HAL
open science

Variational models for color image processing in the RGB space inspired by human vision Mémoire d'Habilitation a Diriger des Recherches dans la spécialité Mathématiques

Edoardo Provenzi

To cite this version:

Edoardo Provenzi. Variational models for color image processing in the RGB space inspired by human vision Mémoire d'Habilitation a Diriger des Recherches dans la spécialité Mathématiques. Signal and Image processing. ED 386 : École doctorale de sciences mathématiques de Paris centre, UPMC, 2016. tel-01339682

HAL Id: tel-01339682

https://hal.science/tel-01339682v1

Submitted on 7 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Variational models for color image processing in the RGB space inspired by human vision

Edoardo Provenzi

Mémoire d'Habilitation à Diriger des Recherches dans la spécialité Mathématiques

Université Paris Descartes

École doctorale de sciences mathématiques de Paris centre, UPMC

Soutenance HDR : 14 Juin 2016

Composition du jury:

- Julie Delon (présidente) - Professeur à l'Université Paris Descartes
- Jean-François Aujol (rapporteur) - Professeur à l'Université de Bordeaux
- Céline Loscos (rapporteur) - Professeur à IUT de Reims-Châlons-Charleville
- Frederic Dufaux - DR Télécom ParisTech
- Pascal Mamassian - DR ENS-Ulm, Paris
- Laurent Perrinet - CR-HDR à l'Institut de Neurosciences de la Timone
- Catalina Sbert - Professeur à l'Université de Palma

Preface

In this document I provide the basic information about my research path. I will start with a brief overview that has the aim to provide a compact resume of the results that will be discussed in deeper detail in the subsequent chapters.

Since a significant part of my research lies in the intersection between phenomenology of human vision and image processing, I will first present the basic information about the human visual system that will be used later.

The variational models for color image processing that I contributed to develop find their foundation in an important result due to Caselles and Sapiro, i.e. the variational interpretation of histogram equalization as the minimization of a functional given by two terms in opposition to each other: the first is minimized when the dispersion around the average value is as small as possible, while the second is minimized when the global contrast of the image is as large as possible.

After discussing the Caselles and Sapiro result, I will show in chapter 4, the most important of this document, how their functional can be modified to approach the operations of the human visual system. The modified energies will be called 'perceptual functionals'.

The consequences of this modification are two-fold. On one side, I will show that famous existing algorithms for perceptual-inspired enhancement of digital images, as e.g. Retinex, can be interpreted as particular instances of this variational framework. On the other side, thanks to suitable numerical approximations, I will show that it is possible to reduce the high computational complexity of these algorithms, thus allowing their applications to high resolution images in a reasonable time. The chapter will be concluded by the wavelet version of the perceptual functionals, which allows a further computational complexity reduction.

The following chapter will be dedicated to high dynamic range (HDR) images, i.e. matrices of values proportionals to the real radiance of a visual scene. The problems discussed are both the generation of HDR images in the presence of motion and the tone mapping, i.e. the reduction of the dynamic range for displaying properties.

The subsequent chapter will be devoted to natural image statistics, in particular to the spatiochromatic properties of natural environment and their relation with human vision.

The last research chapter will show how variational principles can be successfully applied to the problem of histogram transfer and fusion of color images. The flexibility of the variational framework presented allows handling multiple images at once and it opens the possibility to be integrated with the perceptual formalism introduced above.

Finally, I will conclude with a chapter dedicated to the future perspectives of my research.

Dedicated to the memory of Prof. Vicent Caselles, with infinite gratitude for sharing his deep intelligence, culture, passion for science and, last but not least, intellectual honesty.

Contents

1 Overview of the research performed **7**
1.1 Formalization of Retinex-like algorithms and variational framework for perceptual color correction models . . . 7
1.2 HDR imaging and tone mapping . . . 10
1.3 Histogram transfer . . . 11
1.4 Natural image statistics . . . 11

2 Rudiments about human visual system (HVS) features **14**
2.1 The retina . . . 14
2.1.1 Photoreceptors: rods and cones . . . 15
2.2 Adaptation and photo-electrical response of receptors . . . 15
2.3 Locality of vision . . . 16
2.4 Local contrast enhancement . . . 17
2.5 Physical vs. perceived light intensity contrast: Weber-Fechner's law . . . 19
2.6 Color constancy . . . 21

3 Variational formulation of histogram equalization **22**
3.1 The Caselles-Sapiro model . . . 22
3.2 Interpretation of Caselles-Sapiro's functional for histogram equalization . . . 26

4 Perceptually-inspired variational models for color enhancement in the RGB space **27**
4.1 Beyond the Caselles-Sapiro model: Modification of histogram equalization functional to approach visual properties . . . 28
4.1.1 A contrast term coherent with HVS properties . . . 28
4.1.2 Entropic adjustment term . . . 30
4.2 Minimization of perceptual functionals . . . 31
4.2.1 Stability of the iterative semi-implicit gradient-descent scheme . . . 34
4.2.2 A general strategy for the reduction of computational complexity . . . 35
4.2.3 Results . . . 36
4.3 Embedding existing perceptually-inspired color correction models in the variational framework . . . 38
4.4 Perceptual enhancement in the wavelet domain . . . 39
4.4.1 Adjustment to the average value in the wavelet domain . . . 41
4.4.2 Local contrast enhancement in the wavelet domain . . . 41

5 High dynamic range (HDR) imaging **46**
5.1 Generation of HDR images in non-static conditions 46
5.2 HDR generation based on gradient fusion 49
5.2.1 Radiance Registration 49
5.2.2 Refinement step 50
5.2.3 Gradient Fusion 50
5.2.4 Results and comparisons 52
5.3 Tone mapping 53
5.3.1 A two-stage tone mapping 56

