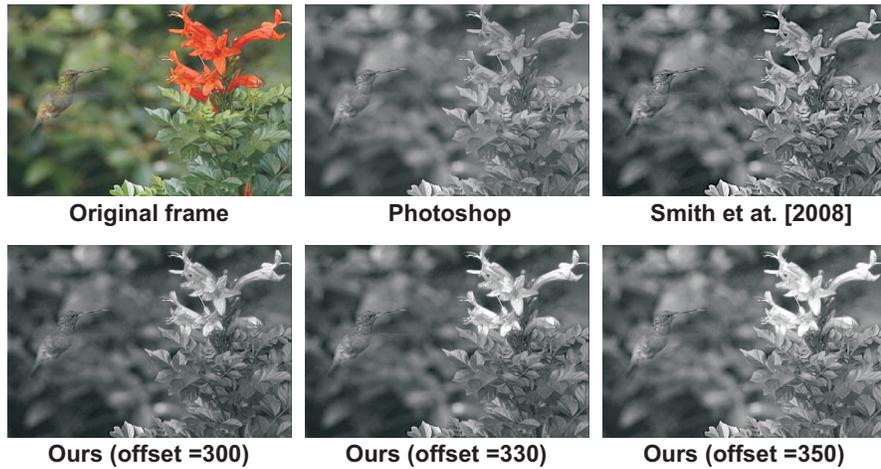


Figure 3.15: *Video decolorization.* From left to right: initial color frame, standard grayscale (Photoshop), Smith. et al [2008] and our result for ϕ values of 300° , 330° and 350° .

in Figure 3.15) the optimal offset angle was found to be $\phi = 300^\circ$. It should be noted that even though an globally optimal value could be found for the video, there will always be a trade-off as in some frames different particular details will be lost. For example in case of this video when increasing the offset angle value ($330^\circ, 350^\circ$) even though the global contrast is enhanced, on a closed inspection some of the details might be lost. For the complete video sequence the reader is referred to the supplementary material.

3.5 Other Applications

As our decolorization is accurate and preserves finest details, we can exploit variations in chromacity as well as luminance to enhance color images simply, blending the processed luminance levels with the initial color. Observed also by Gooch et al. (Gooch 05), the human visual system has better spatial acuity in luminance than in chrominance; by fusing the chrominance with co-located luminance changes, weak chroma variations are rendered sharper and better-localized. To the best of our knowledge we are the first to demonstrate such as wide applicability of a grayscale operator. We obtain comparable results with specialized techniques for tasks like segmentation under different illuminants, detail enhancement and wide-baseline image matching that will be presented and discussed in the following subsections.

3.5.1 Detail Enhancing

The main goal of image detail enhancement is to emphasize the image features for display and analysis. As observed by Gooch et al. (Gooch 05) substituting the luminance channel with the grayscale conversion and recoloring back can yield more pleasant color and grayscale images for printouts. Due to the continuous grayscale mapping and preserving initial details, our technique qualifies for this operation. The LDR image in Figure 3.16 was obtained by tone mapping with the *photoreceptor* (Reinhard 05) and therefore the large variations in luminance among regions can not be solved by employing standard operators (e.g. brightness, contrast adjustment available in the commercial tools) that shed details. For this task we employ an iterative colorizing/decolorizing strategy that starts from the offset angle ϕ , computed as presented in the previous section, and successively blends the resulting luminance after every iteration with the color that has been corrected. For color correction we employed the algorithm of Mantiuk et al. (Mantiuk 09) while the rest of the parameters are kept to their default values.

The method has an effect comparable to contrast filters (Adams 81). Our operator has abilities to manipulate global chromatic contrast without using additional information (e.g. images or hardware). The concept is similar to the *center-surround* (Benavente 03) frequently used by local enhancing methods, where the mapping varies spatially dependent by the neighborhood of the pixel. However, our strategy applies this concept as a global operation employing the same mapping to all image pixels. Because of the global character of our operator, the presence of halving artifacts commonly associated with local methods are significantly reduced.

An alternative solution (Smith 06) is to manipulate the chrominance for detail restoration. In order to measure the quality of the produced results we opted for the recent metric of Aydin et al. (Aydin 08) that identifies three classes of contrast changes relative to the original image: loss of visible contrast (green), amplification of invisible contrast (blue) and reversal of visible contrast (red) (practically (Aydin 08), **green** is related with *blurring* while **blue** and **red** are related with *sharpening* operation).

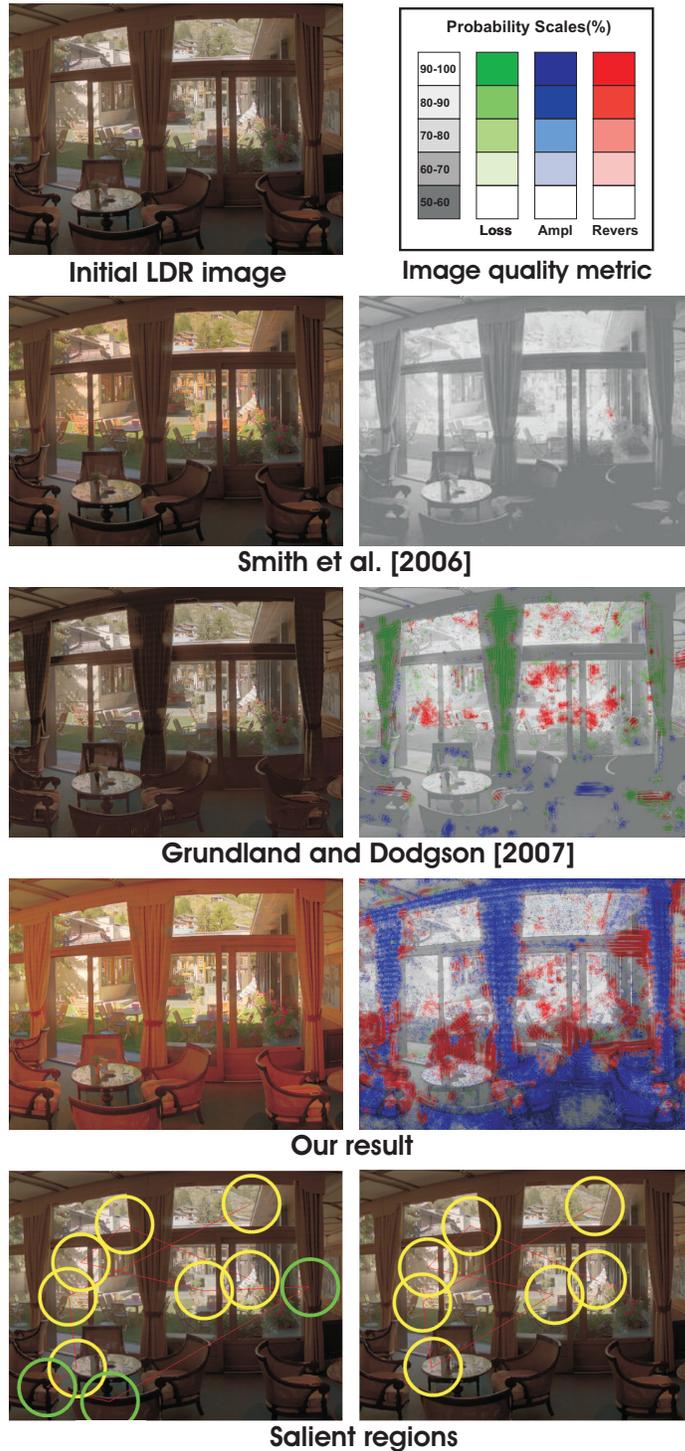


Figure 3.16: *Detail Enhancement*. From top to bottom: LDR initial color image and Aydin et al. [2008] quality metric, details enhanced by Smith et al. [2006], Grundland and Dodgson [2007] and our enhanced results (HSL color space, $\kappa = 2$, $\gamma = 0.7$, $\chi = 0.9$, $\eta = 0.2$ and $\phi = 250^\circ$). Notice that our approach is able to enhance outdoor details while interior details (e.g. curtain texture) become more visible. The right column shows the image quality assessment results while the bottom line shows the selected salient regions when the Itti et al. model is employed with color map (left image) and after color map was suppressed (right image).

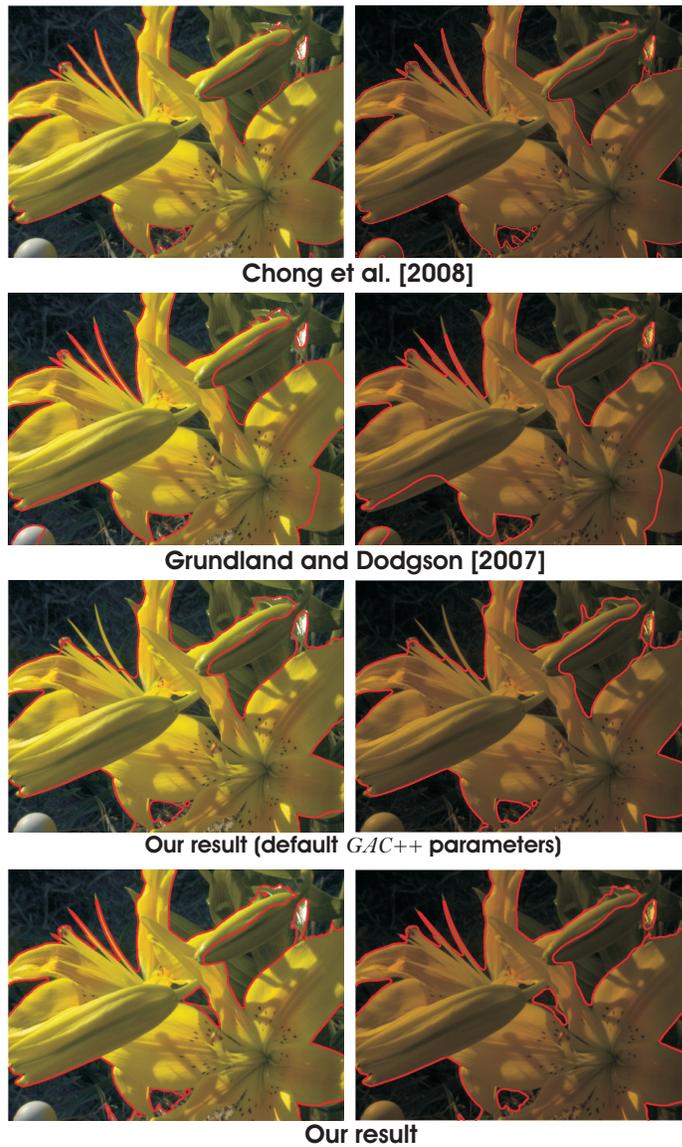


Figure 3.17: *Segmentation under different illuminants.* Considering the same image illuminated differently (left-right), the top line displays results obtained by employing the perceptual color space of Chong et al. [2008] after they tweaked parameters for *GAC++*. Below, the approach of Grundland and Dodgson [2007] yields inconsistent results. The next two lines show our results with and without default *GAC++* parameters (we kept the same parameters for our operator). Note that the rounded white object in the left bottom is correctly unselected in our approaches in comparison with the others since the foreground color is yellow. Additionally, the middle opening in the flowers is observed only by our operator when tweaking the *GAC++* parameters. To generate our results except for the χ we used the default parameters in HSL with $\phi=300^\circ$.

3.5.2 Segmentation under Different Illuminants

Image segmentation, a classical computer vision problem, aims to segregate distinctive foreground objects from the background. Despite of the important progress that has been made in the analysis and formalization of the segmentation it remains an unsolved problem in the general case. A known difficulty in image segmentation is that due to the illumination variation highly inconsistent outputs are observed. The recently introduced perceptual color space (Chong 08) was designed to solve the illumination invariance. By only minor adjustments, our operator may represent a decent solution for this problem. For this particular task, we disable the constraint of our algorithm that limits the illumination to remain in the range $[Min(R, G, B), Max(R, G, B)]$. As presented in the algorithm section parameter χ controls the global intensity of the images. Therefore, for this task we simply set χ to the default value (0.9) for the reference image (the one that is relatively well illuminated) while for the second image in order to compensate the differences of illumination it varies proportionally with the luminance ratio of two corresponding points selected from the foreground.

Since the offset (ϕ) selection depends on the salient regions, the scene is reliably decolorized while favoring objects from the foreground. Running the same algorithm (the geometric active contour *GAC++*³) for the same pair of images used in the work of Chong et al. (Chong 08), rendered by varying the light, we obtained comparable results. The same strategy has been applied for the other grayscale operators (Figure 3.17 displays only the results of the *Decolorize* (Grundland 07) approach). As can be noticed in Figure 3.17, we firstly employed the *GAC++* algorithm with default parameters and afterward we applied the *GAC++* with tweaked parameters for the same grayscale images (the parameters of our operator have been the same as in the first example).

3.5.3 Matching by Local Feature Points

Image matching is one of the fundamental problems in computer vision being mainly performed for grayscale image versions (luminance channel). Given at least two different views of the same scene, the 3D geometry is commonly recovered by matching robust local feature points (Tuytelaars 08). Recent evaluations (Mikolajczyk 05; Moreels 07) disclosed that the most powerful operators are those derived from the well-known SIFT (Lowe 04). However, these studies have been performed only considering grayscale images

³<http://cvsp.cs.ntua.gr/software/GAC++/>

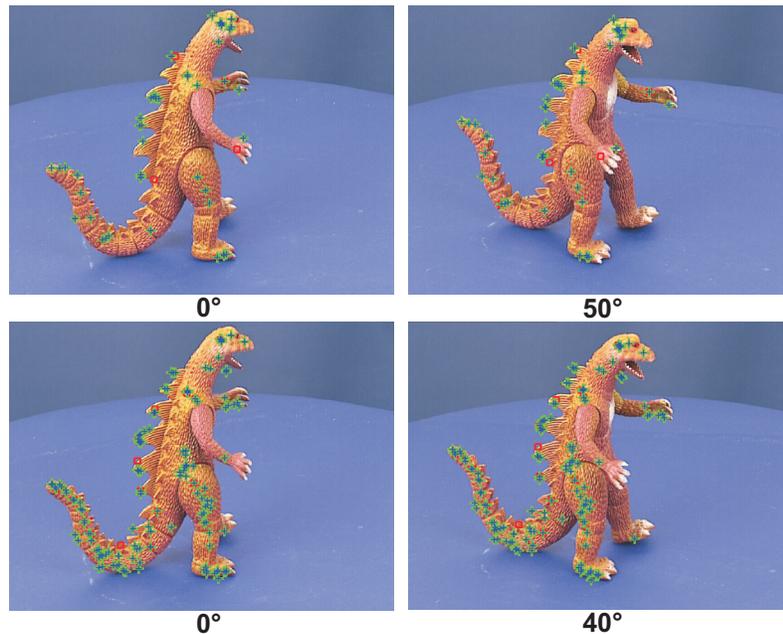


Figure 3.18: *Improving wide-baseline matching performances. Initial image and versions taken under different view-angle differences of the camera. When view-angle between cameras is 50° our approach (green crosses), by only varying the offset angle ϕ between 240° and 360° with 10° step, is able to find 54 good matches compared to only 5 correct matches obtained by the original SIFT approach (red squares). If the difference is 40° , SIFT filters 20 valid matches while our strategy finds 145 correct matches.*

yielded by the standard conversion. However, adding the color information (Abdel-Hakim 06; Huang 08) it does not seem to improve significantly the matching results.

On the other hand, local contrast preservation is crucial in the process of matching by local operators. This has been observed as well by Lowe (Lowe 04) and therefore for the SIFT operator the candidate keypoints with low contrast are rejected in order to decrease the ambiguity of the matching process. In our experiments we considered the well-known SIFT operator (Lowe 04) tested for the images of the OXFORD⁴ *dino toy* model in cases where the view-angles change with 40° and 50° , respectively. Note that the *dino toy* model does not contain large planar surfaces, and as such it is an instance of a challenging

⁴<http://www.robots.ox.ac.uk/vgg/data/data-mview.html>

wide-baseline matching problem.

Kolomenkin and Shimshoni (Kolomenkin 06) observed in their work, where SIFT was also employed in the context of image matching, that the standard contrast enhancement approaches are not able to exploit the photometric information for this task. They presented a complex and computationally expensive approach that after images were segmented assigns with a probability two segments that originate from the same surface. Even then, their experiments have shown only a slight improvement over the results of the original SIFT approach. Our technique, by only varying the chromatic offset parameters (the rest of the parameters are set to their default values) for the same pair of images, is able to filter a significant additional number of correct matches (see Figure 3.18). The accuracy of the geometry is verified for both approaches by the method of Forssen and Lowe (Forssén 07). Note that for a viewing angle difference of 50° , SIFT is unable to find sufficient matches to estimate a reliable epipolar geometry. For this example, our approach is able to produce 54 valid matches. The proportion of mismatches (not shown here) remains constant for both approaches. Since this task relies on the locality of the feature points, the accuracy of our conversion does not distort the finest details, but rather enhances the local and global contrast.

3.5.4 Decolorizing Images for Auditory Substitution Systems

Our ability of mobility and orientation is based on the capacity of mentally mapping the spaces and the possible navigation paths in the environment. Since much of this information is acquired through the sight sense, visually disabled persons face great difficulties to orient in novel environments. Recently, there has been an increasing attention in the development of portable non-invasive substitution systems (Meijer 92; Capelle 98; Pun 07) designed for visually disabled persons. To compensate the deficiency in visual sense, these systems translate the acquired image and make it available to other senses. The aim of such systems is to induce representations or mental images for visually disabled users (in general proficient users) due to imaginary process. Investigations in the field of neural rehabilitation are explaining these phenomena through cross-modal brain plasticity, where large areas in brain cortex (of the visually disabled persons) are recruited to process non-visual tasks (Capelle 98; Auvray 05; Bavelier 02).

In general the existing substitution systems employ decolorized images, obtained by standard grayscale transformation, and modulate the amplitude of signal proportional with the pixel intensity value. However, the standard technique that considers only the luminance channel, may fail to interpret ac-



Figure 3.19: From left to right on the first row: color image, standard grayscale and our decolorized result. On the second row the comparison results after employing the metric of Aydin et al. (Aydin 08). Notice that standard grayscale loses some of the visual contrast, while our decolorization amplifies the visual contrast.

curately the scene appearance due to the fact that the color information is not considered (see Figure 3.19). Separately, the color does not provide enough information about objects shapes and scene geometry. Nevertheless, by modeling a gray translation with color information we can implicitly identify color cues (mostly corresponding to texture or particular classes such as sky, grass or flowers) beside those cues that are provided by intensity variations. Additionally, a good interpretation of the most salient regions overcomes deprived information about the most attractive areas in the scene and leads to focus the attention to the important regions. By this strategy the chances to identify objects and persons into the scene are substantially increased (see Figure 3.20).

On the designing process of the vision substitution system there are some initial assumptions on tailoring the vision sensory over the sound sensory. These constraints can increase the quality and the quantity of the transferred information of the vision sensory. Sound segregation capacity has some similar correspondences with the scene analysis (Wilson 99). Experiments on auditory segregation (Bregman 90) showed that an alternate sequences of high and low frequencies tones played at different rates influence the segregation sensation. When the stream is played at slower rate, the listener is able to follow the entire sequence of tones. At higher rates, the sequence splits into two streams, one high and one low pitch, being difficult to follow the entire sequences of tones. Auditory stream segregation related with the sound frequencies seems to follow the characteristics of apparent motion in human vision systems (Strybel 98).

Since our decolorizing operator is able to preserve the high contrast ap-

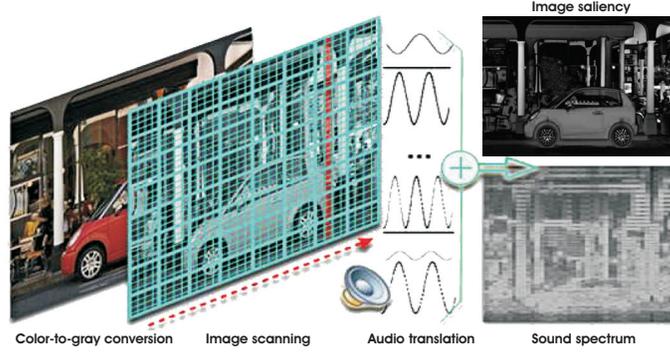


Figure 3.20: Overview of the *vOICe* auditory-substitution system (Meijer 98).

pearance of salient regions we found appropriate to employ our strategy for auditory-substitution systems. Different than existing techniques that employ quantization and expensive optimization (Rasche 05; Gooch 05; Kim 09), our fast decolorization method that enhances the contrast of the grayscale image to optimally reflect the chromatic contrast of the initial color image. Additionally, our goal is to reduce the loss of visual information of the converted image. The utility of our decolorizing operator has been investigated for the sound substitution systems. Among the existing sound substitution systems (Cronly-Dillon 99; y Rita 03; Valazquez 05; Capelle 98; Arno 99), (Auvray 05), we integrate our strategy into the well-known *vOICe* (Meijer 98) system that is described briefly in the remaining part of this section.