6 Statistics of natural images **60**
6.1 Brief state of the art in second order natural color image statistics 61
6.1.1 Chromatic redundancy in natural images 61
6.1.2 Spatio-chromatic redundancy in natural images 63
6.2 Relationship between second order stationarity and the decorrelated spatiochromatic features of natural images 65
6.2.1 The gray-level case 65
6.2.2 Spatial invariance of the covariance 66
6.2.3 The color case 67
6.3 Validations on natural image databases 69
6.3.1 Computation of the covariance matrices 69
6.3.2 Stability of the covariance computation with respect to the number of images and the image content 69
6.3.3 The chromatic autocovariance matrix $c^0(R, G, B)$ and its eigenvalues and eigenvectors 70
6.3.4 The exponential decay of spatiochromatic covariance matrix elements 71

7 Variational histogram transfer **75**
7.1 A short review of midway histogram equalization 76
7.2 A variational approach to histogram transfer 77
7.2.1 The L^2 distance between the cumulative histogram of a gray-scale image and a reference one and its first variation 78
7.2.2 The L^2 distance between the cumulative histogram of a vector-valued image and a reference one and its first variation 79
7.2.3 Symmetrization of the cumulation of histogram 80
7.3 Intermediate histogram equalization of a set of images 80
7.3.1 An energy term to approach the histogram of two images 80
7.3.2 Attachment to data and to the image geometry 81
7.3.3 A model for intermediate histogram equalization of a set of images . . 81
7.4 Experiments 82
7.4.1 Histogram transfer 82
7.4.2 Intermediate histogram equalization of a set of images 84

8 Future perspectives **89**
8.1 Collaboration with cognitive psychologists and psycho-physicists 89
8.1.1 Variational interpretation of context-driven effects in higher-level cognition 89

8.1.2 Development of non-reference quality measures based on human visual properties 90
8.1.3 Analysis of further HVS features in terms of the variational framework 90
8.1.4 Color naming, metamer mismatching and color constancy 91
8.2 Differential geometry of the space of perceived colors 91
8.3 Future research in statistics of natural images and multispectral images . . . 92
8.3.1 Generation of a large unbiased database of multispectral and HDR images 92
8.3.2 Analysis of the effects of HVS transformations on the statistical results 93
8.4 Analysis of problems related to astrophysical images 93
8.5 Development of new pedagogical strategies for scientific teaching 94

Chapter 1

Overview of the research performed

For the sake of a clearer exposition, I will introduce in this chapter a brief discussion about the research achievements that will be discussed with more detail throughout the following chapters.

1.1 Formalization of Retinex-like algorithms and variational framework for perceptual color correction models

The Retinex model of color vision [62] has been proposed by Land and McCann with the will to understand the mechanisms underlying *color constancy*, i.e. the robustness of color perception with respect to changes of illumination and its dependence on spatial context.

Interpreting a series of brilliant experiments with colored sharp-edged patchworks called *Mondrians*, see Figure 1.1, Land and McCann came out with the idea that the human visual system can discount the illumination by performing comparisons among luminance information coming from different points of the scene.

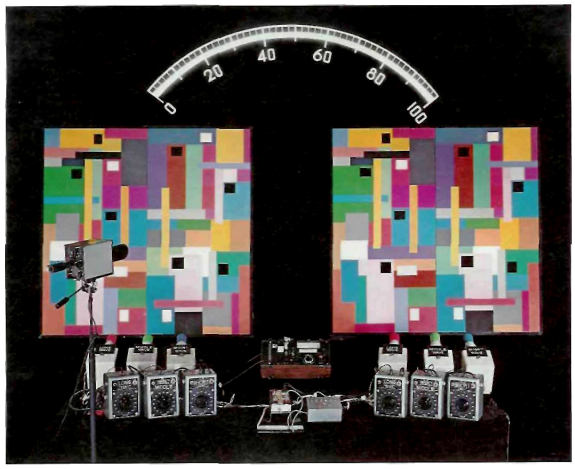


Figure 1.1: The colored Mondrian experiment. Taken from [61].

Moreover, they claimed that these comparison are performed in the separated chromatic channels corresponding to the three distinct cone responses, see chapter 2 for more details.

In order to implement their comparisons, they considered ratios of pixel intensities along *paths* wandering around the digital image representing the visual scene.

The original Retinex model was not written in a rigorous mathematical way and this fact generated a plethora of Retinex interpretations that, even though very interesting, in several cases have little in common with the original ideas that originated the model.

For this reason, in [93], my first work in color image processing, I collaborated at providing the mathematical translation of the original Retinex algorithm described by Land and McCann in [62], pointing out the two mechanisms introduced by the authors: threshold and reset.

As underlined by McCann himself, see e.g. [68], the reset is the heart of the Retinex algorithm, and, if one sets the threshold to zero, then the mathematical formula representing Retinex reduces simply to

$$I_{\text{Retinex}} = \frac{I(x)}{\max_{y\in\gamma}\{I(y)\}} \tag{1.1}$$

where I is the image intensity in a given chromatic channel, x is a fixed pixel and γ is a path ending in x.

Thus, the reset-free original Retinex is proven to be nothing but a local version of the von Kries algorithm [120], locality being given by the use of paths.

In [95], the use of paths has been criticized with a topological argument, which favored the use of the so-called *local sprays*, i.e. sparse sampling of two-dimensional kernels, as in Figure 1.2. The resulting algorithm was called RSR for Random Spray Retinex.

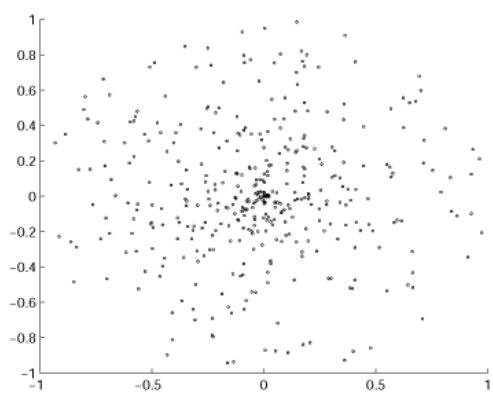


Figure 1.2: A localized spray taken by [95].

The local spray technique has been also used to speed-up another perceptually inspired color processing algorithm, called ACE [102], and to fuse it with RSR in [96].

The proliferation of Retinex-like color correction models over the years and the lack of standard measures for an objective judgment of their performances put in evidence the need to build a mathematical framework where the features of these methods can be analyzed and compared.

With this aim, in the series of papers [15, 77, 14, 92], I collaborated in the construction of a variational framework that has allowed a clear comprehension of the most fundamental intrinsic characteristics of such methods. Another important result is the possibility to reduce

their computational complexity to obtain algorithmic implementations which can be used to filter high resolution images.

The starting point for this research was the identification of the profound similarities, and also of the substantial differences, between histogram equalization and the ACE model. This was possible thanks to the very important result of Caselles and Sapiro [107] about the variational interpretation of histogram equalization. These authors have proven that the argmin I^* of this functional

$$E_{\text{hist eq}}(I) \equiv 2 \int_\Omega \left(I(x) - \frac{1}{2} \right)^2 dx \; - \; \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)| \, dxdy. \tag{1.2}$$

is an image with equalized histogram, i.e. $H(I^*(x)) = I^*(x)$ for all $x \in \Omega$, where Ω is the spatial support of the image, $|\Omega|$ is its area, $x, y \in \Omega$ are pixel positions and H represents the cumulative histogram. Moreover, they have shown that initial value problem for the gradient descent

$$\begin{cases} \partial_t I = -\delta E_{\text{hist eq}}(I) \\ I(0) = I_0 \end{cases}$$

where I_0 is the original image, has a unique solution.

The minimization of the first term of $E_{\text{hist eq}}(I)$ pushes the image towards $1/2$, the center of the normalized dynamic range $[0, 1]$, while, due to the presence of the minus sign, the minimization of the second term implies an amplification of differences among pixel intensities, i.e. a global contrast enhancement.

The balance expressed by the minimization of the functional $E_{\text{hist eq}}(I)$ is reminiscent of what happens in the human visual system: on one side we tend to adjust to the average luminance of the scene that we are observing, but, on the other side, we increase local contrast.

Besides these two properties, there exists also the already mentioned feature of color constancy and the Weber-Fechner's law, i.e. the logarithmic response of human senses to intensity changes of external stimuli.

As proven in [77], when all these features are kept into account simultaneously, there remains only one class of functionals which can incorporate all of them, namely $E^{\varphi}_{w,\alpha,\beta,I_0,\mu} = C^{\varphi}_{w} + D_{\alpha,\beta,I_0,\mu}$, where:

$$C^{\varphi}_{w}(I) = \iint_{\Omega^2} w(x, y) \, \varphi \left(\frac{\min\{I(x), I(y)\}}{\max\{I(x), I(y)\}} \right) dxdy \tag{1.3}$$

$\varphi : [0, 1] \to [0, 1]$ being a positive non-decreasing differentiable function, $w : \Omega \times \Omega \to [0, 1]$ a normalized local kernel, and:

$$\begin{aligned} D_{\alpha,\beta,I_0,\mu}(I) \equiv \alpha \int_\Omega & \left[\mu \log \frac{\mu}{I(x)} - (\mu - I(x)) \right] dx \\ & + \beta \int_\Omega \left[I_0(x) \log \frac{I_0(x)}{I(x)} - (I_0(x) - I(x)) \right] dx, \end{aligned} \tag{1.4}$$

μ being the average intensity of I over Ω.

The minimization of $C^{\varphi}_{w}(I)$ provokes a local (due to w) contrast enhancement independent on changes of illumination (due to the fraction), while the minimization of $D_{\alpha,\beta,I_0,\mu}(I)$ reduces the dispersion of levels around the original image values $I_0(x)$ and around the average μ.

The motivations for this particular type of functional will be explained in detail in chapter 4. Here it is worthwhile underlying that this variational framework allowed explaining in a clear way the action of perceptually-inspired color correction algorithms as the original Retinex and ACE, see [14, 15], respectively.

Moreover, thanks to numerical approximations, it is possible to reduce the computational complexity to the cumbersome $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ and even to $\mathcal{O}(N)$ in the case of the wavelet-based formulation of this framework, see [92].

In Figure 1.3 it can be seen how dramatic is the changes in the performance of the algorithm derived by the histogram equalization functional and by its perceptual modification.

Figure 1.3: *Left*: original image. *Center*: color enhancement obtained via gradient descent minimization of a perceptual functional. *Right*: effect of histogram equalization. Image courtesy of Peter Greenspun.

1.2 HDR imaging and tone mapping

High Dynamic Range (HDR) images are matrices of data proportional to the radiance of a visual scene, in opposition to the more common Low Dynamic Range (LDR) images provided by a typical digital camera, which can represent only a limited amount of the total radiance range.

While already used by specialists, in 1997 HDR images became popular thanks to the breakthrough paper of P. Debevec and J. Malik [23], which allowed a fast generation of HDR images of static scenes starting from a stack of LDR images taken with a tripod at different time exposures. This achievement generated a whole new research field called *HDR imaging*.

Debevec and Malik's technique relies on a suitable fusion of the most salient information of each LDR image of the stack: the ones taken with a short exposure time will show details in the brightest areas of the image, while the ones taken with a long exposure time will provide valuable information for the darkest image areas. The generated HDR image stores the details in every part of the matrix.

When movement is involved in the acquisition of the stack of LDR images, translucent artifacts called 'ghosts' are likely to appear in the final HDR image. In [33] I collaborated at developing a technique to strongly reduce this kind of artifacts based on a suitable fusion of local gradient information.

Another open problem in HDR imaging is *tone mapping*, which consists in finding a local and non-linear transformation to shrink the range of the HDR image to the two orders of

magnitude that a common LDR screen can handle. The aim of tone mapping is to preserve as much as possible the detail visibility and the color sensation of the real scene.

I co-authored a paper [32] in which the tone-mapping technique has been built as an original interplay among the formulae that govern the transduction of electromagnetic energy to neuron activity (discussed in chapter 2) and the variational techniques for perceptual contrast enhancement. This technique not only is coherent with the path of visual signals in the human visual system, but it has also proven to outperform the state of the art algorithms with respect to the HDR image quality measure proposed in [9].

1.3 Histogram transfer

Histogram transfer is a very important technique in image processing, in which two or more image histograms are non-linearly modified in order to approach a specified one.