The *vOICe* (Meijer 92; Meijer 98; Amedi 05) system translates the acquired frontal images into a time-multiplexed auditory representation. Each image is rendered with a resolution of 64×64 pixels in an approximate conversion time of $T = 1.05$ seconds. The translation operation is a per-pixel operation by encoding the vertical position into frequency and the horizontal position into time. The pixel intensity gives the oscillation amplitude, therefore white is mapped into *loudness* and black is mapped into *silence* of its associated oscillator.

Firstly the image matrix elements are associated with one of the G gray tones:

$$\mathcal{P}^k = (p_{ij}^k) \quad , p_{ij}^k \in \{g_i, \dots, g_G\} \quad (3.7)$$

$$i, j = 1 \dots N, N = 64$$

where i and j represent the columns and lines indexes that are limited to the maximum values $N = 64$ (the input image has a resolution of 64×64 pixels).

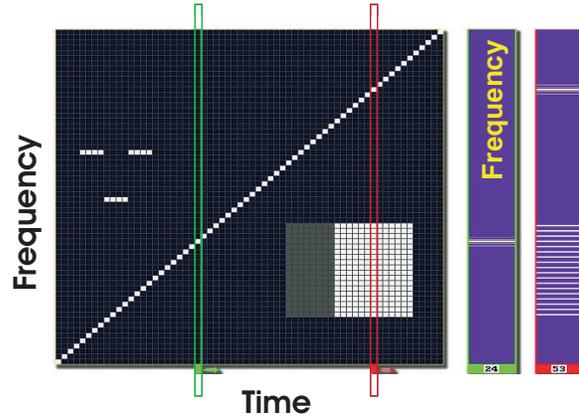


Figure 3.21: Time-frequency multiplexed auditory representation.

Each of the N column that corresponds to the signal $s(t)$ is played in T/N seconds. As already presented, the amplitudes of sinusoidal components of the $s(t)$ signal are proportional with the intensity levels. Considering that $\omega_i = 2\pi f_i$ the sound pattern transformation is mathematical expressed as following:

$$\begin{aligned}
 s(t) &= \sum_{i=1}^N p_{ij}^k \cdot \sin(\omega_i t + \theta_i^k) \\
 t &\in \left\{ t_k + (j-1) \cdot \frac{T}{N}, t_k + j \cdot \frac{T}{N} \right\} \\
 j &= 1 \dots N, k = 1, 2, \dots
 \end{aligned} \tag{3.8}$$

The algorithm computes frequency distribution equidistant as expressed in equation 3.9. In addition to linear frequency distribution the approach allows also exponential distribution of frequency to render the patterns (see equation 3.10):

$$f_i = f_l + \frac{i-1}{N-1} \cdot (f_h - f_l), \quad i = 1 \dots N \tag{3.9}$$

$$f_i = \left(\frac{f_h}{f_l} \right)^{\frac{i-1}{N-1}} \cdot f_l, \quad i = 1 \dots N \tag{3.10}$$

where f_l (default $f_l = 500\text{Hz}$) and f_h (default $f_h = 5\text{KHz}$) are the lowest and respectively the highest frequency.

Finally, after each image, as a distinct end-of-frame mark is inserted a synchronization click sound that indicates the end of the played image, respectively the beginning of a new input.

3.6 Conclusions

In this chapter we described an straightforward yet effective decolorization algorithm that enhances the contrast with respect to the appearance and the quality of the original color image. In our strategy the monochromatic luminance channel is intensified by mixing both saturation and hue channels resulting in enhanced chromatic contrast. This yields a new spatial distribution that finally is re-balanced in order to conserve the amount of glare impression that characterizes the original version. The main advantage of our algorithm is that it generates conversions with higher saturation aspects, while conserving fine detail in the highlighted regions. Besides efficient decolorization, we proved the utility of our technique for several challenging applications. In general, decolorization is very subjective being dependent by how human beings perceive the chroma and color distribution. Therefore, we undertook a perceptual evaluation that demonstrates the potential of the new method to preserve the original saliency of the color images. As future work we intend to investigate the potential of our method using different saliency techniques. We plan to research this method in the context of several other applications such as scene relighting and single image reconstruction.

Chapter 4

Single Image Dehazing

Contents

4.1	Introduction	74
4.2	Related Work	76
4.3	Optical Model	77
4.4	Single Image Dehazing by Per-Pixel Haze Detection	78
4.4.1	Haze Detection	79
4.4.2	Airlight Color (A_∞) Estimation	82
4.4.3	Layer-based Dehazing	83
4.4.4	Experimental Results and Discussion	85
4.4.5	IQA Assesment of Dehazing Methods	90
4.5	Summary	92

An important consideration for any digital image whether taken indoor or outside, it is differences in brightness or colourfulness that have a strong and direct influence to the visual stimulus with which we are presented. This difference, named contrast (presented on chapter 2), represents an important piece in the visual process of the HVS. It allows a viewer to asses the main difference between present surfaces, to detect the object boundaries, to interpret depth relationships and finally, to understand the scene representation. In this chapter we introduce a fast method to classify outdoor images, mainly to detect haze into an image. Image dehazing has been represented an important research topic in the last decade. However, to the best of our knowledge, automatic haze detection has yet not been considered. We demonstrate that it is possible to detect haze into an image based on an per-pixel operation between the image and its inverse image or the complementary image(introduced



Figure 4.1: *Standard techniques limitations*. From left to right: initial foggy images, histogram equalization, local contrast stretching and our restored result.

in chapter 2.10). Finally, the quality of the results is determined by contrast evaluation.

4.1 Introduction

In outdoor environments, light reflected from object surfaces is commonly scattered due to the impurities of the aerosol, or the presence of atmospheric phenomena such as fog and haze. Aside from scattering, the absorption coefficient presents another important factor that attenuates the reflected light of distant objects reaching the camera lens. As a result, images taken in bad weather conditions (or similarly, underwater and aerial photographs) are characterized by poor contrast, lower saturation and additional noise.

Image processing applications commonly assume a relatively transparent transmission medium, unaltered by the atmospheric conditions. Outdoor vision applications such as surveillance systems, intelligent vehicles, satellite imaging, or outdoor object recognition systems need optimal visibility conditions in order to detect and process extracted features in a reliable fashion. Since haze degradation effects depend on the distance, as disclosed by previous studies (Fattal 08; Tan 08) and observed as well in our experiments (see Fig. 4.1), standard contrast enhancement filters such as histogram stretching and equalization, linear mapping, or gamma correction are limited to perform the required task introducing halos artifacts and distorting the color.

The contrast degradation of a hazy image is both multiplicative and additive. Practically, the haze effect is described by two unknown components: the *airlight* contribution and the *direct attenuation* related to the surface radiance. The color ambiguity of the radiance is due to the additive airlight, which

increases exponentially with the distance. Enhancing the visibility of such images is not a trivial task, as it poses an inherently under-constrained problem. A reliable restoration requires an accurate estimation of both the true colors of the scene and the transmission map, closely related to the depth-map.

Recently, there has been an increased interest in the vision and graphics communities in dehazing single images (Fattal 08; Tan 08; He 09; Tarel 09; Kratz 09). In this chapter we present an alternative approach to solving this challenging problem. Our technique is based on the remark that the distance from the observer to the scene objects is highly correlated with the contrast degradation and the fading of the object colors. More specifically, by an extensive study it has been disclosed an important difference between hazy and non-hazy image regions, by performing a per pixel comparison of the hue values in the original image to their values in a 'semi-inversed' image. This 'semi-inversed' image version is obtained by replacing the RGB values of each pixel on a per channel basis by the maximum of the initial channel value (r , g or b) and its inverse ($1 - r$, $1 - g$ or $1 - b$), followed by an image-wide renormalization. This observation has been validated on a large set of images, and allows for the detection of the hazy image regions by applying only a single simple operator. This facilitates the estimation of the airlight constant color, and enables us to compute a good approximation of the haze-free image using a layer-based approach.

Our single image dehazing strategy is characterized by three main contributions. First of all, we are the first to introduce a single image algorithm for the automatic detection of hazy regions. Second, our approach works on a per pixel basis. This makes it suitable for parallelization, and allows us to retain sharp detail near edges. Finally, our layer-based fusion dehazing strategy yields comparative and even better restored results than the existing approaches but performs faster and is suitable for real-time applications.

The remainder of this chapter is organized as follows. We will first review the most recent and influential methods, followed by a brief description of the optical model. Then, we will illustrate our approach to the automatic detection of haze within an image. Building upon this, we elaborate on our strategy of estimating the transmission and airlight. Before concluding the chapter, we present several comparative results together with an evaluation of the existing single image dehazing techniques, using the recent objective image metric of Aydin et al. (Aydin 08).

4.2 Related Work

An important area of application for dehazing algorithms can be found in multi-spectral remote sensing applications, where specialized sensors installed on satellites capture a specific band of the reflected light spectrum. Due to aerosol impurities and cloud obstruction, the recorded images require specific processing techniques (Chavez 88; Moro 06) to recover the original information. The method of Chavez (Chavez 88) searches based on the intensity distribution the value that corresponds to the darkest object. A similar principle has been used as a prior information in the recent method of He et al (He 09).

Many haze-removal techniques have used additional information in order to facilitate the search for a solution to this underconstrained problem. The existing dehazing methods can be grouped in several main classes.

Earlier haze-removal techniques are based on multiple images or supplemental equipment. Such methods (Narasimhan 00; Narasimhan 03a) that employ several input images of the same scene taken under various atmospheric conditions have shown to significantly improve the visibility in the restored image. However, in general scenarios these strategies are not practical since only a single degraded image is available as an input.

Another class of methods are the polarization techniques (Treibitz 09), (Shwartz 06), (Namer 09). These strategies exploit the fact that airlight is partially polarized. By taking the difference of two images of the same scene under different polarization angles, it becomes possible to estimate the magnitude of the polarized haze light component. A strategy with two different angles of the mounted polarization filter (when taken photographs of the same scene) has been used in the recent techniques of (Shwartz 06; Namer 09). Although the results are quite accurate, in situations with dense haze in which the polarization light is not the major degradation factor the results may be less robust. As well these methods seem to have some limitations when dealing with dynamic scenes.

Another category of techniques assume a known model of the scene. Narasimhan and Nayar (Narasimhan 03b) employ an approximated depth-map obtained after collecting information from several users about areas that are degraded or not by the poor weather conditions. *Deep Photo* (Kopf 08) is a more precise system since it uses the existing georeferenced digital terrain and urban models to restore foggy images. An iterative registration method aligns the 3D models with the outdoor images to provide the depth-map information used in the restoration process. Obviously, these methods suffer when no information of the scene geometry is available to the user.

On the other hand, single image resaturation is more challenging. Re-

cently, several single image based methods (Fattal 08; Tan 08; He 09; Tarel 09; Kratz 09) have been introduced. The method of Fattal (Fattal 08) uses a graphical model that solves the ambiguity of airlight color based upon the assumption that image shading and scene transmission are locally uncorrelated. The approach of (Kratz 09) proposed a related method with (Fattal 08) solution. This method models the image with a factorial MRF (Markov random fields) and computes the albedo and depth independently like two statistically independent latent layers. In Tan’s approach (Tan 08), the restoration aims to maximize the local contrast. He et al. (He 09) employ the dark channel image prior, based on statistical observation of haze-free outdoor images, in order to generate a rough estimation of the transmission map. Subsequently, due to the fact that they approximate the scene using patches of a fixed size, a matting strategy is required in order to extrapolate the value into unknown regions, and refine the depth-map. Tarel and Hautiere (Tarel 09) introduced a contrast-based enhancing approach to remove the haze effects, aimed at being faster than the previous approaches.

In this chapter we propose an alternative single image dehazing technique which is able to detect with a decent accuracy the spoiled hazy regions. Our technique has the advantage to be a per-pixel strategy that works relatively fast with no additional postprocessing steps (such as alpha-mating used in He et al. (He 09)) required.

4.3 Optical Model

The optical model used in this thesis is similar to the one employed in previous single image dehazing methods (Fattal 08; Tan 08; He 09; Tarel 09), initially described by Koschmieder (Koschmieder 24). For the sake of completeness, we provide a brief description of this model in this section.

When examining an outdoor scene from an elevated position, features gradually appear lighter and fading as they are closer towards the horizon. Only a percentage of the reflected light reaches the observer as a result of the absorption in the atmosphere. Furthermore, this light gets mixed with the *airlight* (Koschmieder 24) color vector, and due to the scattering effects the scene color is shifted (illustrated in Fig. 4.2). Based on this observation, the captured image of a hazy scene \mathcal{I}_h is represented by a linear combination of *direct attenuation* \mathcal{D} and *airlight* \mathcal{A} contributions:

$$\mathcal{I}_h = \mathcal{D} + \mathcal{A} = \mathcal{I} * t(x) + A_\infty * (1 - t(x)) \quad (4.1)$$

where \mathcal{I}_h is the image degraded by haze, \mathcal{I} is the scene radiance or haze-

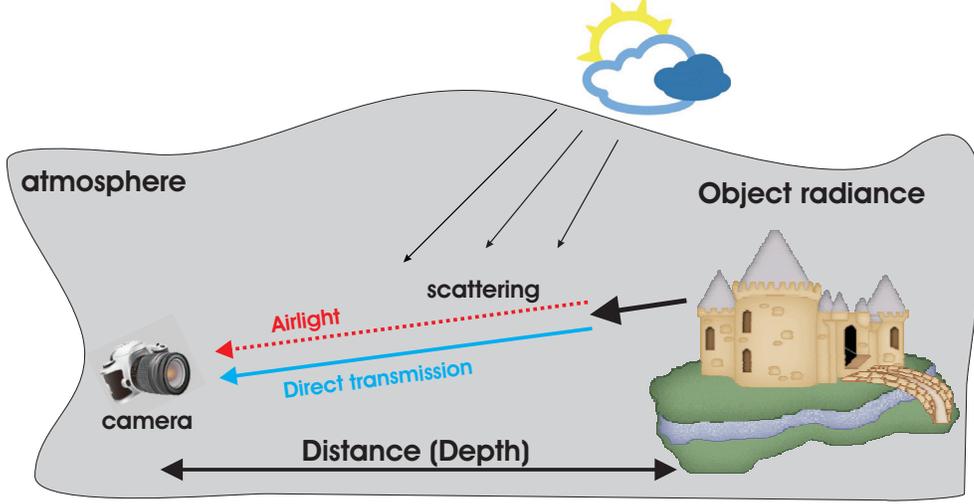


Figure 4.2: The employed optical model.

free image, A_∞ is the constant airlight color vector and t is the transmission along the cone of vision. This problem is clearly ill-posed, and requires us to recover the unknowns \mathcal{I} , A_∞ and $t(x)$ from only a single input image \mathcal{I}_h . In a homogeneous atmosphere, the transmission t is considered to be modulated as:

$$t(x) = \exp(-\beta * d(x)) \quad (4.2)$$

where β is the attenuation coefficient of the atmosphere due to the scattering and d represents the distance to the observer.

From equation 4.1, it becomes apparent that the chrominance attenuation becomes increasingly influenced by the airlight, as the optical depth increases:

$$\frac{\mathcal{A}}{\mathcal{D}} = \frac{A_\infty * (1 - t(x))}{\mathcal{I} * t(x)} \quad (4.3)$$

Theoretically, if the transmission and the airlight are known, the haze-free image can be easily computed:

$$\mathcal{I} = A_\infty - (A_\infty - \mathcal{I}_h) / t(x) \quad (4.4)$$

4.4 Single Image Dehazing by Per-Pixel Haze Detection

In this section is presented our novel single image dehazing technique. First, we introduce our haze detection strategy. Next, we describe how the airlight

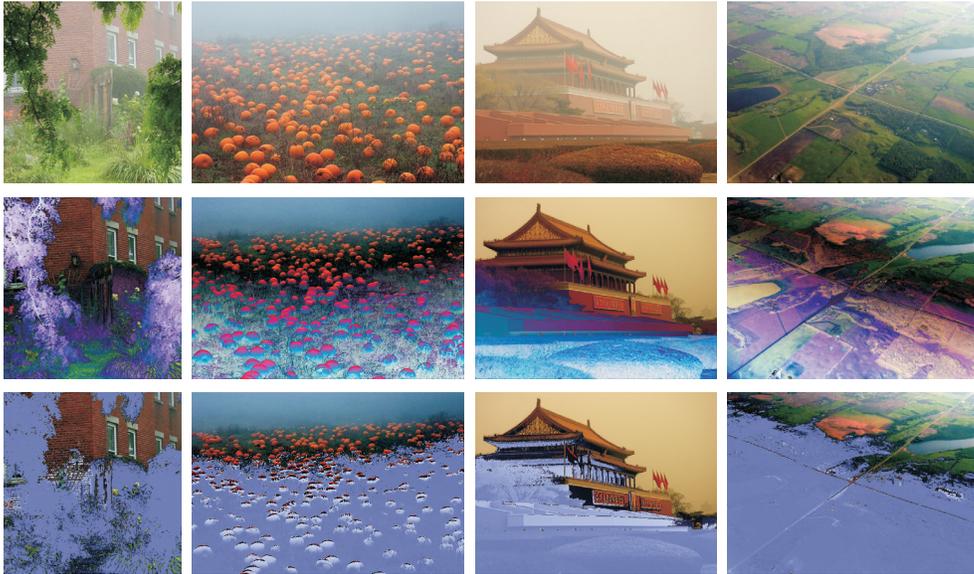


Figure 4.3: *Haze detection*. The first row shows the original hazy images \mathcal{I} . In the second row, we show the yielded semi-inversed image \mathcal{I}_{si} . Finally, in the third row, we label the pixels identified as not under the influence of haze with a blue mask. In these regions, the intensity of the blue color is proportional with the hue disparity.

constant color is estimated and as well how the 'semi-inversed' image version enables us to compute a good approximation of the haze-free image using a layer-based approach.

4.4.1 Haze Detection

The dark object method (Chavez 88) is a well-known technique within the remote sensing community. This technique is employed to remove haze from homogeneous scenes by subtracting an estimated value that corresponds to the darkest object of the image. More recently, He et al. (He 09) have presented a new derivation of this approach, called the dark channel strategy. A disadvantage of this new method is its inability to properly preserve edges, which is caused mainly by the employed erosion filter during the stage of computing the dark channel. In order to recover the refined transmission map and the latent image, this patch-based approach requires a complex postprocessing stage. By employing the dark channel prior (He 09), it has been shown that

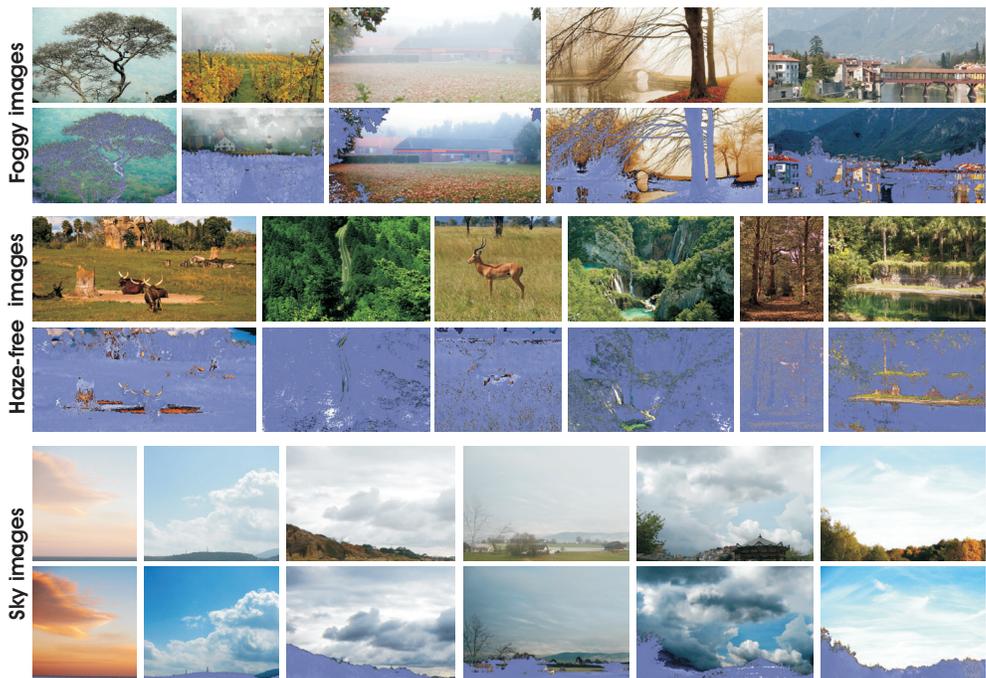


Figure 4.4: Results from applying our haze detection procedure on a large data set of images. Overall, haze-free images contain 96% pixels labeled as haze-free (masked in blue), while hazy and sky images are characterised by a significant decrease in haze-free pixels (less than 13%).

each patch of a natural image contains at least one point that is *dark* for non-sky or haze-free regions. The validity of this observation is mainly motivated by the fact that natural images are colorful and full of shadows (He 09)).