Applications of histogram transfer can be found in color matching of images taken with different time exposures or with different illumination conditions, flicker stabilization of video sequences and numerical comparison of images in order to extract mutual information (e.g. depth or disparity computation in stereo images, or optical flow computations).

I co-authored the paper [78], in which we have proposed an energy functional whose minimization induces histogram transfer of two or more color images, while maintaining colors and geometry as close as possible to the original ones. The functional has three terms: $\mathcal{E} = \mathcal{E}_M + \mathcal{E}_D + \mathcal{E}_S$, where:

- the minimization of $\mathcal{E}_M$ moves closer the histograms, by reducing the L^2 distance between them;
- the minimization of $\mathcal{E}_D$ preserves colors providing an attachment to the original tones;
- the minimization of $\mathcal{E}_S$ preserves the geometry of the images so that the transformed ones share the same level lines.

This method can also transfer the histograms to a midway histogram that does not need to be specified in advance, but it is automatically generated by the set of original images.

Figure 1.4 shows an example of histogram transfer provided by the variational method just described.

1.4 Natural image statistics

The papers [8, 11] pioneered the idea that the human visual system has evolved in order to optimize the elaboration and transmission of visual signals coming from natural scenes by getting rid of redundant information. In fact, there are a hundred million retinal photoreceptors against one million optic nerve neurons, therefore the retinal output must be re-coded to be sure that the salient information passes to subsequent stages of the visual process.

There are two types of redundancies: spatial and chromatic. The first is due to the fact that nearby points in a natural scene are very likely to send very similar information, unless they lie in the proximity of a sharp edge. The second is due to the fact that the spectral sensitivity curves of the retinal cones overlap a lot, in particular those of the L and M cones, see chapter 2 for more details.

Figure 1.4: First row: Original images. Second row: Final images, their histograms have been moved to a common midway histogram.

One of the most influential papers on chromatic correlation is that of Buchsbaum and Gottschalk [18], where the authors approached the problem of finding uncorrelated color features from a purely theoretical point of view. They analyzed the problem of an efficient post-retinal information transmission by performing a Principal Component Analysis (PCA) on the Long, Medium and Short (LMS) cone activation values. Given the spectral distribution of a visual signal, these values are obtained by performing a weighted average, with weights given by the spectral sensitivity functions of the L,M,S cones.

Buchsbaum and Gottschalk's theory has been successfully tested by Rudermann, Cronin and Chiao in [106] by using data coming from a database of 12 multispectral natural images. Most importantly, they had the idea to study the spatio-chromatic redundancy reduction by considering 3×3 patches, treating the 9 patch pixels as a random vector, with each pixel containing a 3-vector color information. They converted every patch in a vector with 27 components to be analyzed with the PCA. The principal axes of these small patches show fluctuations in the achromatic channel, followed by blue-yellow and red-green ones. The pattern of the fluctuations is Fourier-like, as shown in Figure 1.5. No rigorous mathematical explanation of this empirical observation was available.

In the paper [94] that I have co-authored, it has been proven that separated Fourier os-

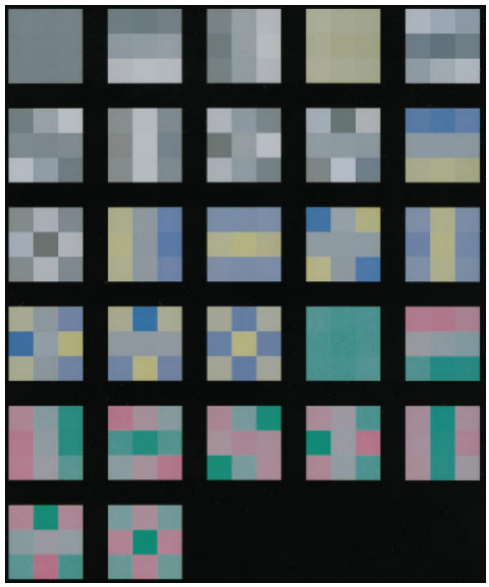


Figure 1.5: Principal axes of patches of dimension 3×3 arranged in order of decreasing eigenvalue, from left to right, top to bottom (adapted from [106], page 2041).

cillations in the achromatic plus color opponent channels are a direct consequence of the translation-invariance of the covariance, i.e. by second order stationariety of natural images, plus two quite technical requirements that we have proven to be satisfied with good approximation by the natural images of two large databases.

These hypotheses imply that a natural color image I can be decomposed on an orthonormal basis given by the tensor product of the achromatic plus color opponent axes (A, P, Q), times the 2D Fourier basis $\mathbf{e}_{m,l}$, as in Formula 1.5:

$$I = \sum_{m=0}^{W-1} \sum_{l=0}^{H-1} \langle I, (A, P, Q) \otimes \mathbf{e}_{m,l} \rangle (A, P, Q) \otimes \mathbf{e}_{m,l}. \tag{1.5}$$

Another achievement of the paper [94] is that we have proven that spatiochromatic covariance decay is exponential, and not a power-law as believed until the end of last century.

As seen in all the previous brief chapter descriptions, several results of my research are based on human visual system features. For this reason, I will present them in the next chapter, that will serve as a basis for the following ones.

Chapter 2

Rudiments about human visual system (HVS) features

In this chapter I recall just the basic facts about the retinal and brain elaboration of a visual signal. The purpose is not to provide an exhaustive treatise, but to introduce some important concepts and formulae that we will have a fundamental role in the development of the models described in chapters 4, 5 and 6 A very clear reference for this chapter is the book [30].

2.1 The retina

Figure 2.1 illustrates a human eye and the cross-sectional representation of its retina.

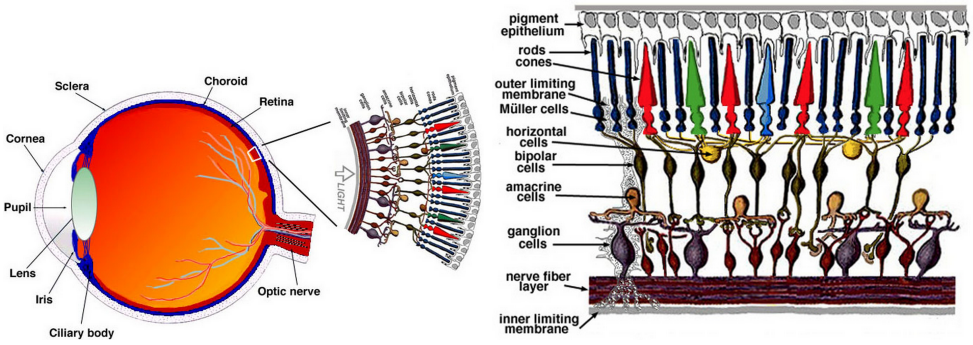


Figure 2.1: *Left*: a human eye. *Right*: cross section of a human retina. Courtesy of [57].

The retina includes several layers of neural cells, beginning with around 130 millions photoreceptors, rods and cones, and ending with about 1 million of ganglion cells. The specific processing that occurs in each type of cell is complex and not yet completely understood.

What we surely know is that retinal cells may respond non-linearly to stimuli and they are connected via links called *synapses*, which are able to perform mathematical operation such as addition, subtraction, multiplication, division, amplification and gain control. Considered as a whole, these operations result in a clever and sophisticated modification of the visual input.

Among all retinal cells, the most important for our purposes are the photoreceptors, rods and cones, to which the next subsection is devoted.

2.1.1 Photoreceptors: rods and cones

Rods and cones derive their respective names from their shape. Rods respond to low luminance levels, typically less than 10^{-3} cd/m^2, the so-called *scotopic region*, while cones repond to luminance levels higher than 10 cd/m^2, called *photopic region*. In the intermediate range, the so-called *mesopic region*, both rods and cones are activated, but with less ability to work.

In the photopic region the rods are saturated and only the cones work, while in the scotopic region the cones are not yet activated. From now on, we will only consider the photopic region and thus the properties of cones.

Color vision in the photopic region is allowed thanks to the existence of three types of cone receptors with peak spectral sensibilities spaced along the visual spectrum, see Figure 2.2. This is due to the existence of three slightly different molecular structures. The three types of cones are referred to as L, M, and S cones. These names refer to the *long*, *middle*, and *short* wavelength sensitive cones, picked at 560 nm, 530 nm and 420 nm, respectively.

The LMS cones are commonly referred to as the RGB cones. Of course, this is suggestive of red, green, and blue sensitivities but, as can be seen in Figure 2.2, this is an abuse of language, in particular because the L cones are picked in the region of monochromatic green-yellow, and not red.

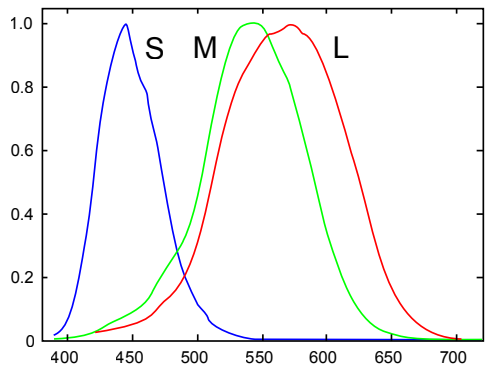


Figure 2.2: The normalized spectral sensitivity functions of the L,M,S cones.

Notice that, while in most physical imaging systems, in particular digital cameras, responsitivities of sensors are separated (see e.g. [53]), the spectral sensibilities of the three cone types are broadly overlapping, in particular those of the L and M cones. This remark will have a particular importance in chapter 6, when chromatic redundancy will be discussed.

The distribution of cones in the retina is not uniform: S cones are relatively sparse and completely absent in the most central area of the *fovea*, the central part of the retina with highest density L and M cones.

2.2 Adaptation and photo-electrical response of receptors

Light adaptation is the name used to describe the fact that the HVS is able to adapt to different light intensities in order to allow detail perception over a range of 10 orders of magnitude.

Before reaching a photoreceptor, rod or cone, light intensity is reduced by the cornea, crystalline lens, the humors and the macula. Moreover, when a light photon is absorbed by

a photoreceptor, a transduction occurs: the electromagnetic energy carried by the photon is passed to the photoreceptor, which changes the electric potential of its membrane. The empirical law that describes the photoreceptor transduction is known as *Michaelis-Menten's equation* [109]:

$$r(I) = \frac{\Delta V}{\Delta V_{\max}} = \frac{I^{\gamma}}{I^{\gamma} + I_S^{\gamma}}, \tag{2.1}$$

where $\Delta V_{\max}$ is the highest difference of potential that the membrane can handle, γ is a constant (measured as 0.74 for the rhesus monkey), I is light intensity and, in particular, I_S is the value at which the photoreceptor response is half maximal, called the *semisaturation level*. When $\gamma = 1$, eq. (2.1) is called *Naka-Rushton's equation*.

I_S is usually associated with the *level of adaptation* of the visual scene where the human observed is embedded. As said before, each type of cone is most sensitive over a particular waveband and the semisaturation constant depends on the amount of light in the particular waveband that reaches it, not on the global luminance of the light source.

The photo-electrical response of photoreceptors, along with other phenomena occurring mainly in the retina, is considered one of the main explainations for the property of *adaptation to the average luminance level* of the HVS. In fact, after the photoreceptors transduction, the dynamic range is centered in $r(I_S) = 1/2$, as can be seen in Figure 2.3, which shows the non-linear compressive effect of eq. (2.1).

Notice that the adaptation property of the HVS is crucial: without it the range of operativity of our vision would be much narrower and sight as we know it would be impossible.