In this work we introduce a novel *per pixel* method that aims at generalizing the previous dark-channel approach. During our experiments, in which we analyzed a large set of natural images degraded by haze, we have observed that in haze-free and non-sky images, pixels in the neighborhood of dark pixels have a low intensity value in at least one color channel (r , g or b). On the dark channel, patches representing sky and hazy regions contain high values, as the local minimal intensity of such patches is high. Similarly, it has been observed that pixels in sky or hazy areas have high values in all color channels. These observations confirm the assumption that values in hazy image patches vary smoothly, except at depth discontinuities.

Based on these observations, we introduce a direct haze detection algo-

rithm that operates in a pixel-wise manner. We create a *semi-inversed* image $\mathcal{I}_{si}(x) = [\mathcal{I}_{si}^r, \mathcal{I}_{si}^g, \mathcal{I}_{si}^b]$. This image can be obtained by replacing the RGB values of each pixel x on a per channel basis by the maximum of the initial channel value and its inverse:

$$\begin{aligned}\mathcal{I}_{si}^r(x) &= \max_{x \in \mathcal{I}} [\mathcal{I}^r(x), 1 - \mathcal{I}^r(x)] \\ \mathcal{I}_{si}^g(x) &= \max_{x \in \mathcal{I}} [\mathcal{I}^g(x), 1 - \mathcal{I}^g(x)] \\ \mathcal{I}_{si}^b(x) &= \max_{x \in \mathcal{I}} [\mathcal{I}^b(x), 1 - \mathcal{I}^b(x)]\end{aligned}\tag{4.5}$$

where $\mathcal{I}^r(x)$, $\mathcal{I}^g(x)$ and $\mathcal{I}^b(x)$ represent the *RGB* channels of a considered image pixel x . Because the operations performed in equation 4.5 map the range of all pixels of the semi-inversed image \mathcal{I}_{si} onto the range $[0.5, 1]$, renormalization is required.

The reason of hue disparity is due to the image characteristics that have been previously described. In haze-free areas at least one-channel is characterized by small values. The operation will replace that value with its inverse. In regions of sky or haze since all values are characterized by high values, the max operation will return the same values. Therefore, by this direct hue comparison of the *semi-inverse* with the original image version, we are able to find pixels that need to be restored while conserving a similar color appearance with the original one.

As illustrated in Figure 4.3, this simple operation produces a semi-inversed image \mathcal{I}_{si} in which hazy areas are rendered with enhanced contrast, while the unaltered areas appear as the inverse of the initial image. To identify the regions affected by haze, we compute the difference between the hue channels of the original image \mathcal{I} and \mathcal{I}_{si} , and threshold it using a predefined value τ . The value of τ facilitates the selection of those pixels that present similar aspect in both the initial and the semi-inverse version. We have generated the results with the default value $\tau = 10^\circ$. Only pixels that have a hue disparity below this threshold τ are labeled as hazy pixels. In our approach the hue information is represented by the h^* channel after the image is transformed into the perceptual CIE $L^*c^*h^*$ color space.

By applying this simple strategy, we are able to estimate the hazy regions with acceptable precision. In order to check the validity of our observation, we collected a large database of natural images from several accessible photo sites (e.g. *Flickr.com*, *Picasaweb.com*, *Photosig.com*). All the selected images have been taken in daily light conditions. We defined three main categories of outdoor images: haze-free images without sky, sky images, and hazy images. After manually selecting 800+ images for each of these classes, we evaluated

the variation of the hue using the strategy previously described. The main conclusion is that the haze-free images are characterized by a vast majority of pixels affected by significant hue variations, while in the other two categories this variation is considerably less. We illustrate this in Figure 4.4. In order to differentiate between the latter categories, it is possible to detect sky regions using existing techniques (Tao 09). In the results and discussion section 4.4.4 we present a comparison of our haze detection component with the dark channel method of He et al. (He 09).

4.4.2 Airlight Color (A_∞) Estimation

One important correlation for dehazing algorithms constitutes the relation between optical depth and airlight (R.C.Henry 00; Treibitz 09). The airlight \mathcal{A} becomes more dominant as the optical depth increases. The optical model (equation 4.1) reveals the fact that two objects with different reflectance properties, located at the same distance from the observer, have identical airlight offsets. Consequently, when observing the values of \mathcal{A} in a small area around a scene point, they usually show only minor variations, except when depth discontinuities occur. Moreover, the A_∞ constant can be acquired from the areas with the highest additive contribution, which are commonly the areas of the image characterized by high intensity.

These properties of the hazy images have been exploited as well in the previous approaches to estimate the airlight constant A_∞ . As observed by Narasimhan and Nayar (Narasimhan 03a), this constant is best estimated in the most haze-opaque areas. He et al. (He 09) choose the 0.1% brightest pixels of the dark channel as their preferred region. Another approach (Tan 08) is to search for this component in regions with the highest intensity, assuming that the sky is present and that there are no saturated pixels in the image.

The key advantage of our approach is that we are able to clearly identify hazy regions. As explained in previous section, these regions are identified in a straightforward manner by observing the hue disparity between the image and its semi-inverse. In order to mask the most haze-opaque areas, we perform the same procedure, but with the intensity of the semi-inverse increased by a factor ξ (with a default value of $\xi = 0.3$). During our experiments, we found that for images where the sky is present, the resulted mask contains mostly the sky region, which decreases the searching space. The extraction of the airlight color vector A_∞ is performed by determining the brightest pixel only in the positive (unmasked) region (see Figure 4.3). The winning value of A_∞ is extracted from the original foggy image from the same location as the brightest pixel. This approach has shown to be more robust than only

searching for the most brightest pixels in the entire image.

4.4.3 Layer-based Dehazing

The contrast within a dehazed image is directly correlated with the estimated airlight \mathcal{A} and the inferred transmission $t(x)$ at each point. Due to the physical additivity of the airlight (see equation 4.1), it is possible to estimate the direct attenuation \mathcal{D} once the A_∞ is known, by varying $\mathcal{A} = A_\infty * (1 - t(x))$ in all possible values of its range (Tan 08). Previous approaches have introduced many constraints and cost functions that favor certain image characteristics based on local image patches, thus limiting this range, and making it possible to compute an approximate transmission map. In previous strategies, the transmission map is commonly refined further using an energy minimization approach, based on the assumption that for local neighborhoods the airlight shows only very minor deviations (an assumption that breaks down at depth discontinuities). The main drawback of such approaches is that the employed search methods, even though they are commonly very expensive, are unable to ensure an accurate transmission map. In contrast, we present a fast method which segments the image in regions that do not contain abrupt depth discontinuities. Our strategy was inspired by the approach of Narasimhan and Nayar (Narasimhan 03a), which aims to normalize the pixels in so-called *iso-depth* neighborhoods. However, the previous method requires two pictures in order to identify such regions for their normalization operation. When using multiple images, it becomes possible to identify such iso-depth neighborhoods, as they are invariant to the weather conditions and do not contain sudden depth discontinuities (Narasimhan 03a).

There are many possible strategies to creating a dehazed image after a per pixel identification of hazy regions. In this work, we propose a layer-based method, which aims to preserve a maximum amount of detail, while still retaining sufficient speed. We initiate our algorithm by creating several new images \mathcal{I}_i , with $i \in [1, k]$ and k layers, in which we remove a decreasingly growing portion of the airlight constant color A_∞ from the initial hazy image \mathcal{I} :

$$\mathcal{I}_i = \mathcal{I} - c_i \cdot A_\infty. \quad (4.6)$$

with the iteratively increasing airlight contribution factor c_i .

After applying our haze detection operation on \mathcal{I}_i , only the pixels with a sufficiently low hue disparity are labeled as being part of layer \mathcal{L}_i . In the absence of the scene geometry, discretization of the image in k distinct layers enables us to estimate the values of c_i that correspond to the most dominant depth layers of the scene. For instance, when the scene contains two objects

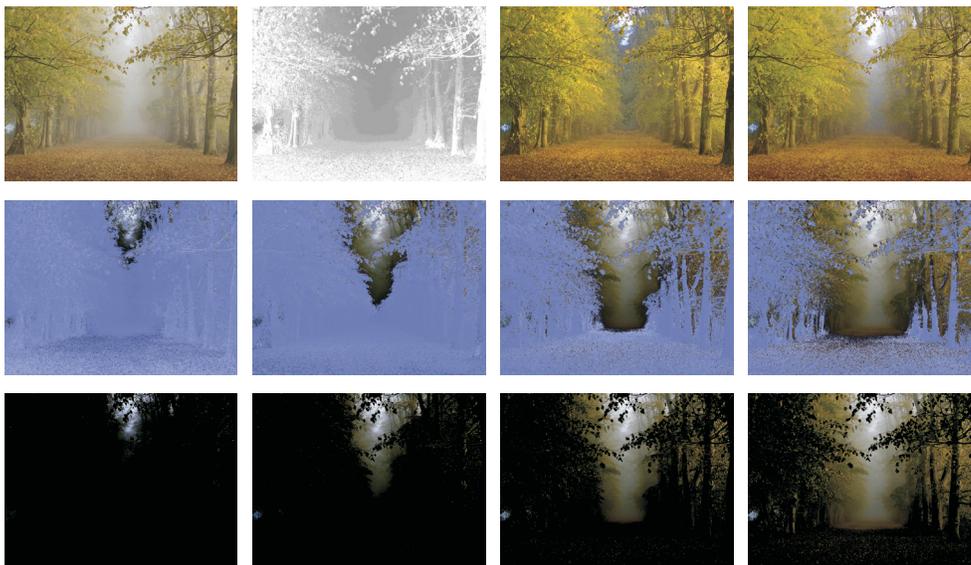


Figure 4.5: *Layer-based dehazing*. **Top line:** the initial foggy image; the rough transmission map that corresponds to \mathcal{I}_0 ; the result of a naive method which simply pastes all layers \mathcal{L}_i upon each other, introducing artifacts; the result of our method, which applies soft blending of the layers. **Middle line:** the mask regions for each layer. **Bottom line:** the computed layers \mathcal{L}_i .

located at different depths, the transmission map will be characterized by two dominant values, as the airlight is correlated with the distance. Finally, these layers are blended into a single composite haze-free image. In order to smooth the transitions between the different layers, the number of extracted layers k needs to be at least 5 or more. As can be observed in Figure 4.5 every layer (except the first one) includes the pixels of the previous layer, but with different levels of attenuation. To obtain the haze-free image \mathcal{I}_0 , the layers are blended in the descending order of the airlight contribution. A naive approach would consist of simply copying the pixels from each layer on the next, but this might generate unpleasing artifacts due to small discontinuities (see Figure 4.5). In order to remove such undesirable transitions artifacts, each layer will contribute a small percentage onto the next layer, according to the following equation:

$$\mathcal{I}_0 = \sum_{i=1}^k \chi_i \mathcal{L}_i. \quad (4.7)$$



Figure 4.6: *Haze removal results* From left to right: initial foggy images, dehazed result using dark channel (He 09), our naive dehazed result using *semi-inverse*, our final refined result and comparative refined result using bilateral filter.

where χ_i weights the contribution of the layers pixels, increasing exponentially according with the layer number. Practically, this is performed by employing alpha channels(or masks) for each layer. To smooth the transition, the algorithm smooths the masks by applying a Gaussian. Splitting the input image into non-uniform neighborhoods that contains approximately uniform airlight generates good results, even when using only a single image. In comparison, algorithms that are based upon fixed-size patches may introduce artifacts because these uniform patches do not consider the intensity distribution of the small region. Moreover, our straightforward pixel based strategy is computationally effective overcoming the existing single image dehazing approaches (see the next section for comparative processing times).

The solution proposed is not unique, and other standard available methods can be used to smooth the transitions and minimize the artifacts. In figure 4.6 are demonstrated the difference between the naive result obtained by applying dark channel and also our naive result using our *semi-inverse* approach. Can be observed that our approach introduces less visible visual artifacts except some transitions. This is the main reason that our refinement and therefore the dehaze method is more efficient than the one proposed by He et al. (He 09). On the right side of the figure is shown the dehazed result obtained after applying the edge preserving bilateral filter.

4.4.4 Experimental Results and Discussion

As we have stated in the introduction, we believe to be the first to present an algorithm for per pixel haze detection. It could be stated that the algorithm of He et al. (He 09), an extension of the dark object technique common in the remote sensing community (Chavez 88), could also be regarded as a method for

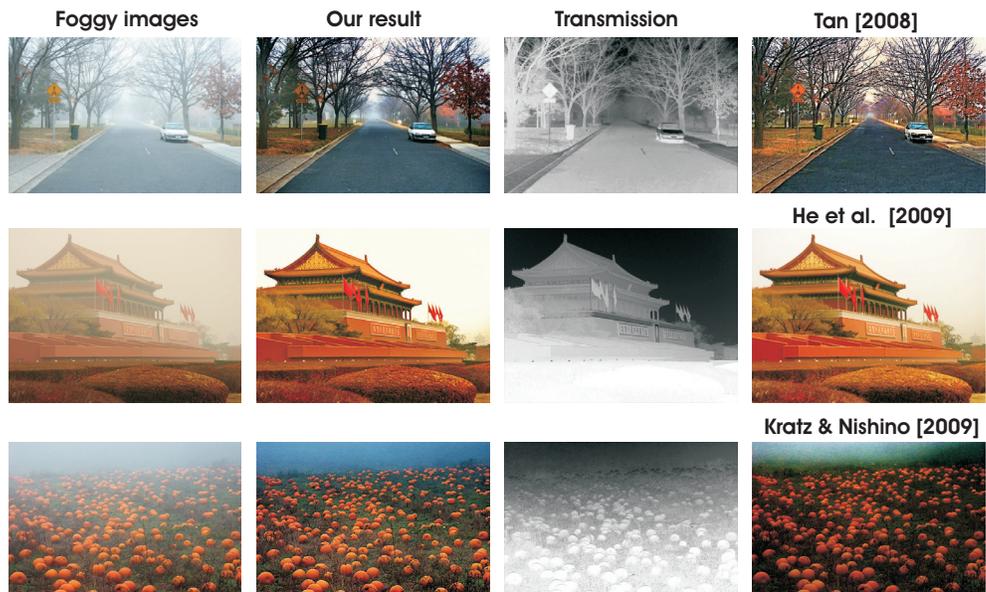


Figure 4.7: *Comparative haze removal results.* From left to right: initial foggy images, our restored results, our estimated transmissions (depth map) and the corresponding results obtained by Tan (Tan 08) , He et al. (He 09) and Kratz and Nishino (Kratz 09) respectively.

haze area detection. However, in order to create the dark channel, their technique employs a patch-based approach which is unable to properly preserve fine detail. When taking a closer look at Figure 4.8, the patch-like structure of the dark channel images immediately becomes apparent. The black regions associated with haze-free areas do not reflect the true haze-free area borders. It is important to note that these transitions need to be recovered properly, as edges between two regions characterized by large illumination differences can generate prominent halo artifacts. This is the result of using patches with a constant dimension, which do not consider the intensity distribution of the small region. In contrast, our method employs a pixel-wise multi-layer strategy, which decomposes the image in regions that are characterized by small illumination variations. This approach does not suffer from halo artifacts, because the regions are non-uniform and respect the image intensity distribution.

We have tested our approach on a large data set of natural hazy images. Figure 4.7 illustrates results (our dehazed image and the computed transmis-

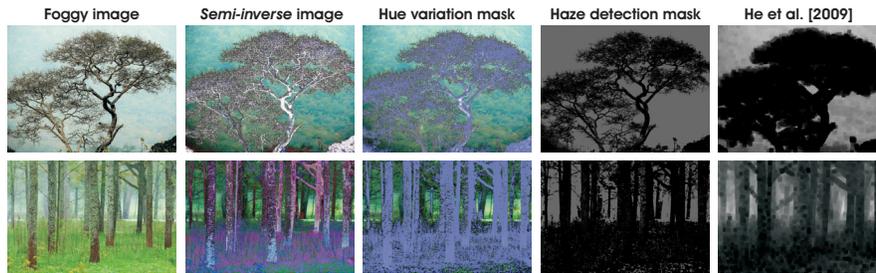


Figure 4.8: Our technique is able to identify hazy regions on a per pixel basis. Comparing the haze mask produced by our technique to the dark channel mask of He et al. (He 09) clearly shows that we are able to preserve significantly more detail.

sion map) obtained for three foggy images by our technique, compared to the methods of Tan (Tan 08), He et al. (He 09) and Kratz and Nishino (Kratz 09). All the figures presented in this thesis contain the original restored images provided by the authors. As can be observed, we are able to enhance the images while retaining even very fine details. Furthermore, our method accurately preserves the color of the objects in the scene. Another set of images will be provided in Figure 4.12, where the top lines show comparative results obtained by the techniques of Fattal (Fattal 08), He et al. (He 09), Tarel and Hautiere (Tarel 09) and our method. It should be noted that in the discussion above, we have limited ourselves to images in which the scene is sufficiently illuminated. Even though the method performs generally well, for poorly lit scenes (an extreme case of such problem), like previous single image dehazing techniques, our approach is limited to accurately detect hazy regions. This limitation is due to the hypothesis of the considered optical model that assumes that regions characterized by small intensity variations contain no depth discontinuities. As can be seen in the example presented in Figure 4.9 in the absence of significant intensity variations of distant regions our approach may have trouble to accurately restore images.

As previously mentioned our approach has the advantage to perform faster than related algorithms. Our method implemented in Matlab processes an 600×800 image in approximately 4-5 seconds.



Figure 4.9: **Limitation of our approach.** In the absence of significant intensity variations of distant regions our approach may have trouble to accurately restore images. Considering the initial foggy image (a) our method using only 5 layers yields the result shown in the middle (b). In spite of increasing the degree of visibility, our algorithm may introduce several artifacts. By increasing the number of layers our method (c) amplifies these artifacts (see the yellow box region). This is due to the general assumption that regions characterized by small intensity variations contain no depth discontinuities.



Figure 4.10: Additional restored images by our technique.

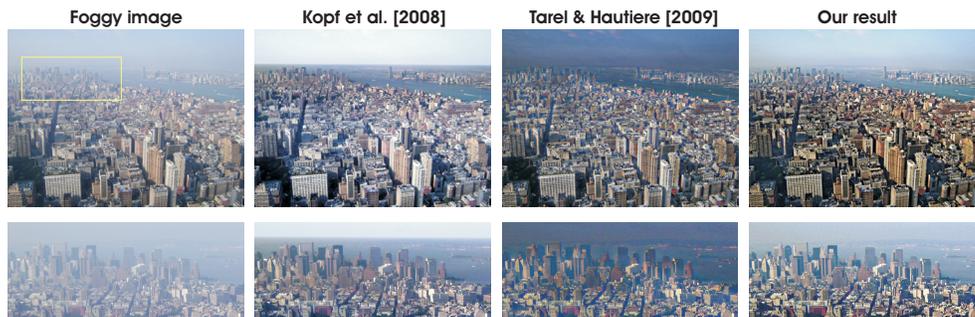


Figure 4.11: From left to right: initial foggy image, the result obtained by *Deep Photo* (Kopf 08) that employs additionally an approximated 3D model of the scene, the result of Tarel and Hautiere (Tarel 09) and our result.

The method of Tan (Tan 08) requires more than 5 minutes per image while the technique of Fattal (Fattal 08) computes an 512×512 image in 35 seconds. The algorithm of He et al. (He 09) takes approximately 20 seconds per image while the computation times of the techniques of Kratz and Nishino (Kratz 09) were not reported.

4.4.5 IQA Assesment of Dehazing Methods

Since there is no specialized evaluation procedure of the dehazing techniques we searched the recent literature for an appropriate method for this task. Tarel and Hautiere (Tarel 09) evaluate the quality of dehazing techniques based on a visibility resaturation procedure (Hautiere 08). Because this procedure only applies to grayscale images and is mainly focused on finding the most visible edges, we searched for a more general method that is able to perform a pixelwise evaluation of the dehazing process.

In this work, we have employed the *Image Quality Assessment (IQA)* quality measure introduced recently by Aydin et al. (Aydin 08). The IQA metric is sensitive to three types of structural changes: *loss of visible contrast* (**green**), *amplification of invisible contrast* (**blue**) and *reversal of visible contrast* (**red**). As a general interpretation (please refer to chapter 2 for more details about this metric), contrast loss (*green*) has been related with image blurring, while contrast amplification (*blue*) and reversal (*red*) have been connected to image sharpening.

Since these modifications are closely related to our problem, we have found this measure to be more appropriate as a means to evaluate the resaturation of hazy regions after applying different dehazing techniques.

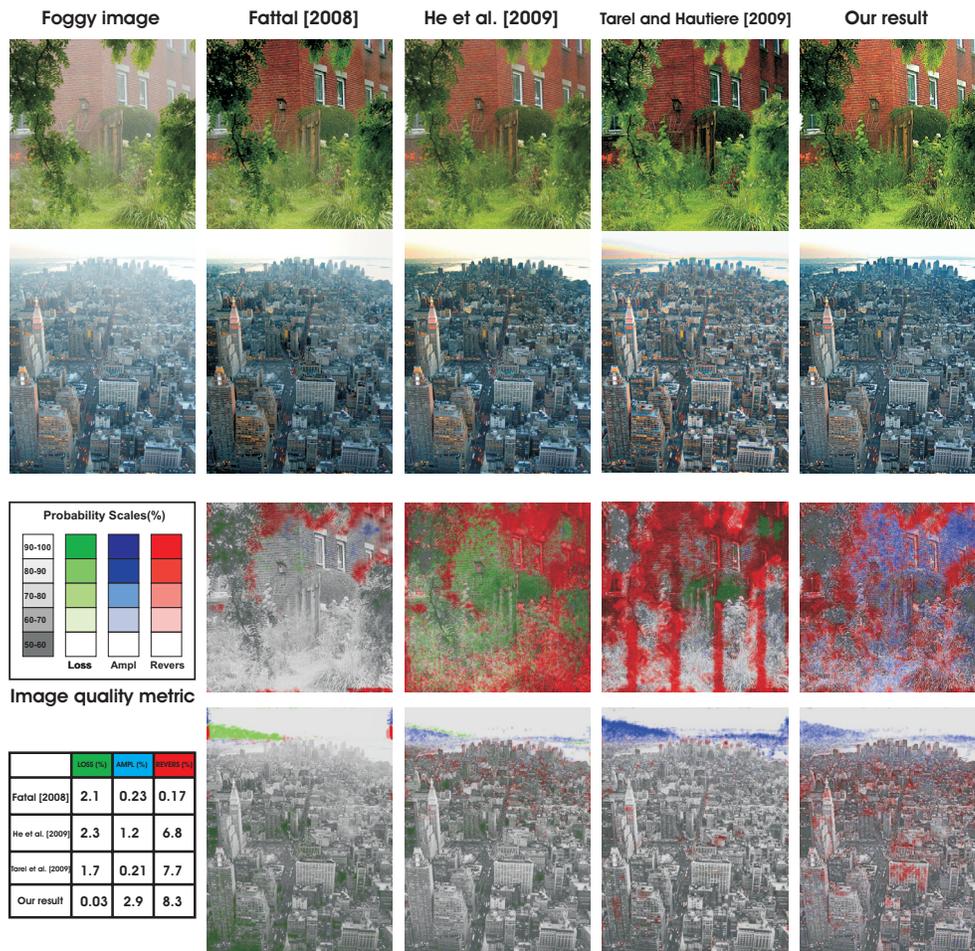


Figure 4.12: *Evaluation of the results using the IQA metric.* The top two lines present the comparative results obtained by Fattal (Fattal 08), He et al. (He 09), Tarel and Hautiere (Tarel 09) and our method. The bottom two lines show the results after applying the IQA between initial image and the restored version. The left bottom table presents the ratio of the color pixels (with a probability scale higher than 70%) counted for each method.

Figure 4.12 shows the comparative results of applying the IQA metric on two foggy images and their dehazed versions, using the method of Fattal (Fattal 08), He et al. (He 09), Tarel and Hautiere (Tarel 09) and ours. The bottom-left table of Figure 4.12 displays the comparative ratios of the (colored) pixels yielded by the IQA measure when applied to the results of

the considered dehazing methods. Following the recommendation of Aydin et al. (Aydin 08), in order to reduce the possibility of misclassification, only the pixels with a probability scale higher than 70% have been considered. Based on the results from the table, it becomes clear that compared with the other techniques, the structural changes yielded by our algorithm are more closely related to sharpening operations (**blue** and **red** pixels) and less related with blurring (**green** pixels).

4.5 Summary

In this chapter we have focused to the single image dehazing problem. We have presented a single-image dehazing strategy which does not make use of any additional information (e.g. images, hardware, or available depth information). Our approach is conceptually straightforward. Based on a per pixel hue disparity between the observed image and its semi-inverse, we are able to identify the hazy regions of the image. After we have identified these regions, we are able to produce a haze-free image using a layer-based approach. The limitations of this technique are inherited from the dark object, meaning that when the color information is absent the technique may fail. The processing time of our optimized technique is very low when compared to previous methods who were designed and optimized for speed. As future work we plan to investigate the potential of our method for video-dehazing. Additionally, we intend overcome the noticed limitations of our algorithm for the more complex case of non-homogeneous haze images. A possible application of the haze-detector is to develop the technique to be suitable for sky detection. Existing methods are still limited to recover in many cases the images degraded by haze because the hazy regions are often characterized by noise. To increase the quality of the results, we propose to combine the image dehazing with other techniques such as image denoising and superresolution techniques. As also mentioned in (He 09) the optical model would greatly benefit by further investigations. We plan to address this problem and also we intend to extend the visibility enhancement problem for other research areas such as underwater images.

Chapter 5

Enhancing by Fusion-based Techniques

Contents

5.1	Introduction	94
5.2	Image Decolorization by Fusion	96
5.2.1	Overview	96
5.2.2	Our Algorithm	97
5.2.3	$H - K$ Chromatic Adapted Lightness	98
5.2.4	Weight Maps Assignment	99
5.2.5	Multi-scale Fusion of the Inputs	101
5.2.6	Results and Discussion	104
5.2.7	Video Decolorization	105
5.2.8	Evaluation of Grayscale Operators	105
5.3	Fusion-based Single Image Dehazing	111
5.3.1	Overview	111
5.3.2	Definition of the Inputs	112
5.3.3	Weight Maps	113
5.3.4	Fusion Process	114
5.3.5	Results and Discussion	116
5.4	Summary	119

Two or more images can be combined into a single image, that in some way can be more suitable for some intended purposed. *Image fusion* is a well-known technique that is designed to maximize the amount of relevant information into fused image. The most relevant example for this application is tone-mapping by combining multiple multiple exposures (Mertens 07). In this chapter we use the fusion technique for single image manipulation, mainly for image decolorization and single image dehazing. We show that this technique can be adapted to solve these two different problems. In the preceding chapters we have presented that contrast plays a key role in image decolorization and single image deahzing. We show that fusion algorithm can be engineer also based on various contrast measurements, such that fusion result incorporate the relevant information. For the best of our knowledge this is the first work that employ fusion for decolorization purpose. Hyperspectral-image fusion is similar with other fusion techniques such as exposure-fusion (Mertens 09) and depth-of-field extension (Burt 83). We show, in particular, that fusion can be used as an effective mode for single image dehazing, even in the absence of additional information. We opted for multiresolution fusion-approach because it proves effective at avoiding seems.

5.1 Introduction

Image fusion is a fundamental technique that blends data from multiple sources and has been successfully applied in different fields such as remote sensing, medical imaging, microscopic imaging, robotics, enhancing and restoration in image processing. A great deal of effort has been made in recent years to develop efficient image fusion methods for specific applications. All-in-focus imagery is obtained by combining images captured when the camera focus is varied (Haeberli 94). In multi-spectral imagery (Socolinsky 02) fusion has been applied by merging satellite images obtained as an output of different sensors and wavelengths. As shown by Raskar et al. (Raskar 03), context enhancement could be achieved by blending nighttime and daytime images into day-night imagery. Their basic idea (Raskar 04) was to increase the information density in a set of low quality images by exploiting the context from a higher quality image captured under different conditions from the same view. Blending images has shown utility for image editing (Perez 03) and interactive photo montage (Agarwala 04). Image fusion has been applied as well for image compositing (Brinkmann 99; Grundland 06). Grundland et al. (Grundland 06) adapted the classical linear interpolation in order to preserve the original saliency and global contrast of the input images. Fusing images with multiple exposures has been employed for tone-mapped depiction

of HDR images (Burt 93; Mertens 07; Mertens 09).

In this chapter we present two novel fusion-based techniques. Firstly, we present a fusion-based strategy to decolorize images accurately. The algorithm employs the three independent *RGB* channels and an additional image that conserves the color contrast, based on Helmholtz-Kohlrausch effect, as image inputs. This fourth input better preserves the global appearance of the image, as it enforces a more consistent gray-shades ordering. Our algorithm uses Three weights based on three different forms of local contrast: a saliency weight map (which assesses the saliency of the input); a second weight map that advantages well-exposed regions; and a chromatic weight map (which increases color contrast in addition to the effect of $H - K$ (Helmholtz-Kohlrausch) input). This approach is designed in a multi-scale fashion, using a Laplacian pyramid representation of the inputs combined with Gaussian pyramids of normalized weights. This solution is preferred as it is capable to reduce the possible artifacts that may appear due to the weight maps.

To the best of our knowledge we are the first that introduce a fusion-based decolorization technique. Our method performs faster than existing color-to-gray methods since it does not employ color quantization (Rasche 05), that tends to introduce artifacts, or cost function optimization, which commonly is computationally expensive (e.g. Gooch et al. (Gooch 05) approach) and risks not converging to a global extremum. The new operator has been tested on a large dataset of both natural and synthetic images. In addition, we demonstrate that our operator is able to decolorize videos. Our multi-scale fusion approach demonstrates consistency over varying palettes, and is able to maintain temporal coherence within videos. Furthermore, we have performed a comparative evaluation of the contrast enhancement qualities of the recent state-of-the-art color-to-grayscale techniques.

Secondly, we introduce a novel single image strategy that is able to accurately dehaze images using only the original degraded information. Our approach is based as well on a fusion strategy that takes two inputs derived from the original image. These inputs are weighted by three normalized weight maps and finally blended in a multi-scale fashion that avoids introducing artifacts. The method is fast being straightforward to implement and shows to outperform the related operators in the contrast-based evaluation where IQA measure (Aydin 08) has been used. Several advantages advocate our technique against the previous single image dehazing methods. First, our approach performs an effective per-pixel computation, different than the majority of the previous methods (Fattal 08; Tan 08; He 09) that consider patches. A proper per-pixel strategy reduces the amount of artifacts since patch-based methods have some limitations due to the assumption of a constant airlight in every

patch. In general this assumption is not true and therefore additional post processing is required (e.g. the method of He et al. (He 09) needs to smooth the transmission map by alpha-matting). Secondly, the complexity of our approach is more reduced than the previous strategies. Finally, our technique performs faster being suitable for real-time applications.

Even though we used the same principle as the most of the previous fusion methods, as will be presented in the next two sections, our approaches present several distinctive features that allow to compress (the decolorization approach) the three-dimensional color image in a grayscale version that preserves the original contrast and fine details and as well to restore accurately the hazy regions.

5.2 Image Decolorization by Fusion

5.2.1 Overview

Discussed as well in chapter 3, although color plays an important role in images, applications such as compression, visualization of medical imaging, aesthetical stylization, and printings require reliable decolorized image versions. The widely-used standard color-to-grayscale conversion employs the luminance channel only, disregarding the important loss of color information.

In this section we present a novel decolorization method, built on the principle of image fusion. The main difference between fusion methods, that makes them application-specific, is the choice of inputs and weights. Our algorithm employs the three independent *RGB* channels and an additional image that conserves the color contrast, based on Helmholtz-Kohlrausch effect, as image inputs. This fourth input better preserves the global appearance of the image, as it enforces a more consistent gray-shades ordering. The weights used by our algorithm are based on three different forms of local contrast: (a) a saliency weight map which aims to preserve the saliency of the original color image; (b) a second weight map that advantages well-exposed regions; and (c) a chromatic weight map which enhances the color contrast in addition to the effect of $H - K$ input. In order to minimize artifacts introduced by the weight maps, our approach is designed in a multi-scale fashion, using a Laplacian pyramid representation of the inputs combined with Gaussian pyramids of normalized weights.

Our fusion-based decolorization algorithm, the way of choosing the weights and the inputs and as well how we blend them is detailed in the next subsections.

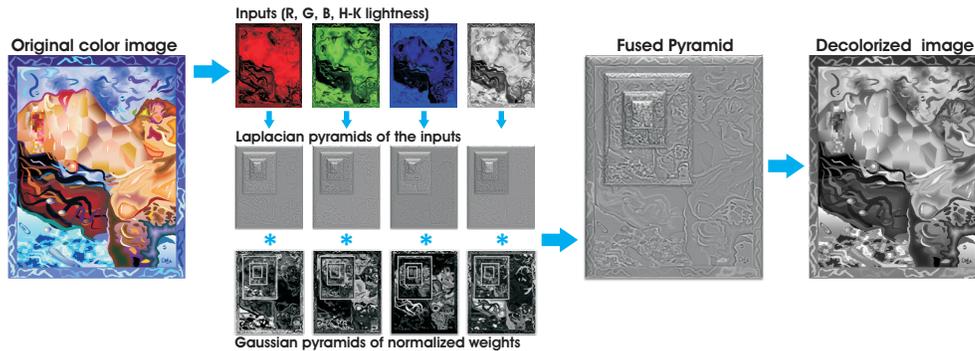


Figure 5.1: An overview of our fusion-based approach. Based on the original input image, we derive four input images (R , G , B and $H - K$ lightness) and three weight maps that blended by a multi-scale image fusion strategy yields the decolorized output.

5.2.2 Our Algorithm

The standard grayscale transformation tends to reduce the amount of variations and sharpness within an image. Qualitatively, the dull appearance is due to the loss of contrast that is more visually noticeable on dimmed highlights and shadows. In order to obtain pleasing decolorized images, photographers might compensate these limitations by tedious work in the darkroom, applying elaborate lighting techniques or using photo-editor programs to manually adjust the contrast, luminance or histogram distribution.

We argue that the image appearance in black-and-white is tightly connected with models of color appearance, and that measurable values like salient features and color contrast are difficult to integrate by simple per pixel blending, without introducing artifacts into the image structure. For this reason, we have opted for the multi-scale approach of image fusion, combining the Helmholtz-Kohlrausch lightness predictor (Fairchild 91) with a set of pixel weights depending on important image qualities. This will ensure that regions with superior gain are well depicted in the decolorized image. Practically, the resulted grayscale image is obtained by fusing four input images (a lightness image that incorporates the $H - K$ effect, and the R , G , B color channels), weighted by normalized coefficients maps determined by saliency, pixel exposure, and chromatic weights. An overview of our approach is given in Figure 5.1.

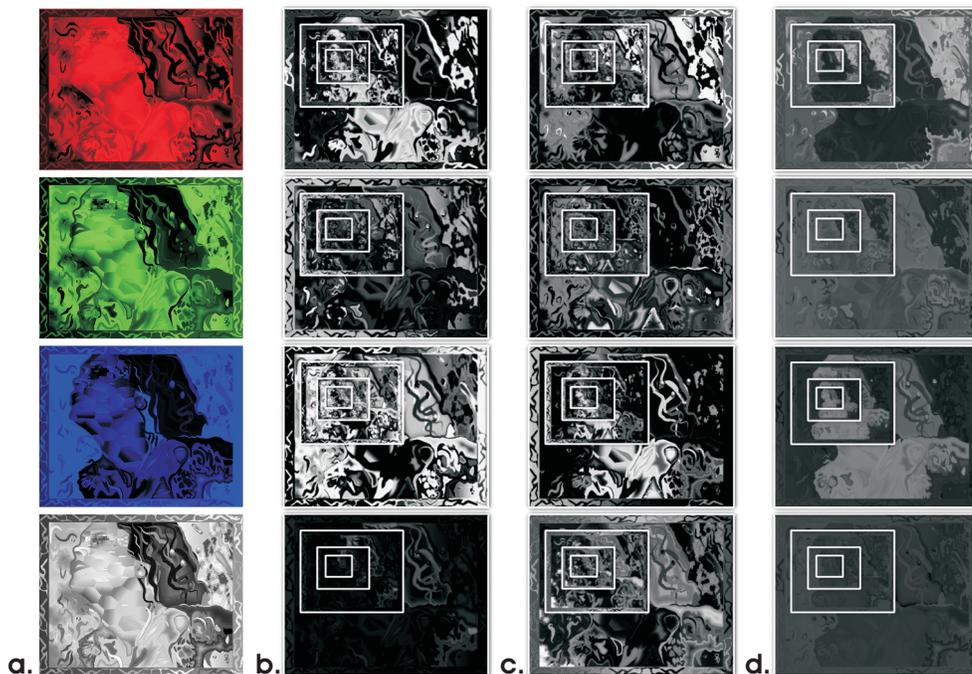


Figure 5.2: a. The four image fusion inputs (R , G , B and $H - K$ Lightness) and the corresponding Gaussian pyramids of the image fusion weights (b. saliency; c. exposedness ; d. chromatic).

5.2.3 $H - K$ Chromatic Adapted Lightness

Our algorithm requires four input images to be used in the fusion process. Besides the color channels R , G , B , we define an additional input that preserves the global contrast, based on Helmholtz-Kohlrausch effect. As observed by Smith et al. (Smith 08), the $H - K$ effect can be used to resolve potential ambiguities regarding the difference between the isoluminant colors. Therefore, given two isoluminant patches, the most colorful one will be mapped onto a brighter output intensity. For this fusion input channel, we used Fairchild's chromatic lightness metric (Fairchild 91), which predicts the $H - K$ effect, defined in the $CIEL^*c^*h^*$ color space by the expression:

$$L_{H-K} = L^* + (2.5 - 0.025L^*)(0.116 \left| \sin \left(\frac{h^* - 90}{2} \right) \right| + 0.085)c^* \quad (5.1)$$

This L_{H-K} predictor has also been used in the work of Smith et al. (Smith 08), in which it was demonstrated to be more appropriate for the task of image decolorization than the chromatic lightness metric of Nayatani (Nayatani 98).

The Nayatani (Nayatani 98) predictor often tends to map bright colors to white, which makes it harder to discriminate between images that contain bright isoluminant colors. However, as can be observed in the comparative results (see the image but as well video results of Smith et al. (Smith 08)), relying only to the Helmholtz-Kohlrausch effect the decolorized outputs might not preserve accurately the original saliency. This feature is mostly ensured in our operator by integrating the saliency weight map.

5.2.4 Weight Maps Assignment

In the following section, we present how the weight maps are defined in our fusion-based decolorization algorithm. Our approach is based on the principle that the output decolorized image needs be both visually pleasing and meet the application requirements: In the case of grayscale conversion, aside from the luminance which is the main contributor to the perceived lightness, there are also several other image qualities that guide our visual system during its analysis of the incoming light. Practically, the attention of an observer tends to be focused on the salient regions that stand-out within their neighborhood. In order to maintain this focus, it is desirable that these predominant regions are well preserved by the grayscale version. Therefore, in order to meet this requirement, we first introduce a *saliency weight map*. Furthermore, as commonly the over- and underexposed regions are advantaged by the saliency map, we also define an *exposedness weight map* that overcomes perception degradation in these regions. Finally, we assign a third weight map, the *chromatic weight map*, which has the main goal of balancing the influence of chromatic stimuli into the perception of lightness. Practically, by smoothly fusing the input channels weighted by these weight maps, the original consistency of the image is well preserved, while ghosting and haloing artifacts are reduced. Moreover, we believe that these weight maps are intuitive concepts for the users.

The maps are explained in more detail in the following paragraph:

Saliency weight map (\mathcal{W}_S) reveals the degree of conspicuousness with respect to the neighborhood regions. For this measurement, our algorithm employs the recent saliency algorithm of Achanta et al. (Achanta 09). Their strategy is inspired by the biological concept of center-surround contrast. The saliency weight at pixel position (x, y) of input I^k is defined as:

$$\mathcal{W}_S(x, y) = \left\| I_\mu^k - I_{\omega_{hc}}^k \right\| \quad (5.2)$$

where I_μ^k represents the arithmetic mean pixel value of the input I^k while $I_{\omega_{hc}}^k$ is the blurred version of the same input that aims to remove high frequency

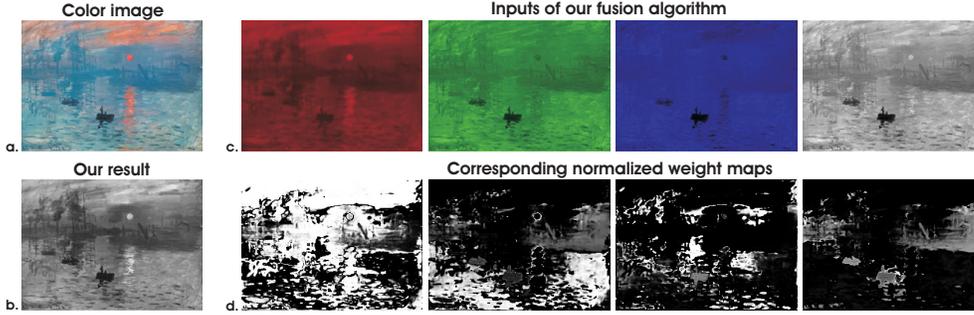


Figure 5.3: From color to gray: from the original color image (a), we obtain our decolorized result (b) by applying an image fusion approach, using the four inputs (c), weighted by the corresponding normalized weight maps (d).

noise and textures. $I_{\omega_{hc}}^k$ is obtained by employing a small 5×5 ($\frac{1}{16}[1, 4, 6, 4, 1]$) separable binomial kernel with the high frequency cut-off value $\omega_{hc} = \pi/2.75$. For small kernels the binomial kernel is a good approximation of its Gaussian counterpart, but it can be computed more effectively. The approach of Achanta et al. (Achanta 09) is very fast, and has the additional advantage of the extracted maps being characterized by well-defined boundaries and uniformly highlighted salient regions, even at high resolution scales. Based on extensive experiments, we found that this saliency map tends to favor highlighted areas. In order to increase the accuracy of results, we introduce the exposedness map to protect the mid tones that might be altered in some specific cases.