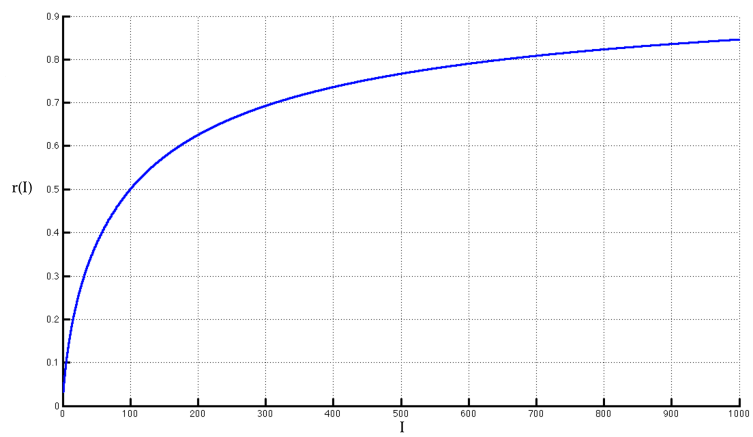


Figure 2.3: Compressive effect of Michaelis-Menten's equation in arbitrary units between 1 and 1000 and with $\gamma = 0.74$. The semisaturation constant I_S has been arbitrarily set to 100.

2.3 Locality of vision

Transduction curves as in Figure 2.3 represent the very first stage of visual processing. The electrical signals generated by the photoreceptors are then processed by the retinal neurons, synapses and ganglion cells, until finally transmitted to the brain via the optic nerve. In the brain, the visual signal is elaborated in several zones each of which is devoted to process different characteristics, e.g. shape, orientation, spatial frequency, size, color, motion, and so on [126].

Our present understanding of post-photoreceptors physiological operating principles is far from being precise: not only the brain, but also the retinal functions still present some unknown features.

For the purposes of this memoire, what is important to stress about post-photoreceptor physiology is that the signals transmitted from the photoreceptors to higher levels of the visual path are not simple point-wise representations, but they consist of sophisticated combinations of the receptors responses to photons coming from different parts of the visual scene.

Even when we fixe a single point, our eyes are constantly moving and capturing light information from all over the visual scene. These movements are called *saccadics* and they are the fastest of our body.

A conventional nomenclature has been introduced to rigorously define the local neighborhood of a point in a visual field, see e.g. [50] and [30]:

- **Stimulus**: is the visual element of about 2° angular subtense corresponding to foveal vision;
- **Proximal field**: is the immediate environment of the stimulus considered, extending isotropically for about 2° from the edge of the stimulus;
- **Background**: is defined as the environment of the stimulus considered, extending isotropically for about 10° from the edge of the proximal field;
- **Surround**: is a field outside the background. In practical situations, the surround can be considered to be the entire room or the environment in which a scene is viewed;
- **Adapting field**: is the total environment of the stimulus considered, the proximal field, the background, and the surround, until the limit of vision in all directions.

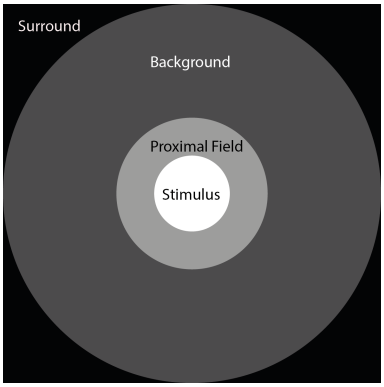


Figure 2.4: Components of the adaptive field.

2.4 Local contrast enhancement

The eye's optical system and the response of photoreceptors strongly reduces the range of light intensity that can be processed. To *compensate* this reduction, the HVS has developed a

system to enhance contrast perception already in the retina with lateral inhibition and further in the brain with higher perceptual features.

Ganglion cells are among the most studied retinal cells and we know that they spontaneously *fire* action potentials[1] at a base rate while at rest. *Excitation* of retinal ganglion cells results in an increased firing rate while *inhibition* results in a depressed rate of firing. For this reason the magnitude of the signal is represented in terms of the *number of spikes of voltage per second fired by the cell* rather than by the voltage difference across the cell membrane.

To represent the physiological properties of these cells, it is useful to consider the concept of *receptive field*, which is a *graphical representation of the area in the visual field to which a given cell responds*. The positive or negative response is indicated in the receptive field, as in Figure 2.5, which represents a prototypical receptive field for ganglion cells.

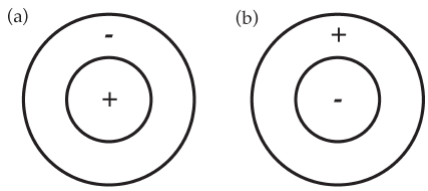


Figure 2.5: Typical center-surround antagonistic receptive fields: (a) on-center, (b) off-center.

Receptive fields illustrate *center-surround antagonism*: the receptive field in Figure 2.5 (a), called *on-center*, describes a positive central response, surrounded by a negative surround response, while Figure 2.5 (b), called *off-center*, illustrates a ganglion cell response of opposite polarity.

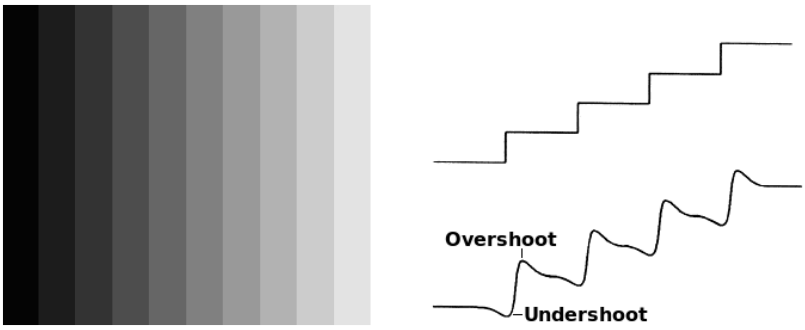


Figure 2.6: *Left*: Mach bands effect. *Right*: real and apparent luminance pattern.

The excitation-inhibition processing can explain some *local contrast enhancement* effects as Mach bands, see Figure 2.6: as we approach the vertical band on the right the gray level appears lighter, as we approach the vertical band on the left the gray level appears darker, in spite of the fact that the luminance value in each vertical gray band is constant.

Consider Figure 2.7: in the situation 1 excitation and inhibition fire signals equally, so a uniform patch is perceived; in the situation 2 a part of the inhibition component of the receptive field is activated by a region of highest luminance and so it prevails, generating a sensation of darker gray; on the contrary, in the situation 3, a part of the inhibition component

[1] In physiology, an *action potential* is a short-lasting event in which the electrical membrane potential of a cell rapidly rises and falls.

of the receptive field is activated by a region of lowest luminance and so it is dominated by the excitation component, which produces a sensation of lighter gray.

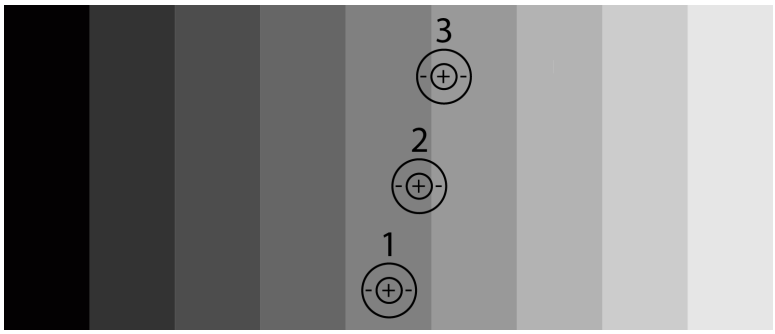


Figure 2.7: Excitation/inhibition explanation of the Mach band effect.

Another well known local contrast enhancement effect is simultaneous contrast, shown in Figure 2.8

Figure 2.8: Simultaneous contrast: even thought the two inner squares have the same luminosity, our perception is different, due to the different background.

2.5 Physical vs. perceived light intensity contrast: Weber-Fechner's law

In the previous section we have seen that the HVS rearranges spatial information in order to produce a peculiar response to spatial contrast. The HVS responds non-linearly also to intensity contrast of light coming from fixed positions.

Psychophysics is the science that aims at modeling in a mathematically rigorous way the magnitude of human perception in response of external stimuli. The German physicist E. Weber, with the results obtained in the second half of the 19th century, was one of the first scientists in history to develop some psychophysical experiments to test intensity contrast perception.

To avoid unwanted biases, he worked in a very constrained setting: a dark-adapted human observer was put in a dim room in front of a white screen on which a narrow beam of light was thrown in the center of the visual field. The luminous intensity I of the beam was

increased very slowly and the observer was asked to tell whether he/she could perceive an intensity change. The least perceptible intensity change ΔI is called *JND* for *Just Noticeable Difference*.

Weber found out that the JND increased proportionally with the luminous intensity[2], i.e. $\Delta I/I = K$, called Weber's law, that can be re-written as $\Delta I = K \cdot I$, $K \simeq 0.08$ is called *Weber's constant*.

Weber's law says that, as we increase the background light I, the difference ΔI must increase proportionally in order to be able to appreciate $I + \Delta I$ as being different from I. This partially explains why we are more sensitive to noise in dark areas of a visual scene and thus why it is more important to perform a good denoising in dark areas of digital images rather than in bright ones. This last consideration is a practical application of a psychophysical phenomenon.

The founder of psychophysics, the German experimental psychologic G. Fechner, gave the following interpretation of Weber's law: he introduced the adimensional quantity $s(I)$ called light *sensation* and stated that the difference of sensation $\Delta s(I)$ is proportional to a slightly modified Weber's ratio, i.e.

$$\Delta s(I) = k\frac{\Delta I}{n + I} \tag{2.2}$$

where $k > 0$ is a constant and $n > 0$ is a quantity often interpreted as *internal noise* in the visual mechanism. Fechner transformed this finite difference equation into the following differential equation [125]:

$$ds(I) = k\frac{dI}{n + I}, \tag{2.3}$$

by integrating both sides from I_0, the threshold above which luminous intensity is perceivable, i.e. such that $s(I_0) = 0$ and $s(I_0 + \varepsilon) > 0$ for all $\varepsilon > 0$, to a generic value of I, we obtain

$$\int_{I_0}^{I} ds(I) = k\int_{I_0}^{I} \frac{dI}{n + I} \iff s(I) - s(I_0) = k[\log(n + I) - \log(n + I_0)], \tag{2.4}$$

by using the properties of the logarithm and using the fact that $s(I_0) = 0$ we get the so-called *Weber-Fechner's law*:

$$s(I) = k\log\left(\frac{n + I}{n + I_0}\right) = s_0 + k\log(n + I), \tag{2.5}$$

where $s_0 = -k\log(n + I_0)$.

Weber-Fechner's law says that *the sensation of luminous differences*, in the very constrained context of Weber's experiment, *grows at the logarithm of the luminous intensity*.

We must stress the *limitations of Weber-Fechner's law*:

1. Firstly, it is valid only for very simple visual scenes, as those considered by Weber in his experiments, because, as discussed in the previous section, the presence of a non-trivial spatial context may change significantly contrast perception;

2. Secondly, even for very simple visual scenes, Fechner's assumption that we can maintain the validity of Weber's law passing from finite to infinitesimal light intensity differences

[2]Weber's law is approximately valid not only for the visual sense, but also for all the other senses, with different values of Weber's constant.

is correct only for luminous intensities intermediate between the minimum and the maximum perceivable light. As we approach these extreme situations, this assumption fails dramatically due to strong non-linearities in the visual mechanism, so that Weber-Fechner's law doesn't hold anymore.

2.6 Color constancy

The last HVS feature that it is important to recall is color constancy, already defined as the ability of human observers to perceived colors very robustly with respect to changes of illumination.

The precise causes behind color constancy are not yet known. Phenomenologically speaking, it is known that cognition and physiological effects as color memory and chromatic adaptation, respectively, play a major role in color constancy [30].

Color memory consists in the fact that when our brain recognize an object that usually has a prototypical color, it tends to automatically associate that color to the object.

To explain how chromatic adaptation works it is worth considering the typical example of a white paper illuminated by different illuminants, e.g. daylight, fluorescent, and incandescent. Daylight contains more short-wavelength photons than fluorescent light, and the incandescent one contains more long-wavelength photons than fluorescent light.

The paper retains approximately its white appearance because the cones are able to vary independently the height of their spectral sensitivity curves, so that the S-cones become relatively less sensitive under daylight to compensate for the additional short-wavelength photons and the L-cones become relatively less sensitive under incandescent illumination to compensate for the additional long-wavelength photons.