Exposedness weight map (\mathcal{W}_E) estimates the degree to which a pixel is exposed. The function of this weight map is to maintain a constant appearance of the local contrast, neither exaggerated nor understated. Practically, this weight avoids an over- or underexposed look by constraining the result to match the average luminance. Pixels are commonly better exposed when they have normalized values close to the average value of 0.5. Inspired by the approach of Mertens et al. (Mertens 09), who employ a similar weight in the context of tone mapping, the exposedness weight map is expressed as a Gaussian-modeled distance to the average normalized range value (0.5):

$$\mathcal{W}_E(x, y) = \exp\left(-\frac{(I^k(x, y) - 0.5)^2}{2\sigma^2}\right) \quad (5.3)$$

where $I^k(x, y)$ represents the value of the pixel location (x, y) of the input

image I^k , while the standard deviation is set to $\sigma = 0.25$. This mapping conserves those tones that are characterized by distance close to zero, while larger distance values are related with the over- and underexposed regions. As a result, the impact of over- and underexposed regions filtered by the saliency map is tempered, keeping the original image appearance well preserved.

Chromatic weight map (\mathcal{W}_C) controls the saturation contribution of the inputs in the decolorized image. This is expressed as the *standard deviation* between every input and the saturation S (in *HSL* color space) of the original image. Due to the fact that in general humans prefer increased saturation, it is desirable that more saturated areas are mapped onto brighter tones. This balances the chromatic contrast loss with the desired amount of enhancement. We have observed that the impact of this gain is reduced for the $H - K$ chromatic adapted lightness input.

In our framework these weight maps (saliency, exposedness and chroma) have the same contribution to the resulted decolorized images. As an example, Figure 5.2 shows the computed weights for the considered inputs.

5.2.5 Multi-scale Fusion of the Inputs

Having defined the inputs (R , G , B color channels and $H - K$ chromatic adapted lightness) and the weight maps, in the following section we present how this information is blended by our fusion strategy. As previously mentioned, during the fusion process the inputs are weighted by specific maps in order to conserve the most significant features, and finally combined into a single output image:

$$\mathcal{F}(x, y) = \sum_{k=1}^K \bar{\mathcal{W}}^k(x, y) \mathcal{I}^k(x, y) \quad (5.4)$$

where the value of every pixel location (x, y) of the fused result \mathcal{F} is obtained by taking the sum of the corresponding locations of the inputs \mathcal{I}^k (k is the input index), weighted by the normalized weight maps $\bar{\mathcal{W}}^k$. The number of the inputs is counted by the index k (in our case $K = 4$). The normalized weights $\bar{\mathcal{W}}$ are obtained by normalizing over the M weight maps \mathcal{W} ($M = 3$) in order that the value of each pixel (x, y) weights to sum up to unity ($\sum \mathcal{W}^k = 1$ for each pixel location) (see Figure 5.3).

Unfortunately, applying eq. 5.4 directly sometimes introduces haloing artifacts, mainly in locations close to strong transitions between weight maps. In order to solve this problem, a more effective strategy needs to be devised.

Generally, this task is solved by multi-scale decomposition strategies that use linear (Burt 83; Rahman 97) or non-linear filters (Durand 02; Farbman 08; Subr 09). While the class of non-linear filters has shown to be better at preserving edges, the linear filters are computationally more effective. Even though more refined multi-scale solutions might be applied as well, we have opted for the classical multi-scale Laplacian pyramid decomposition (Burt 83). In this linear decomposition, every input image is represented as a sum of patterns computed at different scales based on the Laplacian operator. The inputs are convolved by a Gaussian kernel, yielding a low pass filtered versions of the original. In order to control the cut-off frequency, the standard deviation is increased monotonically. To obtain the different levels of the pyramid, initially we need to compute the difference between the original image and the low pass filtered image. From then on, the process is iterated by computing the difference between two adjacent levels of the Gaussian pyramid. The resulting representation, the Laplacian pyramid, is a set of quasi-bandpass versions of the image.

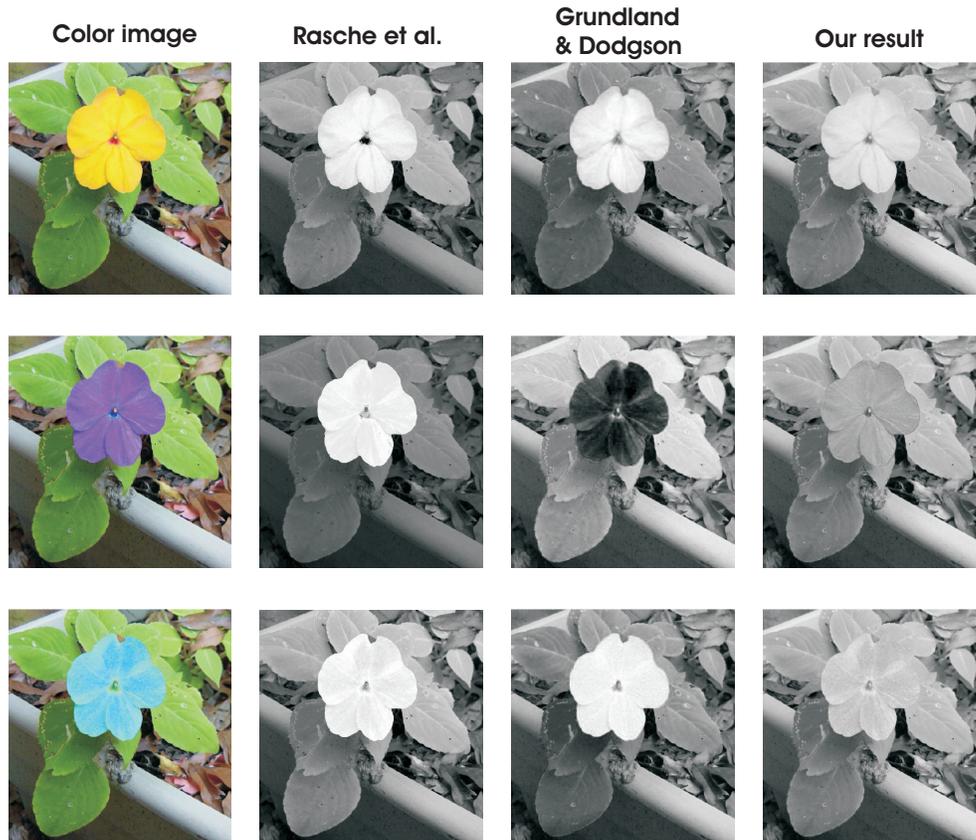


Figure 5.4: Coherence of color-to-gray methods. Note how differently the methods map the background (e.g. leaves) and the flower. Compared with the methods of (Rasche 05; Grundland 07) our operator maps into the same grayscale level the leaves while the flower is converted into different grayscale levels.

In our case, each input is decomposed into a pyramid by applying the Laplacian operator to different scales. Similarly, for each normalized weight map \bar{W} a Gaussian pyramid is computed. Considering that both the Gaussian and Laplacian pyramids have the same number of levels, the mixing between the Laplacian inputs and Gaussian normalized weights is performed at each level independently yielding finally the fused pyramid:

$$\mathcal{F}^l(x, y) = \sum_{k=1}^K G^l \left\{ \bar{W}^k(x, y) \right\} L^l \left\{ \mathcal{I}^k(x, y) \right\} \quad (5.5)$$

where l represents the number of the pyramid levels (determined by the image dimensions), $L \{ \mathcal{I} \}$ is the Laplacian version of the input \mathcal{I} , and $G \{ \bar{W} \}$ represents the Gaussian version of the normalized weight map \bar{W} . This step is performed successively for each pyramid layer, in a bottom-up manner. Basically, this approach solves the *cut and paste* problem among the inputs with respect to the normalized masks. A similar adjustment of the Laplacian pyramid, but in the context of exposure fusion has been applied in other contexts (Mertens 09). The final decolorized image is obtained by summing the fused contribution of all inputs.

This linear multi-scale strategy performs relatively fast (takes approximately 1.4 seconds per image in our unoptimized MATLAB implementation) representing a good trade off between speed and accuracy. By employing independently a fusion process at every scale level the potential artifacts due to the sharp transitions of the weight maps are minimized. Multi-scale techniques are broadly used due to their efficiency in image compression, analysis and manipulation. This operation has the advantage that it respects the perceptual system of the human eye, which is known to be more sensitive to modifications into high frequencies than changes in low frequencies.

5.2.6 Results and Discussion

Our fusion-based approach addresses the preservation of several important image features: saliency, well-exposedness and chromatic contrast. One major benefit of fusing the inputs guided by weight maps is that this principle allows for a direct transfer of the important characteristics of the color image to the decolorized version. We believe that strong perceptual similarity between colorized and decolorized images can be obtained by algorithms that consider both global and local impressions. In our approach, the global appearance of the image is preserved by imposing a gray-shades order that respect the $H - K$ color appearance model. The weight maps contribute to the local preservation of the original relations between neighbor patches. A similar idea has been

experimented with, using Poisson solvers (Fattal 02) in a related approach of Gooch et al. (Gooch 05). However, their approach performs poorly for images with extended disconnected regions that represent isoluminant features. The main reason for this is that the Poisson solver ignores differences in gradients over distances larger than one pixel. Our fusion technique proves that by employing well defined quality measures and inputs, consistent results can be produced even for these difficult cases.

The new operator has been tested extensively for a large set of images. Figure 5.6 presents several comparative results against recent grayscale operators.

5.2.7 Video Decolorization

Video decolorization adds another dimension to the problem of image decolorization, as temporal coherence needs to be guaranteed for the entire video sequence. In order to speak of consistency, an algorithm has to map similar regions from the color input onto similar areas in the decolorized output. Recently Smith et al. (Smith 08) have shown that local approaches are suitable for this task. As our strategy retains both global and local characteristics, it is able to maintain consistency over varying palettes (see Figure 5.4), yielding temporal coherence for videos (see as well Figure 5.5). Figure 5.4 shows several versions of the same image, in which the flower is colored differently at each instance. Global pallet mapping techniques like the one of Grundland and Dodgson (Grundland 07) generate dissimilar gray levels for the same region on different instances (note the leaves and the background mapping). Our operator and the method of Smith et al. (Smith 08) yield more consistent outputs.

A similar limitation can be observed by analyzing Figure 5.5 that displays the frames of a synthetically generated footage with isoluminant color patches. By a close inspection it might be seen that even though the technique of Grundland and Dodgson (Grundland 07) decolorizes each frame perceptually accurate, this technique is not able to preserve the same grayscale level corresponding to the same color patch along the entire sequence of the frames. On the other hand, the method of Smith et al. (Smith 08) introduces some non-homogeneity artifacts along edges.

5.2.8 Evaluation of Grayscale Operators

In order to measure the quality of the conversions, we performed a contrast-based evaluation of the recent state-of-the-art operators. For this task we adapted the recent technique of Aydin et al. (Aydin 08), which is used to

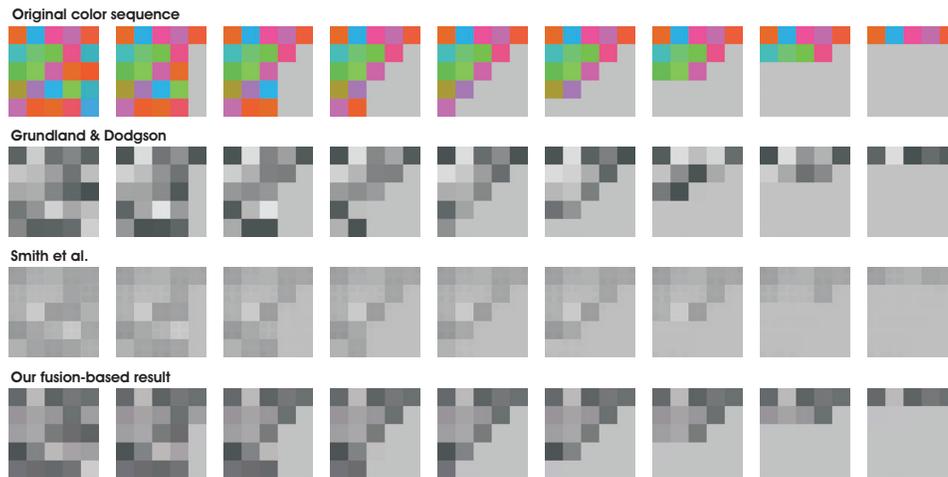


Figure 5.5: Isoluminant video test. For a well consistency the same color patch needs to be converted into a similar gray level in all images. Notice the artifacts introduced by Smith et al. (Smith 08) approach but also the different grayscale mapping of the same colored patch yielded by Grundland and Dodgson method (Grundland 07) (please refer to the supplementary material for the entire sequence).

compare a pair of images with significantly different dynamic ranges. Instead of detecting only contrast changes, this metric is sensitive to three types of structural changes: *loss of visible contrast* (**green**) - a contrast that was visible in the reference image becomes invisible in the transformed version, *amplification of invisible contrast* (**blue**) - a contrast that was invisible in the reference image becomes visible in the transformed version and *reversal of visible contrast* (**red**) - a contrast is visible in both images, but has different polarity. They observed that the contrast loss (green) is related with blurring, while contrast amplification (blue) and reversal (red) with sharpening. An online implementation¹ of this metric is made available by the authors.

We tested several grayscale operators for a set of 24 images that have also been used in the perceptual evaluation of Cadik (Cadik 08). Besides the *CIEY*, Bala and Eschbach (Bala 04), Gooch et al. (Gooch 05), Rasche et al. (Rasche 05), Grundland and Dodgson (Grundland 07), Coloroid (Neumann 07), Smith et al. (Smith 08) methods, we also reviewed the recent technique of Kim et al. (Kim 09) and our fusion-based decolorization operator. The measure of Aydin et al. (Aydin 08) is applied using the

¹http://www.mpi-inf.mpg.de/resources/hdr/vis_metric/

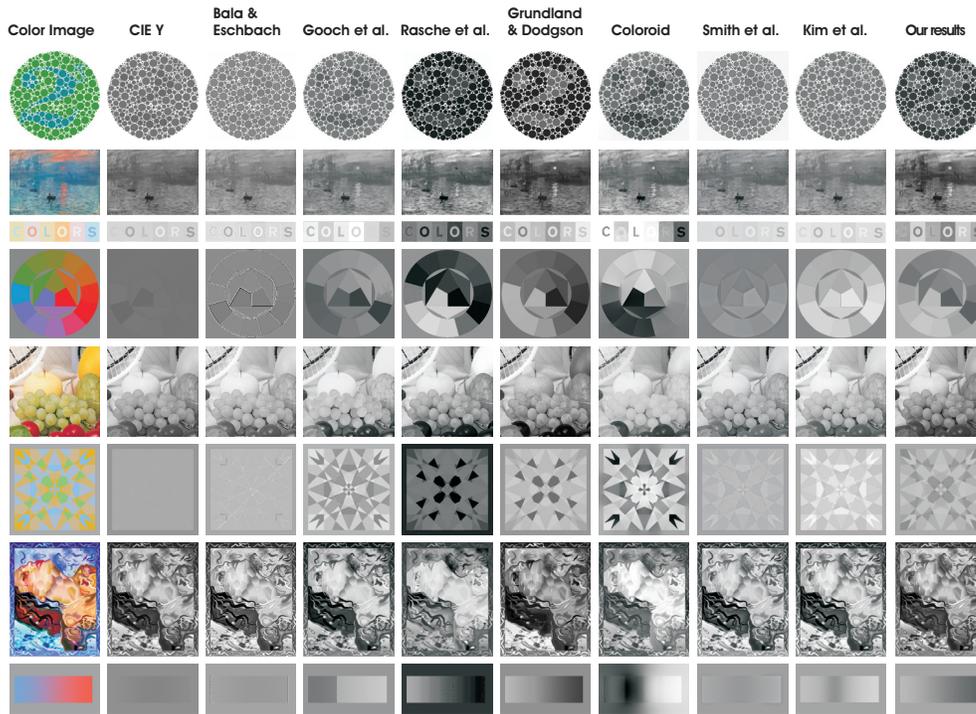


Figure 5.6: Comparative results. From left to right the grayscale results obtained by applying *CIEY*, Bala and Eschbach (Bala 04), Gooch et al. (Gooch 05), Rasche et al. (Rasche 05), Grundland and Dodgson (Grundland 07), Coloroid (Neumann 07), Smith et al. (Smith 08), Kim et al. (Kim 09) methods and our fusion-based operator.

default parameter set of the authors, and considering the original color image as a reference. The results of applying the *IQA* measure are shown in Figure 5.7. The graphics in the figure display the average ratio over the 24 *IQA* images of the pixels with the contrast changed after applying the corresponding transformation. Only the pixels with a probability higher than 70% have been counted.

Based on these graphics, we can observe that our operator, together with Smith et al. (Smith 08) and Kim et al. (Kim 09), shows the minimal amount of produced blurring artifacts (rendered with the *green* after applying *IQA*). Regarding the sharpening effects (*blue* and *red* pixels), in general all methods, except the one of Smith et al. (Smith 08) and *CIEY*, perform in a relative similar range of values.

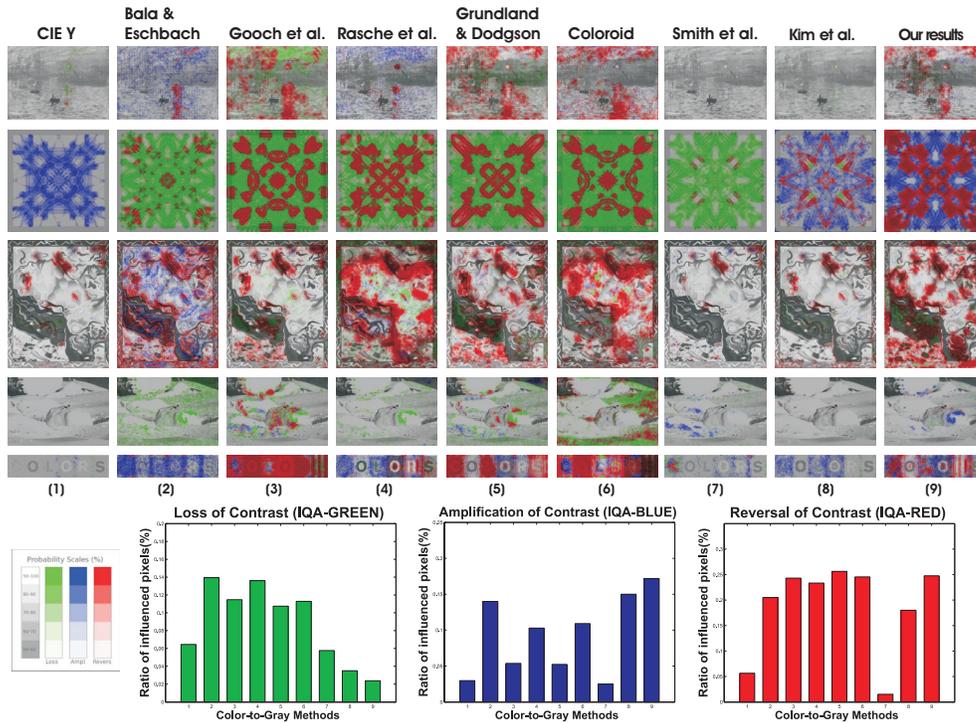


Figure 5.7: IQA evaluation of operators. In the top are shown the results obtained by applying the contrast-based measure of Aydin et al. (Aydin 08) between the original color and decolorized images. In the bottom part are displayed the graphics that plot the average ratio (over the complete set of images) of the pixels with the contrast changed after applying the corresponding transformation. *Green* is related with blurring while *blue* and *red* are related with sharpening.

Furthermore, we present as well a direct comparison between the fusion-based decolorization method presented in this chapter and our method described previously in the chapter 3 (see Figure 5.8).

In our extensive experiments, the fusion-based operator performed generally well. However, we observed that in some situation due to the inexact selection of the saliency map, our technique is unable to improve substantially the results yielded by standard conversion. In general, the chosen weights can yield proper results.

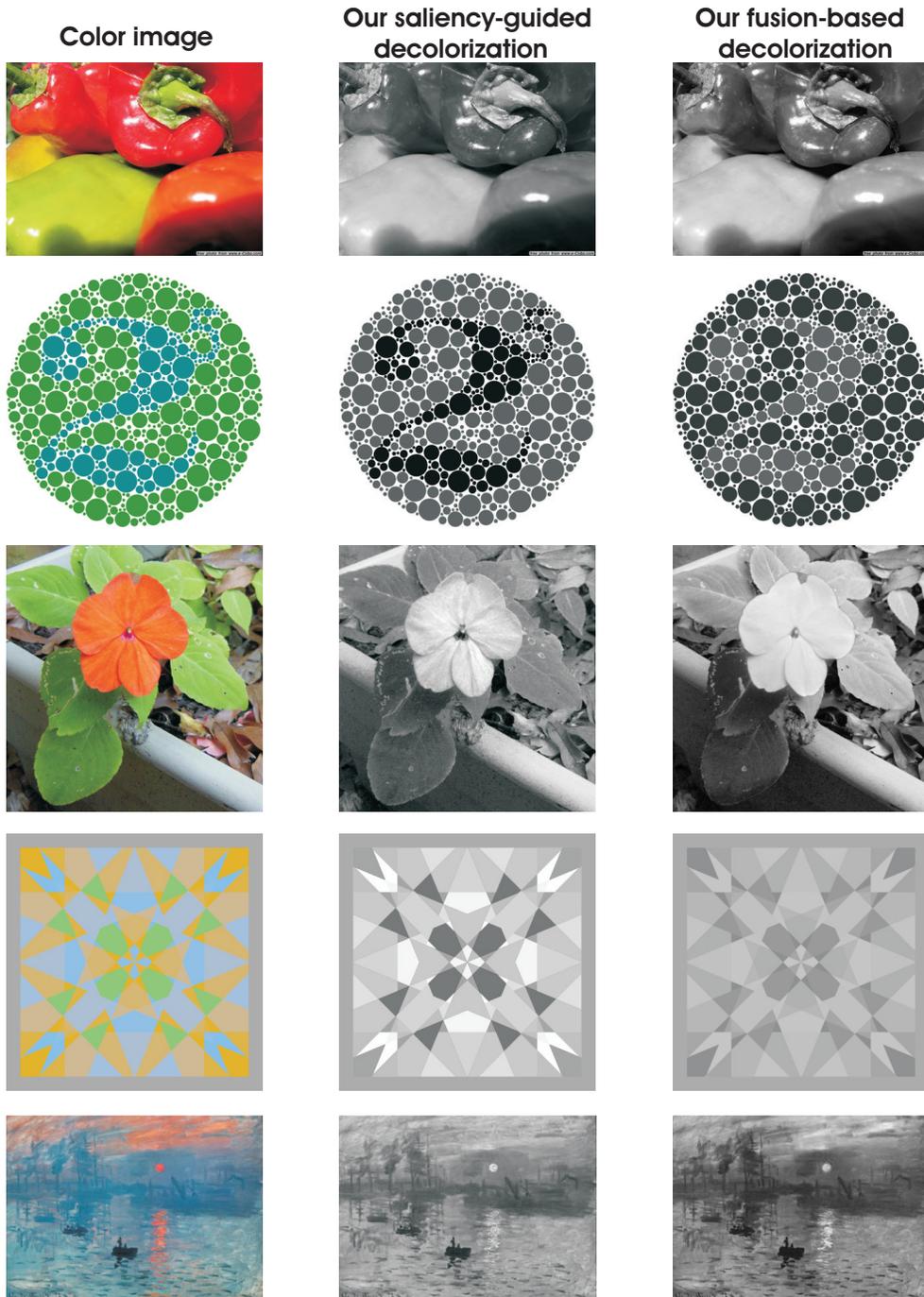


Figure 5.8: Comparison between our fusion-based color-to-grayscale method and the results yielded by our saliency-guided method presented in chapter 3.

As well, the algorithm shares a weight map related problem in common with other fusion algorithms. As noticed by previous methods (Mertens 09), the exposedness map may generate an artificial appearance of the image when its gain is exaggerated.

Our algorithm is computationally effective (our unoptimized implementation takes approximately 2.5 seconds for a 800x600 image), having a processing time comparable to recent CPU approaches (e.g. Smith et al. (Smith 08) method takes 6.7 - 10.8 seconds for 570x593 image, Decolorize (Grundland 07) -unoptimized code - 3.5 seconds for a 800x600 image and the (extremely) optimized code of Kim et al. (Kim 09) decolorizes a 800x600 image in 1-2 seconds). However, even relatively fast since it employs an effective nonlinear global mapping optimization, the method of Kim et al. (Kim 09) did not solve the rendering limitations of the related technique of Gooch et al (Gooch 05), tending to diminish the global contrast and to lose the original saliency. In addition, we believe that an optimized CPU implementation would make our operator suitable for real-time applications.

5.3 Fusion-based Single Image Dehazing

5.3.1 Overview

As discussed as well in the chapter 4, the single image dehazing is an ill-posed problem that can be solved by a regularized optimization that maximizes the contrast (Tan 08). He et al. (He 09) built their approach on the statistical observation of the dark channel (Chavez 88) that allows a rough estimation of the transmission map. Fattal (Fattal 08) assumed that image shading and scene transmission are locally uncorrelated and formalize them using a Gauss-Markov random field model. This approach has been generalized recently (Kratz 09) by a factorial Markov random field.

Since in general the previous methods are computationally burdensome we searched for a different solution that processes very fast with minimum loss in accuracy.

In this section we introduce a fusion techniques that employs only the inputs and weights derived from the original hazy image. Image fusion has a wide applicability (e.g. remote sensing, medical imaging, microscopic imaging, robotics) and the main idea is to combine several images into a single one, keeping only the most significant features of them. By choosing appropriate weight maps and inputs, our fusion-based method is able to effectively dehaze images. Our strategy bears some similarity with the recent methods of He et al. (He 09) and Tarel and Hautiere (Tarel 09).



Figure 5.9: Comparison with the fast method of Tarel and Hautiere (Tarel 09). Our method performs faster yielding as well more accurate results than (Tarel 09). Notice the sky and sea regions.

Both of this methods can be seen as filtering solutions since the dark channel can be related with an erosion problem while in (Tarel 09) employed their defined *median of median filter* in order to preserves both edges and corners. However, our approach is fundamentally different since it removes haze from

images by simply blending the inputs weighted by several maps. Our strategy combines the input information in a per-pixel fashion minimizing the loss of the image structure. Moreover, even though both approaches perform relatively fast, compared with the results of Tarel and Hautiere (Tarel 09), our strategy yields more pleasing results with low degree of artifacts (see Figure 5.9).

5.3.2 Definition of the Inputs

Our fusion method takes two inputs derived from the original image. The **first input** \mathcal{I}_1 is obtained by white balancing the original hazy image. The white balance step ensures the natural rendition of images by eliminating chromatic casts that are caused by the atmospheric color. Due to the fact that haze is dominating the image, an average value is computed for the entire image. Similar as in (Tan 08; Tarel 09) a straightforward biasing of the image average color toward pure white is employed. This step assures that atmospheric light color constant V_∞ is equal to one and the normalized image values are in the range $[0, 1]$. As well observed in (Tarel 09), when the light color varies in the image it is more robust to perform this bias operation using the local average value.

The **second input** \mathcal{I}_2 is selected in order to increase the contrast in hazy regions. In our approach this is obtained automatically by subtracting from the original image \mathcal{I} the average luminance value of the entire image $\bar{\mathcal{I}}$. This operation has the effect to amplify the visibility in regions degraded by haze but yielding some degradation in the rest of the image. A similar effect may be obtained by general contrast enhancing operators (e.g. gamma correction, histogram stretching) that also amplify the visibility in the hazy parts while destroying the details in the rest of the image. However, this degradation is solved by employing proper weight maps (please refer to the next subsection and Figure 5.10). Practically the second input is calculated by the expression:

$$\mathcal{I}_2 = \gamma (\mathcal{I} - \bar{\mathcal{I}}) \quad (5.6)$$

where γ is the factor that increases linearly the luminance in the hazy regions (default value is $\gamma = 2.5$). In general hazy images are too dark and therefore is preferred to increase the global luminance. The parameter γ has a similar impact as the tone mapping stage of (Tarel 09) applied on the the haze-free regions assumed to be in the bottom third part of the original image. Basically, the equation 5.6 first identifies the regions that have greater values compared to the average and afterward the values of detected range is redistributed onto the entire luminance range by multiplication with γ . The γ can also be computed as $\gamma = 2(0.5 + \bar{\mathcal{I}})$. Thanks to the haze dominance in the image, the

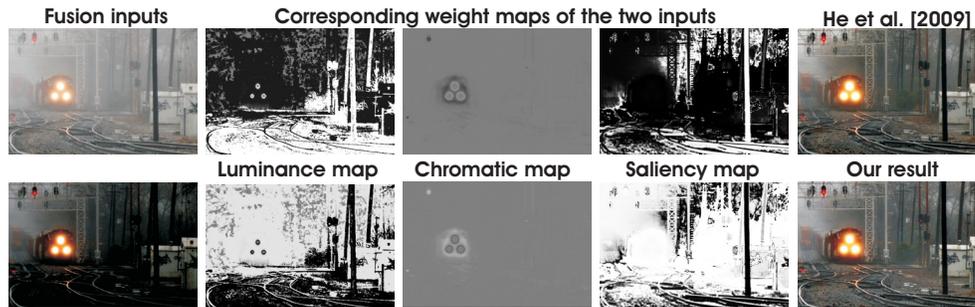


Figure 5.10: In the left side are shown the two inputs. In the middle are displayed the three weight maps corresponding to the inputs. Finally, the results of He et al. (He 09) and ours are shown.

regions with values higher than the mean, are identified as hazy. To improve the exposedness, this operation centers the values of the hazy regions around the original mean range value (0.5).

5.3.3 Weight Maps

Luminance weight map manages the luminance gain in the output image. This gain map computes at each pixel the standard deviation between every R, G and B color channel and luminance L of the input. This overcomes the degradation induced by \mathcal{I}_2 in the haze-free regions ensuring a seamless transition between the inputs $\mathcal{I}_1, \mathcal{I}_2$. This map also tends to reduce the global contrast and colorfulness. However, these effects are overcome by defining two additional weights: chromatic (colorfulness) and saliency (global contrast).

Chromatic weight map controls the saturation gain in the output image. To obtain this map, for every pixel is computed the distance between its saturation value S and the maximum of the saturation range using a Gauss curve: $d = \exp\left(-\frac{(S-S_{max})^2}{2\sigma^2}\right)$ with a standard deviation $\sigma = 0.3$. Thus, weights close to zero are assigned to the pixels with smaller saturation while the most saturated pixels have weights close to one. This weight map is motivated by the fact that in general humans prefer increased saturation, being desirable that more saturated areas to be better depicted in the haze free image.

Saliency weight map identifies the degree of conspicuousness with respect to the neighborhood regions. In our approach is used the recent saliency algorithm of Achanta et al. (Achanta 09) mainly because due to its computation-

ally efficiency but also due to the fact that the yielded map has well-defined boundaries and uniformly highlighted salient regions even at high resolution scales. The impact of this gain is to increase the global contrast appearance since it increases the contrast in highlighted and shadowed parts.

To yield consistent results, we employed the normalized weight values $\bar{\mathcal{W}}$, by constraining that the sum at each pixel location of the weight maps \mathcal{W} to equal one.

5.3.4 Fusion Process

In the fusion process, the inputs are weighted by specific computed maps in order to conserve the most significant detected features. Each pixel (i, j) of the output \mathcal{F} is computed by summing the inputs \mathcal{I}_k weighted by corresponding normalized weight maps $\bar{\mathcal{W}}_k$: $\mathcal{F}^{(i,j)} = \sum_k \bar{\mathcal{W}}_k^{(i,j)} \mathcal{I}_k^{(i,j)}$. The naive solution to implement directly this equation might introduce halos artifacts, mostly in the locations characterized by strong transitions of the weight maps. To prevent these image degradation problems, we opted for the adapted solution that employs a classical multi-scale pyramidal refinement strategy (Burt 83). We tested as well several more recent edge preserving techniques (e.g. WLS (Farbman 08)) but did not obtain significant improvement. However, these methods need in general to tweak the parameters and are more computationally expensive. Practically, in our case, each input is decomposed into a pyramid by applying Laplacian operator at different scales. Similarly, for each normalized weight map $\bar{\mathcal{W}}$ a Gaussian pyramid is computed. Considering that both the Gaussian and Laplacian pyramids have the same number of levels, the mixing between the Laplacian inputs and Gaussian normalized weights is performed at each level independently yielding the fused pyramid:

$$\mathcal{F}_l^{(i,j)} = \sum_k G_l \{ \bar{\mathcal{W}}_k^{(i,j)} \} L_l \{ \mathcal{I}_k^{(i,j)} \} \quad (5.7)$$

where l represents the index of the pyramid levels and $L \{ \mathcal{I} \}$ is the Laplacian version of the input \mathcal{I} while $G \{ \bar{\mathcal{W}} \}$ represents the Gaussian version of the normalized weight map of the $\bar{\mathcal{W}}$. This step is performed successively for each pyramid layer, in a bottom-up manner. The final haze-free image \mathcal{J} is obtained by summing the fused contribution of all inputs.

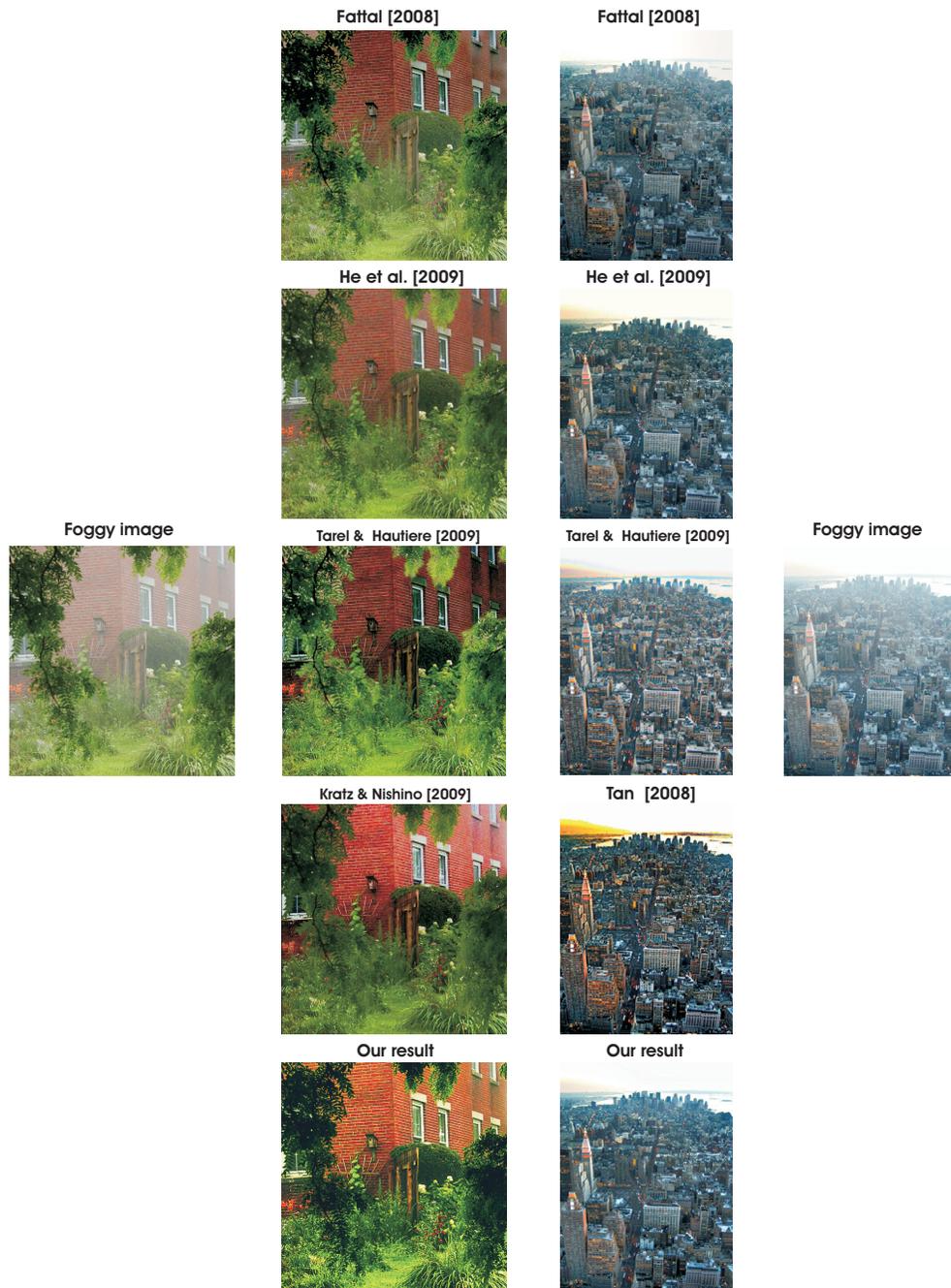


Figure 5.11: Comparison results among the recent single image dehazing methods.

5.3.5 Results and Discussion

In order to prove the robustness of our method we have tested a large dataset of natural hazy images. We also considered the complete sets of images provided by the authors of the previous single image dehazing methods. As can be seen in Figure 5.11 our operator is able to perform competitive with more complex methods. Some comparative results are shown as well in Figure 5.12.

	Tan	Fattal	He	Tarel	Kratz	Ours
Ampl.(%)	0.15	0.34	1.4	0.23	0.38	1.56
Loss(%)	2.7	1.9	2.2	1.74	2.47	1.34

Table 5.1: Amplification and loss of contrast induced by several operators by applying IQA measure

However, compared with most of the existing techniques, an advantage of our strategy is the computation time since our unoptimized implementation (Matlab) processes an image in approximately 3-4 seconds. In comparison, the method of Tan (Tan 08) needs more than 5 minutes per image while He et al. (He 09) requires 20 seconds.

In addition we performed a contrast preservation evaluation. For this task we employed the measure of Aydin et al. (Aydin 08). By comparing the original hazy image with the haze-free versions restored by different operators, this quality measure finds the regions where the contrast has been amplified (represented with *blue* pixels in their scheme) and regions where the contrast has been lost (*green* pixels). For more details please refer section 2.6.1. Table 5.1 displays the average ratio (%) of the pixels that have been filtered by applying IQA measure on several images provided by the authors. As can be observed compared with the other methods our operator is able to amplify better the original contrast (first row) while the loss of details is reduced.

Even though the method performs generally well, as the previous methods, a limitation of this algorithm is when the images are characterized by non-homogenous haze layers.

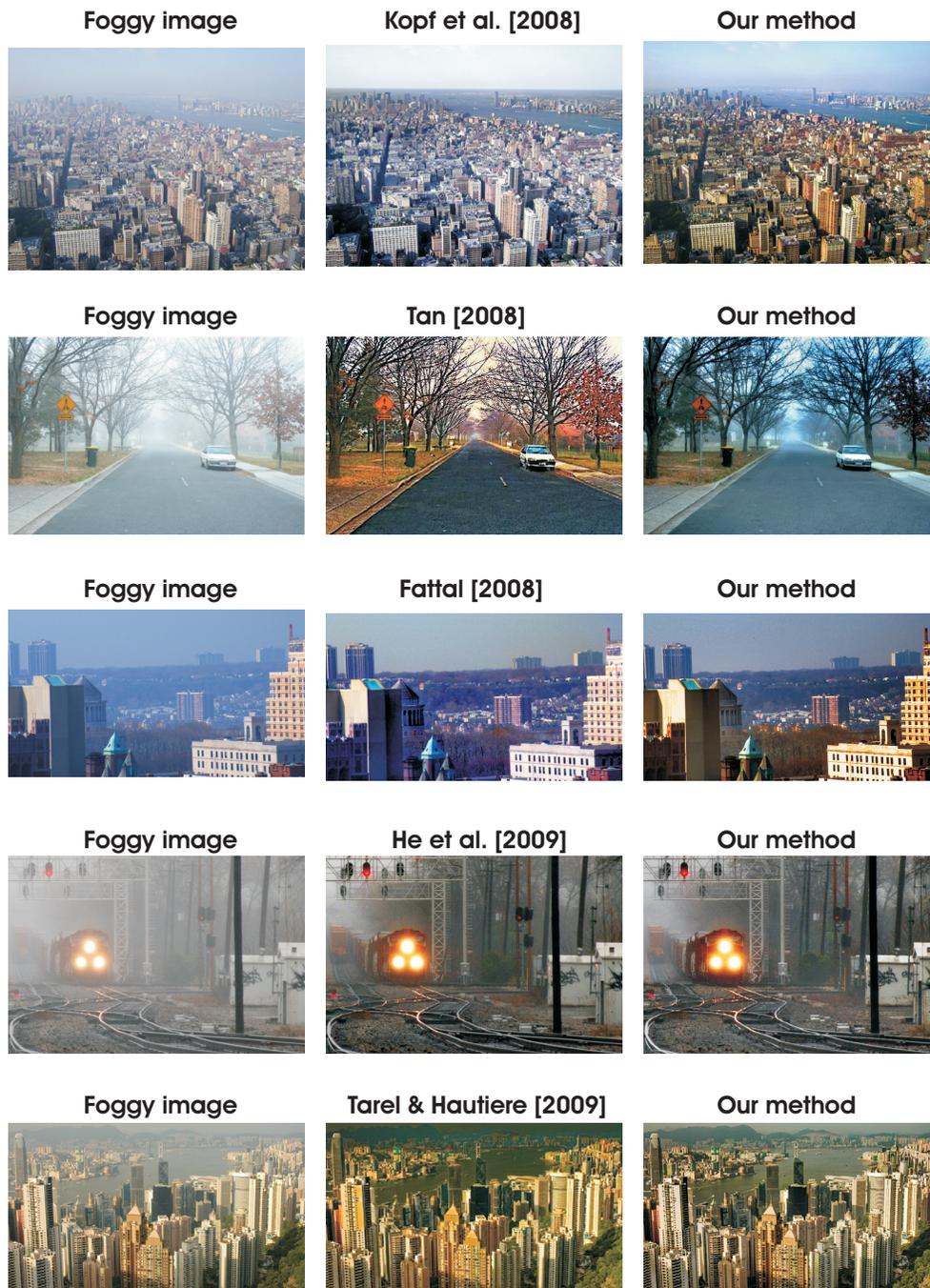


Figure 5.12: Additional comparative results against the recent single image dehazing methods (except **Deep Photo**, the method of Koph et al. (Kopf 08), that employs additionally an approximate 3D model).



Figure 5.13: Comparative results against our single image dehazing method presented in chapter 4.

On the other hand, by comparing our fusion-based dehazing technique with our method presented previously in chapter 4, a general observation would be that both are able to restore relatively accurate the hazy regions (see Figure 5.13) while processing relatively fast.

5.4 Summary

In this chapter we analyzed the classical concept of image fusion. As a main difference among fusion methods that makes them application-specific, is the choice of inputs and weights. We first demonstrates that the concept of image fusion can be used to decolorize accurately color images while preserving the original appearance and contrast. The algorithm employs the three independent *RGB* channels and an additional image that conserves the color contrast, based on Helmholtz-Kohlrausch effect, as image inputs. The weights used by our algorithm are based on three different forms of local contrast: saliency weight map, exposureness weight map and chromatic weight map.

Next, we demonstrate a novel fusion strategy to restore hazy images while using no additional information about the scene. The method is fast being straightforward to implement and shows to outperform the related operators in the contrast-based evaluation where IQA measure (Aydin 08) has been used.

For both methods we perform extensive experiments and provide comprehensive comparison against the existing state-of-the-art techniques.

Chapter 6

Conclusions

This thesis presents several techniques that manipulate the contrast of images, aiming to overcome the depiction limitations, due to the medium or image content. Contrast plays a vital role in achieving vivid, crisp and natural images with reduced amount of visual artifacts. Inspired by the work in perceptual image processing and color appearance, in this dissertation we addressed mainly two specific problems: image and video decolorization and restoration of images degraded by haze.

Perceptual image decolorization aims to maintain the original contrast appearance and the fine details, including those visible differences that are contained on the color channels. There are several proposed grayscale conversion algorithms from straightforward global mapping to much more complex and expensive optimizations. Generally, they are tightly dependent on the local image content and color palettes. Most of them are not suitable for video conversion due to the fact that color palettes change frequently.

The process of enhancing the visibility of the images is a challenging case that requires restoring both color and details from foggy regions. It has a wide applicability such as satellite imaging, surveillance systems and intelligent vehicles, all requiring optimal visibility conditions. General contrast enhancement techniques, such as histogram equalization or local contrast stretching are unable to produce reliable results. Due to the fact that the restoring process requires estimation of both airlight and direct attenuation, the problem is not trivial, as it is an under-constrained problem. Due to this fact, earlier haze-removal techniques make use of multiple images. Recently the complex case of single image dehazing has been addressed based on several well-defined assumptions or found prior information.

6.1 Summary of Contributions

Image and Video Decolorization

We presented a straightforward yet effective *perceptual decolorization algorithm* (Ancuti 11a) that enhances the contrast of the images while maintaining the initial appearance and quality. Aiming to preserve the original saliency, the monochromatic luminance channel is intensified by mixing both saturation and hue channels. This yields a new spatial distribution, which then is re-balanced in order to conserve the amount of glare impression that characterizes the original version. In general, decolorization is very subjective as it is dependent on how human beings perceive the chroma and color distribution. The perceptual experiments demonstrate the potential of the new method to preserve the original saliency of the color images. The novelty of this approach is that besides for effective decolorization, our operator is appropriate for several applications such as video decolorization, detail enhancement, segmentation under different illuminants, image matching by local feature points and auditory substitution systems.

In addition, we introduced a *fusion based color-to-grayscale* conversion strategy, in which we employ a multi-scale fusion algorithm (Ancuti 10g). We have shown that by choosing appropriate weight maps and inputs, an image fusion strategy can be used to effectively decolorize images. We performed an extensive evaluation against the recent decolorization operators. Moreover, our fusion-based operator is able to transform color videos into a decolorized version that preserves the original discriminability and appearance in a consistent manner. The originality of this algorithm is that it fuses the three independent R, G and B colors with an additional image that reflects the Helmholtz-kohlensch effect. The algorithm furthermore accounts for three different types of contrast which aims to measure and to preserve the color image appearance after decolorization.

Single Image Dehazing

We have presented a *layer-based single-image dehazing* strategy (Ancuti 10e) that has the advantage to not using any additional information (e.g. images, hardware, available depth information). Based on a per pixel hue disparity between the observed image and its semi-inverse, we are able to identify the hazy regions of the image. After we have identified these regions, we are able to produce a haze-free image using a layer-based approach. Our approach is conceptually straightforward while the processing time is very low, even

when compared to previous methods which were designed and optimized for speed. The algorithm has the main advantages: it is able to detect hazy regions, performs on a per-pixel basis and retains sharp details near transmission discontinuities.

Furthermore, we developed a *fusion-based image dehazing* solution for the same problem (Ancuti 10c). Our approach takes two inputs derived from the original image. These inputs are weighted by three normalized weight maps and finally blended in a multi-scale fashion that avoids introducing artifacts. Our technique has been tested for a large data set of natural hazy images. In general, the method performs faster than existing single image dehazing strategies yielding accurate results.

6.2 Future Work

In the recent years tremendous interest has been witnessed in the utilization of the contrast for image content manipulation purposes. The success of the existing methods shows that the depiction medium can be overcome by modifying the algorithms to detect the contrast changes and to adjust it to fit for each specific problem. We hope that our results will fuel further research on the perception-based contrast images manipulation field. We believe that there is still room for new research, especially in tone-reproduction based on scenes classification. Potential research direction that we intend to address in the future work (but not limited to):

Image decolorization:

Since image decolorization is quite subjective, we believe that one single solution will hardly satisfy all the expectations. One of the major drawbacks is caused by the saliency algorithm, it independently needs to identify the most salient regions in the image. In many situations the image content plays a key role and therefore saliency detection may fail. We would like to extend the saliency algorithm by including additional information such as scene classification, or objects identification. During tests of the fusion algorithm, we have noticed that the employed saliency can fail to produce a consistent regional contrast since only the global contrast differences are accounted. We want to extend this technique to consider simultaneously the global contrast and the spatial coherence. We believe that the algorithm can be included by other applications such as edge preserving filtering, single image depth estimation and segmentation. Additionally, we would like to investigate the potential of our algorithm for several other pattern recognition applications that are still

relying to the standard decolorizing technique. We intend to develop an objective color-to-grayscale measure that is able to evaluate both the potential of preserving finest details and the original discriminability.

Image dehazing:

As a future direction, we intend as well to extend our dehazing strategies to the problem of videos. Additionally, we plan to address the more complex case when the images are characterized by non-homogenous haze. We would like to investigate a more comprehensive optical model that would allow to restore more generally the images. We plan to extend the algorithm for visibility recovery for underwater images. We believe that a promising direction is to combine the dehazing with superresolution and image denoising techniques, since in many cases the recovered regions require further enhancement mainly due to the noise.

Appendices

Appendix A

Scientific Contributions and Publications

The following list of publications, presented at scientific international conferences, contains work that is part of this dissertation:

- (**Ancuti 11a**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Enhancing by Saliency-guided Decolorization. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2011*
- (**Ancuti 10g**) *Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. Image and Video Decolorization by Fusion. In Proceedings of 10th Asian Conference on Computer Vision (ACCV) (**oral presentation**), 2010*
- (**Ancuti 10e**) *Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. A Fast Semi-Inverse Approach to Detect and Remove the Haze from a Single Image. In Proceedings of 10th Asian Conference on Computer Vision (ACCV), 2010*
- (**Ancuti 10c**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Effective Single Image Dehazing by Fusion. In Proceedings of 17th IEEE International Conference on Image Processing (IEEE ICIP), 2010*
- (**Ancuti 10b**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Decolorizing Images for Robust Matching. In Proceedings of 17th IEEE International Conference on Image Processing (IEEE ICIP), 2010*
- (**Ancuti 10f**) *Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. Fusion-based Image and Video Decolorization. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2010*
- (**Ancuti 10h**) *Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. Layer-based Single Image Dehazing by Per-Pixel Haze Detection. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2010*

- (**Ancuti 10a**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. CC-SIFT: Exploiting Chromatic Contrast for Wide-Baseline Matching. In Proceedings of 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP), 2010*
- (**Ancuti 09b**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. An effective grayscale conversion with applications to image enhancement. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2009*
- (**Ancuti 09a**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. ColEnViSon: Color Enhanced Visual Sonifier. A Polyphonic Audio Texture and Salient Scene Analysis. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), February 2009*
- (**Ancuti 09d**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Preserving Visual Saliency in Image to Sound Substitution Systems. In Proceedings of SPIE Human Vision and Electronic Imaging (HVEI 2009), January 19-21 2009*

Several publications that have not been included in this dissertation:

- (**Ancuti 08b**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. Robust Matching of Images Across Large Viewpoint Changes. In Proceedings of Visualization, Imaging and Image Processing (VIIP 2008), 2008*
- (**Ancuti 08a**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. Image Deblurring by Corresponding Regions. In Conference Abstracts and Applications ACM SIGGRAPH, August 11-15 2008*
- (**Ancuti 09c**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Effective Sound-Vision Substitution System Guided by Salient Regions. 2009*
- (**Ancuti 09f**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. An Efficient Two Steps Algorithm for Wide Baseline Image Matching. In Journal of Visual Computer, (Computer Graphics International Conference), May 2009*
- (**Ancuti 09e**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. Deblurring by Matching. In Computer Graphics Forum, (**Eurographics**), April 2009*
- (**Ancuti 10i**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. A Framework to Improve Matching Results of Widely Separated Views. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), 2010*

- (**Ancuti 10d**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Robust Grayscale Conversion for Vision-Substitution Systems. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), 2010*
- (**Ancuti 10j**) *Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. Video Super-Resolution using High Quality Photographs. In Proceedings of 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP), 2010*
- (**Gerrits 11**) *Mark Gerrits, Bert de Decker, Cosmin Ancuti, Tom Haber, Codruta O. Ancuti & Philippe Bekaert. Stroke-based Creation of Depth Maps. In IEEE International Conference on Multimedia and Expo (IEEE ICME), 2011*
- (**Ancuti 11b**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Enhancing Underwater Images by Fusion. In Conference Abstracts and Applications ACM SIGGRAPH- Posters, 2011*
- (**Ancuti 11c**) *Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. Fusion-based Restoration of the Underwater Images. In Proceedings of 18th IEEE International Conference on Image Processing (IEEE ICIP), 2011*

Appendix B

Samenvatting (Dutch Summary)

Ondanks de recentetecnologische vooruitgang ondervinden digitale camera's tot op heden vaak vele beperkingen. Hierdoor vereisen de meeste systemen nog nabewerking van de gemaakte foto's. Een groot deel van ons werk is geïnspireerd op de visuele perceptie van de mens. Dit onderzoeksgebied houdt zich bezig met de wijze waarop onze hersenen signalen afkomstig van de ogen interpreteren. De kennis van het menselijke gezichtsvermogen maakt het voor wetenschappers mogelijk om de inhoud van afbeeldingen te begrijpen, afbeeldingen op een meer accurate wijze voor te stellen op basis van de feitelijke perceptie van een scène of zelfs het visuele waarneembare detail te verbeteren (beeldherstelling van nevel of mist). Naast het gebruik van het menselijke gezichtsvermogen, is ons werk ook gebaseerd op perceptuele veronderstellingen en observaties. In dit proefschrift onderzoeken we verschillende nieuwe technieken. Deze zijn voornamelijk gebaseerd op het analyseren en manipuleren van contrast in een afbeelding. Contrast is niet enkel van belang voor het waarnemen van de 'lichtheid' van een scène, maar ook voor het waarnemen van kleur. Het gebruik van een geschikte techniek voor contrastmanipulatie zal er dus voor zorgen voor dat er een gewenste verbetering is in de algemene weergave van de afbeelding. Dit werk presenteert nieuwe technieken voor de transformatie van kleur, stijl en weergave van een afbeelding. We hebben de impact van contrastmanipulatie onderzocht voor twee belangrijke toepassingen: kleur naar grijschaal transformatie (decolorization) van afbeeldingen/videos en herstelling van nevelige of mistige afbeeldingen (dehazing). Onze bedoeling is om de bestaande beperkingen, opgelegd door de inhoud van een beeld, te voorkomen. Daar afbeeldingen met laag contrast de

hoeveelheid visuele zichtbare informatie beperken, worden afbeeldingen met hoger contrast doorgaans meer accuraat waargenomen. Bovendien is aangetoond dat een hoger contrast zorgt voor een vereenvoudiging van de visuele stimulus en een verhoging van de dynamic range van een afbeelding zoals waarneembaar in de scene zelf. In het eerste deel introduceren we een algoritme voor decolorization van afbeeldingen of videos gebaseerd op de oorspronkelijke saliency. Deze methode is geïnspireerd op de opponent process theorie van Hering en heeft als bedoeling het contrast enkel te verhogen in regio's die belangrijk zijn. In het daaropvolgende deel presenteren we een nieuwe dehazing techniek die gebruik maakt van een enkele afbeelding en het gekende dark object principe. Tot slot presenteren we twee nieuwe fusion-based technieken die toepasbaar zijn voor decolorization en dehazing van afbeeldingen. Eerst demonsteren we dat, door het definiëren van geschikte invoer en gewichtsmappen, een fusion-based strategie accurate resultaten kan opleveren en zo de waarneming en verscheidenheid van de visuele informatie in kleurenbeelden behouden blijft of verbeterd wordt. Uitgebreide experimenten en vergelijkingen met bestaande state-of-the-art technieken zijn opgenomen in dit proefschrift en tonen de accuraatheid en bruikbaarheid van onze nieuwe methodes aan.

Bibliography

- [Abdel-Hakim 06] A. E. Abdel-Hakim & A. A. Farag. *CSIFT: A SIFT Descriptor with Color Invariant Characteristics*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.
- [Achanta 09] R. Achanta, S. Hemami, F. Estrada & S. Susstrunk. *Frequency-tuned Salient Region Detection*. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
- [Adams 81] A. Adams. *The Negative*. Little, Brown and Company, 1981.
- [Agarwala 04] A. Agarwala, M. Dontcheva, M. Agrawala, S. M. Drucker, A. Colburn, B. Curless, D. Salesin & M. F. Cohen. *Interactive digital photomontage*. SIGGRAPH, ACM Transactions on Graphics, 2004.
- [Amedi 05] A. Amedi & P.B.L. Meijer. *Neural correlates of visual-to-auditory sensory substitution in proficient blind users*. 6th Annual Meeting of the International Multisensory Research Forum, 2005.
- [Ancuti 08a] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *Image Deblurring by Corresponding Regions*. In Conference Abstracts and Applications ACM SIGGRAPH, August 11-15 2008.
- [Ancuti 08b] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *Robust Matching of Images Across Large Viewpoint Changes*. In Proceedings of Visualization, Imaging and Image Processing (VIIP 2008), 2008.

- [Ancuti 09a] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *ColEnViSon: Color Enhanced Visual Sonifier. A Polyphonic Audio Texture and Salient Scene Analysis*. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), February 2009.
- [Ancuti 09b] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *An effective grayscale conversion with applications to image enhancement*. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2009.
- [Ancuti 09c] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Effective Sound-Vision Substitution System Guided by Salient Regions*. 2009.
- [Ancuti 09d] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Preserving Visual Saliency in Image to Sound Substitution Systems*. In Proceedings of SPIE Human Vision and Electronic Imaging (HVEI 2009), January 19-21 2009.
- [Ancuti 09e] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *Deblurring by Matching*. In Computer Graphics Forum, (**Eurographics**), April 2009.
- [Ancuti 09f] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *An Efficient Two Steps Algorithm for Wide Baseline Image Matching*. In Journal of Visual Computer, (Computer Graphics International Conference), May 2009.
- [Ancuti 10a] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *CC-SIFT: Exploiting Chromatic Contrast for Wide-Baseline Matching*. In Proceedings of 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP), 2010.
- [Ancuti 10b] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Decolorizing Images for Robust Matching*. In Proceedings of 17th IEEE International Conference on Image Processing (IEEE ICIP), 2010.
- [Ancuti 10c] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Effective Single Image Dehazing by Fusion*. In Proceedings

- of 17th IEEE International Conference on Image Processing (IEEE ICIP), 2010.
- [Ancuti 10d] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Robust Grayscale Conversion for Vision-Substitution Systems*. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), 2010.
- [Ancuti 10e] Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. *A Fast Semi-Inverse Approach to Detect and Remove the Haze from a Single Image*. In Proceedings of 10th Asian Conference on Computer Vision (ACCV), 2010.
- [Ancuti 10f] Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. *Fusion-based Image and Video Decolorization*. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2010.
- [Ancuti 10g] Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. *Image and Video Decolorization by Fusion*. In Proceedings of 10th Asian Conference on Computer Vision (ACCV) (**oral presentation**), 2010.
- [Ancuti 10h] Codruta O. Ancuti, Cosmin Ancuti, Chris Hermans & Philippe Bekaert. *Layer-based Single Image Dehazing by Per-Pixel Haze Detection*. In Conference Abstracts and Applications ACM SIGGRAPH ASIA - Sketches, 2010.
- [Ancuti 10i] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *A Framework to Improve Matching Results of Widely Separated Views*. In Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP), 2010.
- [Ancuti 10j] Cosmin Ancuti, Codruta O. Ancuti & Philippe Bekaert. *Video Super-Resolution using High Quality Photographs*. In Proceedings of 35th IEEE International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP), 2010.
- [Ancuti 11a] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Enhancing by Saliency-guided Decolorization*. In IEEE

- Conference on Computer Vision and Pattern Recognition (CVPR), 2011.
- [Ancuti 11b] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Enhancing Underwater Images by Fusion*. In Conference Abstracts and Applications ACM SIGGRAPH- Posters, 2011.
- [Ancuti 11c] Codruta O. Ancuti, Cosmin Ancuti & Philippe Bekaert. *Fusion-based Restoration of the Underwater Images*. In Proceedings of 18th IEEE International Conference on Image Processing (IEEE ICIP), 2011.
- [Arno 99] P. Arno, C. Capelle, D.M. Wanet, A.M. Catalan & C. Verbaart. *Auditory coding of visual patterns for the blind*. In Perception, volume 28(8), pages 1013–1029, 1999.
- [Auvray 05] M. Auvray, S. Hanneton & C. Lenay. *There is something out there: distal attribution in sensory substitution, twenty years later*. Journal of Integrative Neuroscience, vol. 4, pages 505–521, 2005.
- [Aydin 08] T. O. Aydin, R. Mantiuk, K. Myszkowski & H.-S. Seidel. *Dynamic range independent image quality assessment*. In SIGGRAPH, ACM Transactions on Graphics (TOG), volume 27, pages 1–10, 2008.
- [Bala 04] Raja Bala & Reiner Eschbach. *Spatial Color-to-Grayscale Transform Preserving Chrominance Edge Information*. In Color Imaging Conference, pages 82–84, 2004.
- [Bavelier 02] D. Bavelier & H. Neville. *Cross-modal plasticity: where and how?* Nature Reviews Neuroscience, vol. 3, no. 6, pages 443–452, 2002.
- [Benavente 03] Robert Benavente, Ramon Baldrich, Maria Vanrell & Anna Salvatella. *Color image enhancement based on perceptual sharpening*. In IEEE International Conference on Image Processing (ICIP), 2003.
- [Bregman 90] A. Bregman. *Auditory scene analysis*. MIT Press, Cambridge, MA, 1990.

- [Brinkmann 99] R. Brinkmann. *The Art and Science of Digital Compositing*. Morgan Kaufmann, 1999.
- [Brown 07] M. Brown & D. Lowe. *Automatic Panoramic Image Stitching using Invariant Features*. International Journal of Computer Vision, vol. 74, no. 1, pages 59–73, 2007.
- [Burt 83] P. Burt & T. Adelson. *The Laplacian Pyramid as a Compact Image Code*. IEEE Transactions on Communication, 1983.
- [Burt 93] P. Burt, K. Hanna & R. Kolczynski. *Enhanced image capture through fusion*. IEEE International Conference on Computer Vision, 1993.
- [Cadik 08] Martin Cadik. *Perceptual Evaluation of Color-to-Grayscale Image Conversions*. Computer Graphics Forum, vol. 27, no. 7, 2008.
- [Calabria 03] A.J. Calabria & M.D. Fairchild. *Perceived image contrast and observer preference I: The effects of lightness, chroma, and sharpness manipulations on contrast perception*. Journal of Imaging Science and Technology, vol. 47, pages 479–493, 2003.
- [Capelle 98] C. Capelle, C. Trullemans, P. Arno P. & Veraart C. *A real-time experimental prototype for enhancement of vision rehabilitation using auditory substitution*. IEEE Transactions on Biomedical Engineering, vol. 45, no. 10, pages 1279–1293, 1998.
- [Chavez 88] P.S. Chavez. *An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data*. Remote Sensing of Environment, 1988.
- [Chong 08] Hamilton Y. Chong, Steven J. Gortler & Todd Zickler. *A perception-based color space for illumination-invariant image processing*. SIGGRAPH, Transactions on Graphics (TOG), vol. 27, no. 3, 2008.
- [Cohen-Or 06] Daniel Cohen-Or, Olga Sorkine, Ran Gal, Tommer Leyvand & Ying-Qing Xu. *Color Harmonization*. SIGGRAPH, ACM Transactions on Graphics (TOG), 2006.

- [Cohen 64] J. Cohen. *Dependency of the Spectral Reflectance Curves of the Munsel Color Chips*. Psychonimic Science, vol. 1, pages 369–370, 1964.
- [Col 86] Colorimetry, second edition: Publication 15.2 (e-1.3.1). Central Bureau of the CIE, 1986.
- [Conway 02] Bevil Richard Conway. *Neural mechanisms of color vision: double-opponent cells in the visual cortex*. Kluwer Academic, 2002.
- [Cronly-Dillon 99] J. Cronly-Dillon, K. Persaud & R. Gregory. *The Perception of Visual Images Encoded in Musical Form: a Study in Cross-Modality Information*. Biological Sciences, pages 2427–2433, 1999.
- [Dannemiller 89] J.L. Dannemiller. *A test of color constancy in 9-and-20-week-old human infants following simulated illuminant changes*. Developmental Psychology, vol. 25, no. 2, pages 171–184, 1989.
- [Debevec 97] Paul E. Debevec & Malik Jitendra. *Recovering high dynamic range radiance maps from photographs*. In Proc. ACM SIGGRAPH, 1997.
- [Durand 02] Frédo Durand & Julie Dorsey. *Fast bilateral filtering for the display of high-dynamic-range images*. SIGGRAPH, ACM Transactions on Graphics (TOG), vol. 21, pages 257–266, 2002.
- [Durette 08] B. Durette, N. Louveton, D. Alleysson & J. Herault. *Visio-auditory sensory substitution for mobility assistance*. In ECCV Comp. Vis. Applic. for the Visually Impaired, 2008.
- [Engel 97] S. Engel, X. Zhang & B. Wandell. *Colour Tuning in Human Visual Cortex Measured With Functional Magnetic Resonance Imaging*. Nature, vol. 338, no. 6, pages 68–71, 1997.
- [Fairchild 91] M. Fairchild & E. Pirrotta. *Predicting the lightness of chromatic object colors using CIELAB*. Color Research and Application, 1991.

- [Fairchild 05] Mark D. Fairchild. *Color appearance models*, 2nd edition. Wiley-IS&T, Chichester, 2005.
- [Farbman 08] Z. Farbman, R. Fattal, D. Lischinski & R. Szeliski. *Edge-preserving decompositions for multi-scale tone and detail manipulation*. ACM SIGGRAPH, 2008.
- [Fattal 02] R. Fattal, D. Lischinski & M. Werman. *Gradient domain high dynamic range compression*. SIGGRAPH, ACM Transactions on Graphics, 2002.
- [Fattal 08] Raanan Fattal. *Single image dehazing*. SIGGRAPH, ACM Transactions on Graphics (TOG), 2008.
- [Forssén 07] Per-Erik Forssén & David Lowe. *Shape Descriptors for Maximally Stable Extremal Regions*. In In IEEE International Conference on Computer Vision (ICCV), pages 1–8, 2007.
- [Gerrits 11] Mark Gerrits, Bert de Decker, Cosmin Ancuti, Tom Haber, Codruta O. Ancuti & Philippe Bekaert. *Stroke-based Creation of Depth Maps*. In IEEE International Conference on Multimedia and Expo (IEEE ICME), 2011.
- [Gooch 05] Amy A. Gooch, Sven C. Olsen, Jack Tumblin & Bruce Gooch. *Color2Gray: salience-preserving color removal*. SIGGRAPH, ACM Trans. Graph., vol. 24, no. 3, pages 634–639, 2005.
- [Grundland 06] M. Grundland, R. Vohra, G. P. Williams & N. A. Dodgson. *Cross dissolve without cross fade : Preserving contrast, color and salience in image compositing*. Computer Graphics Forum (EUROGRAPHICS), 2006.
- [Grundland 07] Mark Grundland & Neil A. Dodgson. *Decolorize: Fast, Contrast Enhancing, Color to Grayscale Conversion*. Pattern Recognition, vol. 40, no. 11, 2007.
- [Guild 31a] John Guild. *The colorimetric properties of the spectrum*. Philosophical Transactions of the Royal Society of London, vol. 230, pages 149–187, 1931.

- [Guild 31b] John Guild & Thomas Smith. *The C.I.E. colorimetric standards and their use*. Transactions of the Optical Society, vol. 33, no. 3, pages 73–134, 1931.
- [Haeberli 94] P. Haeberli. *A Multifocus Method for Controlling Depth of Field*. 1994.
- [Hautiere 08] N. Hautiere, J.-P. Tarel, D. Aubert & E. Dumont. *Blind contrast enhancement assessment by gradient ratioing at visible edges*. Image Analysis and Stereology, 2008.
- [He 09] K. He, J. Sun & X. Tang. *Single Image Haze Removal Using Dark Channel Prior*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
- [Huang 08] Chun-Rong Huang, Chu-Song Chen & Pau-Choo Chung. *Contrast context histogram-An efficient discriminating local descriptor for object recognition and image matching*. Pattern Recognition, 2008.
- [Hubel 88] David H. Hubel. *Eye, brain, and vision*. W H Freeman & Co, 1988.
- [Hunt 98] R. W. G. Hunt. *Measuring colour*, 3rd edition. West Sussex: Fountain Press, 1998.
- [Itti 98] L. Itti, C. Koch & E. Niebur. *A model of saliency-based visual attention for rapid scene analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 20, no. 11, pages 1254–1259, 1998.
- [Judd 64] D. B. Judd, D. L. MacAdam & G. Wyszecki. *Spectral distribution of typical daylight as a function of correlated color temperature*. Journal of Optical Society of America, vol. 54, 1964.
- [Judd 66] Deane B. Judd. *Fundamental studies of color vision from 1860 to 1960*. Proceedings of the National Academy of Sciences of the USA, vol. 55, no. 6, 1966.
- [Kim 09] Y. Kim, C. Jang, J. Demouth & S. Lee. *Robust color-to-gray via nonlinear global mapping*. ACM TOG (SIGGRAPH Asia), 2009.

- [Koch 85] C. Koch & S. Ullman. *Shifts in selective visual attention: towards the underlying neural circuitry*. Human Neurobiology, no. 4, pages 219–227, 1985.
- [Kolomenkin 06] M. Kolomenkin & I. Shimshoni. *Image Matching Using Photometric Information*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pages 2506–2514, 2006.
- [Kopf 08] J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele & D. Lischinski. *Deep Photo: Model-Based Photograph Enhancement and Viewing*. In Siggraph ASIA, ACM Transaction on Graphics (TOG), 2008.
- [Koschmieder 24] H. Koschmieder. *Theorie der horizontalen sichtweite*. In Beitrage zur Physik der freien Atmosphere, 1924.
- [Kratz 09] L. Kratz & K. Nishino. *Factorizing Scene Albedo and Depth from a Single Foggy Image*. In IEEE International Conference on Computer Vision (ICCV), 2009.
- [Land 71] Edwin Land & John McCann. *Lightness and Retinex Theory*. Journal of the Optical Society of America, vol. 61, 1971.
- [Liu 07] T. Liu, J. Sun, N.-N. Zheng, X. Tang & H.-Y. Shum. *Learning to detect a salient object*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2007.
- [Livingstone 08] Margaret Livingstone & David Hubel. *Vision and art: The biology of seeing*. Abrams, Harry N., Inc., New York, NY, USA, May 2008.
- [Lowe 04] D. Lowe. *Distinctive image features from scale-invariant keypoints*. In International Journal of Computer Vision, volume 20, pages 91–110, 2004.
- [Lubin 95] J. Lubin. *Vision Models for Target Detection and Recognition*. World Scientific, pages 245–283, 1995.
- [Madden 93] B. Madden. *Extended Intensity Range Imaging*. Technical Report MS-CIS-93-96, University of Pennsylvania, 1993.

- [Magnussen 75] S. Magnussen & A. Glad. *Temporal frequency characteristic of spatial interaction in human vision*. Experimental Brain Research, 1975.
- [Maloney 86] L.T. Maloney & B.A. Wandell. *Color Constancy: A method for recovering surface spectral reflectance*. Journal of Optical Society of America, vol. 3, no. 1, pages 29–33, 1986.
- [Mann 95] S. Mann & R.W. Picard. *Being undigital with digital cameras: Extending dynamic range by combining differently exposed pictures*. In Proc. IS&T 46th annual conference, pages 422–428, 1995.
- [Mantiuk 06] Rafal Mantiuk, Karol Myszkowski & Hans-Peter Seidel. *A perceptual framework for contrast processing of high dynamic range images*. ACM Trans. Appl. Percept., 2006.
- [Mantiuk 09] Radoslaw Mantiuk, Rafal Mantiuk, Anna Tomaszewska & Wolfgang Heidrich. *Color Correction for Tone Mapping*. In Computer Graphics Forum, volume 28, 2009.
- [Marchesotti 09] L. Marchesotti, C. Cifarelli & G. Csurka. *A framework for visual saliency detection with applications to image thumbnailing*. In IEEE International Conference on Computer Vision (ICCV), 2009.
- [Marr 80] Marr & E. Hildreth. *Theory of Edge Detection*. 1980.
- [Meijer 92] P. Meijer. *An experimental system for auditory image representations*. IEEE Transactions on Biomedical Engineering, vol. 39, pages 112–121, 1992.
- [Meijer 98] P. Meijer. *Cross-modal sensory streams*. In Conference Abstracts and Applications, ACM SIGGRAPH, 1998.
- [Mertens 07] T. Mertens & J. Kautz. *Exposure Fusion*. In Pacific Graphics, pages 382–390, 2007.
- [Mertens 09] T. Mertens, J. Kautz & Frank Van Reeth. *Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography*. Computer Graphics Forum, vol. 28, no. 1, pages 161–171, 2009.

- [Michael 78] C.R. Michael. *Color vision mechanism in monkey striate cortex: Dual-opponent cells with concentric receptive fields*. vol. 41, pages 572–588, 1978.
- [Mikolajczyk 05] K. Mikolajczyk & C. Schmid. *A Performance Evaluation of Local Descriptors*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 27, no. 10, 2005.
- [Milanese 93] R. Milanese. *Detecting Salient Regions in an Image: From Biological Evidence to Computer Implementation*. PhD thesis, 1993.
- [Mitsunaga 99] T. Mitsunaga & S. K. Nayar. *Radiometric self calibration*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 2, pages 374–380, 1999.
- [Moreels 07] Pierre Moreels & Pietro Perona. *Evaluation of Features Detectors and Descriptors based on 3D Objects*. International Journal of Computer Vision, vol. 73, no. 3, pages 263–284, 2007.
- [Morimura 93] A. Morimura. *Imaging method for a wide dynamic range and an imaging device for a wide dynamic range*. 1993.
- [Moro 06] G. Dal Moro & L. Halounova. *Haze Removal and Data Calibration for High-Resolution Satellite Data*. International Journal of Remote Sensing, 2006.
- [Namer 09] E. Namer, S. Shwartz & Y.Y. Schechner. *Skyless polarimetric calibration and visibility enhancement*. Optic Express, no. 17, pages 472–493, 2009.
- [Narasimhan 00] S.G. Narasimhan & S.K. Nayar. *Chromatic Framework for Vision in Bad Weather*. In In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2000.
- [Narasimhan 03a] S.G. Narasimhan & S.K. Nayar. *Contrast Restoration of Weather Degraded Images*. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2003.
- [Narasimhan 03b] S.G. Narasimhan & S.K. Nayar. *Interactive de-weathering of an image using physical models*. ICCV Workshop CP-MVC, 2003.

- [Nayatani 97] Y. Nayatani. *Simple estimation methods for the Helmholtz-Kohlrausch effect*. Color Research and Application, vol. 22, pages 385–401, 1997.
- [Nayatani 98] Y. Nayatani. *Relations between the two kinds of representation methods in the Helmholtz-Kohlrausch effect*. Color Research and Application, 1998.
- [Nemcsis 87] A. Nemcsis. *Color Space of the Coloroid Color System*. In Color Research and Applications, volume 12, 1987.
- [Neumann 07] Laszlo Neumann, Martin Cadik & Antal Nemcsics. *An Efficient Perception-Based Adaptive Color to Gray Transformation*. In Proceedings of Computational Aesthetics 2007, pages 73–80, 2007.
- [Palmer 99] Stephen Palmer. Vision science: Photons to phenomenology. The MIT Press, Cambridge, Massachusetts, first edition edition, May 1999.
- [Pattanaik 98] Sumanta N. Pattanaik, James A. Ferwerda, Mark D. Fairchild & Donald P. Greenberg. *A multiscale model of adaptation and spatial vision for realistic image display*. In SIGGRAPH, ACM Transactions on Graphics (TOG), 1998.
- [Peli 90] E. Peli. *Contrast in Complex Images*. Journal of the Optical Society of America, pages 2032–2040, 1990.
- [Peli 00] E. Peli. *Contrast in complex images*. Journal of Optical Society of America, 2000.
- [Pelli 02] D. G. Pelli, J. G. Robson & A. J. Wilkins. *The design of a new letter chart for measuring contrast sensitivity*. Clinical Vision Sciences, pages 187–199, 2002.
- [Perez 03] P. Perez, M. Gangnet & A. Blake. *Poisson image editing*. SIGGRAPH, ACM Transactions on Graphics, 2003.
- [Pun 07] Thierry Pun, Patrick Roth, Guido Bologna, Konstantinos Moustakas & Dimitrios Tzovaras. *Image and video processing for visually handicapped people*. Journal Image Video Processing, vol. 5, pages 1–12, 2007.

- [Rafael 08] Gonzalez C. Rafael & Woods E. Richard. Digital image processing, 2nd edition. 2008.
- [Rahman 97] Z.U. Rahman & G.A. Woodell. *A multi-scale retinex for bridging the gap between color images and the human observation of scenes*. In IEEE Transactions on Image Processing, 1997.
- [Rasche 05] Karl Rasche, Robert Geist & James Westall. *Re-coloring Images for Gamuts of Lower Dimension*. Computer Graphics Forum, vol. 24, no. 3, pages 423–432, 2005.
- [Raskar 03] R. Raskar, J. Yu & A. Ilie. *Video Surveillance with NPR Image Fusion*. MERL Techn. Report, 2003.
- [Raskar 04] R. Raskar, A. Ilie & J. Yu. *Image Fusion for Context Enhancement and Video Surrealism*. International Symposium on Non-Photorealistic Animation and Rendering (NPAR), pages 85–152, 2004.
- [Raskar 06] Ramesh Raskar, Jack Tumblin, Ankit Mohan, Amit Agrawal & Yuanzen Li. *State of The Art Report (STAR): Computational Photography*. ACM/Eurographics, 2006.
- [Ratlif 65] F. Ratlif. *Mach bans: Quantitative studies in neural networks in the retina*. Floyd Ratliff. Holden-Day, San Francisco, 1965.
- [R.C.Henry 00] R.C.Henry, S.Mahadev, S. Urquijo & D. Chitwood. *Color perception through atmospheric haze*. Journal of Optical Society of America, vol. 17, pages 831–835, 2000.
- [Reinhard 05] E. Reinhard & K. Devlin. *Dynamic range reduction inspired by photoreceptor physiology*. IEEE Transactions on Visualization and Computer Graphics, 2005.
- [Rosenfeld 82] Azriel Rosenfeld & Kak Avinash. *Digital picture processing*. 1982.
- [Schaul 09] L. Schaul, C. Fredembach & S. Ssstrunk. *Color Image Dehazing using the Near-Infrared*. In IEEE International Conference on Image Processing (ICIP), 2009.

- [Schubert 92] J. Schubert & A. L. Gilchrist. *Relative luminance is not derived from absolute luminance*. Investigative Ophthalmology and Visual Science, vol. 33, no. 4, pages 161–185, 1992.
- [Shepard 92] R. N. Shepard. The perceptual organization of color: An adaptation to the regularities of the terrestrial world? 1992.
- [Shwartz 06] S. Shwartz, E. Namer & Y.Y. Schechner. *Blind haze separation*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006.
- [Smith 78] Alvy Ray Smith. *Color gamut transform pairs*. 1978.
- [Smith 06] K. Smith, G. Krawczyk, K. Myszkowski & H.-S. Seidel. *Beyond Tone Mapping: Enhanced Depiction of Tone Mapped HDR Images*. In Computer Graphics Forum, volume 25, pages 427–438, 2006.
- [Smith 08] Kaleigh Smith, Pierre-Edouard Landes, Jelle Thollot & Karol Myszkowski. *Apparent Greyscale: A Simple and Fast Conversion to Perceptually Accurate Images and Video*. Computer Graphics Forum, vol. 27, no. 3, pages 193–200, 2008.
- [Socolinsky 02] D. Socolinsky & L. Wolff. *Multispectral image visualization through first-order fusion*. IEEE Transactions on Image Processing, 2002.
- [Strybel 98] T. Z. Strybel & M.L. Menges. *Auditory apparent motion between sine waves differing in frequency*. Perception, vol. 27(4), pages 483–495, 1998.
- [Subr 09] Kartic Subr, Cyril Soler & Fredo Durand. *Edge-preserving Multiscale Image Decomposition based on Local Extrema*. ACM Transactions on Graphics (SIGGRAPH ASIA), vol. 28, pages 1471–1479, 2009.
- [Tabachnick 05] B. G. Tabachnick & L. S. Fidell. Using multivariate statistics. Allyn & Bacon, 5th edition, 2005.

- [Tan 08] Robby T. Tan. *Visibility in Bad Weather from A Single Image*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008.
- [Tao 09] Litian Tao, Lu Yuan & Jian Sun. *Sky Finder: Attribute-based Sky Image Search*. Siggraph, ACM Transactions on Graphics (TOG), 2009.
- [Tarel 09] J.P. Tarel & N. Hautiere. *Fast Visibility Restoration from a Single Color or Gray Level Image*. IEEE International Conference on Computer Vision (ICCV), 2009.
- [Treibitz 09] T. Treibitz & Y. Y. Schechner. *Polarization: Beneficial for Visibility Enhancement?* In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
- [Treisman 85] A. Treisman. *Preattentive processing in vision*. Computer Vision, Graphics, and Image Processing, no. 31, pages 156–177, 1985.
- [Tsai 93] Y. T. Tsai. *Method and apparatus for extending the dynamic range of an electronic imaging system*. 1993.
- [Tuytelaars 08] T. Tuytelaars & K. Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Found. and Trends in Comp. Graph. and Vis., 2008.
- [Valazquez 05] R. Valazquez, E. E. Pissaloux, J. C. Guinot & F. Maingreud. *Walking using touch: Design and preliminary prototype of a noninvasive ETA for the visually impaired*. In in Proceedings of the the Medicine and Biology 27th Annual Conference, 2005.
- [Valois 90] Russell L. De Valois & Karen K De Valois. *Spatial vision*. Oxford University Press, 1990.
- [Wang 02] Z. Wang & A. C. Bovik. *A universal image quality index*. IEEE Signal Processing Letters, vol. 9, no. 3, pages 81–84, 2002.
- [Wang 04] Z. Wang, A. C. Bovik, H. R. Sheikh & E. P. Simoncelli. *Image quality assessment: From error visibility to structural similarity*. IEEE Transactions on Image Processing, vol. 13, no. 4, 2004.

- [Wang 09] G. Wang & D. Forsyth. *Joint learning of visual attributes, object classes and visual saliency*. IEEE International Conference on Computer Vision (ICCV), 2009.
- [Ware 83] C. Ware & W.B. Cowan. *The chromatic Cornsweet effect*. Vision Research, vol. 23, pages 1075–1077, 1983.
- [Wilson 99] Robert A. Wilson & Frank C. Keil. *The MIT Encyclopedia of the Cognitive Sciences*. Cerebral Cortex: Top-Down Processing in Vision, 1999.
- [Wright 28] William David Wright. *A re-determination of the trichromatic coefficients of the spectral colors*. Transactions of the Optical Society, vol. 30, pages 141–164, 1928.
- [Wyszecki 00] G. Wyszecki & W. S. Stiles. *Color science: Concepts and methods, quantitative data and formulae*, 2nd edition. Wiley-Interscience, 2000.
- [y Rita 03] P. Bach y Rita & S. W. Kerchel. *Sensory substitution and the human-machine interface*. Trends in Cognitive Sciences, vol. 7, no. 12, pages 541–546, 2003.
- [Yoshida 05] Akiko Yoshida, Volker Blanz, Karol Myszkowski & Hans-Peter Seidel. *Perceptual evaluation of tone mapping operators with real-world scenes*. In Human Vision and Electronic Imaging(HVEI), SPIE, vol. 5666, pages 192–203, 2005.