The von Kries diagonal model [120] represents chromatic adaptation through the following diagonal matrix:

$$K = \mathrm{diag}(k_L, k_M, k_S) \tag{2.6}$$

where k_c, $c \in \{L, M, S\}$ are the adaptation coefficients. The von Kries transformation is used to implement the *simplest white-balance* in digital image processing with the choice of the values $k_c = \max_{x\in\Omega}\{I_c(x)\}$, $c \in \{R, G, B\}$.

For the purposes of chapter 4, it is significant to end this chapter by recalling the simplest image formation model and the consequences of color constancy on it.

Let λ represent the intensity of a uniform illuminant and let $\rho(\xi)$ be the the reflectance of an object in the physical point ξ. If x is the projection of ξ on the camera sensor, then the simplest way to interpret the intensity I of the pixel x in terms of illuminant and reflectance is by taking their product, i.e.

$$I(x) = \lambda \cdot \rho(x). \tag{2.7}$$

Of course this image formation model is very simplified, but it is often considered as a good starting point because it contains the essential ingredients of any other image formation model, for more details about this topic see e.g. [38].

A perfect color constancy would imply the invariance of $I(x)$ with respect to changes of λ, for all $x \in \Omega$, i.e. that $\lambda_1\rho(x) \sim \lambda_2\rho(x)$, $\forall\lambda_1 \neq \lambda_2$, where $\sim$ means perceptual match.

In the following chapter I will present the variational interpretation of histogram equalization provided by Caselles and Sapiro in [107], then, in chapter 4 I will fuse the information of the present chapter with Caselles-Sapiro variational model to build a perceptual variational framework for color correction.

Chapter 3

Variational formulation of histogram equalization

Variational principles amount at defining a suitable functional, i.e. a scalar-valued function defined on a certain functional space, so that its minima (ideally, its *unique* minimum) provide the optimal solution of the problem under analysis.

In 1997, Caselles and Sapiro used variational principles to give a novel interpretation of histogram equalization of a digital image. This work is not only a profound achievement by its own, but it is also the main theoretical result on which the variational model of perceptual enhancement of color images is based. For this reason, this entire chapter will be dedicated to the description and interpretation of Caselles-Sapiro's model of histogram equalization.

3.1 The Caselles-Sapiro model

It is worthwhile starting with the notation that will be used from now on. The functional space that we will consider here is that of RGB continuous image functions. To introduce these functions we will first denote with $\Omega \subset \mathbb{R}^2$ the *spatial domain* of a digital image, with $|\Omega|$ its area and with $x \equiv (x_1, x_2)$ and $y \equiv (y_1, y_2)$ the coordinates of two arbitrary pixels in Ω. Unless otherwise specified, we will always consider a *normalized dynamic range* in $[0, 1]$, so that a *RGB image function* will be denoted with

$$\begin{array}{rccl} \vec{I}: & \Omega & \longrightarrow & [0,1] \times [0,1] \times [0,1] \\ & x & \mapsto & (I_R(x), I_G(x), I_B(x)) \end{array}$$

where each scalar component $I_c(x)$ defines the intensity level of the pixel $x \in \Omega$ in the red, green and blue channel, respectively.

We stress that we will perform every computation on the scalar components of the image, thus *treating each chromatic component separately.* Therefore, we will avoid the subscript c and write simply $I(x)$ to denote the intensity of the pixel x in a given chromatic channel.

From the point of view of functional analysis, we will implicitly consider the space of image functions as a subspace of $L^2(\Omega)$, the space of square-integrable (finite-energy) functions from Ω to $[0, 1]$.

Let us also recall very briefly what histogram equalization is. Let $\lambda \in [0, 1]$ be a generic

intensity level, then the *histogram* of I computed in λ is:

$$h(\lambda) = \frac{1}{|\Omega|} \text{Area}\{x \in \Omega \,|\, I(x) = \lambda\} \qquad \lambda \in [0,1], \tag{3.1}$$

i.e. the *occurrence probability* of the level λ in the image.

The *cumulative histogram* of I computed in λ, $H(\lambda)$, is:

$$H(\lambda) = \frac{1}{|\Omega|} \text{Area}\{x \in \Omega \,|\, I(x) \leq \lambda\} \qquad \lambda \in [0,1], \tag{3.2}$$

i.e. the *probability to find a pixel with intensity less than (or equal to)* λ.

Of course, the relationship between h and H is:

$$H(\lambda) = \int_0^\lambda h(t)\,dt, \;\; H'(\lambda) = h(\lambda), \tag{3.3}$$

i.e. *H is the integral function of h* in the interval $[0,1]$ and the first derivative of H in each level gives the histogram of that level.

It will be useful for later purposes to notice that the relationship $H(\lambda) = \int_0^\lambda h(t)\,dt$ can be re-written as follows

$$H(\lambda) = \int_0^\lambda h(t)\,dt = \frac{1}{|\Omega|} \int_0^1 \text{sign}^+(\lambda - I(t))\,dt \tag{3.4}$$

where

$$\text{sign}^+(s) = \begin{cases} 1 & \text{if } s \geq 0 \\ 0 & \text{if } s < 0 \end{cases}$$

and its *spatial version*

$$H(I(x)) = \frac{1}{|\Omega|} \int_\Omega \text{sign}^+(I(x) - I(y))\,dy. \tag{3.5}$$

An image is said to be *equalized* if *each level has the same occurrence probability*, i.e. if $h(\lambda) \equiv 1$ (recall that the histogram is normalized) $\forall\lambda$, which of course can be translated to the following condition on the cumulative histogram: $H(\lambda) = \lambda$, $\forall\lambda$.

It is easy to prove (see e.g. [40]) that the transformation from $[0,1]$ to itself given by $\lambda \longmapsto H(\lambda)$ is the easiest application that implements histogram equalization.

However, this is not the only histogram equalization transformation available in literature. In particular, a variational interpretation has been provided by Caselles and Sapiro in the paper [107]. For the purposes of this document, the Caselles-Sapiro's results can be resumed in the following theorem.

Theorem 3.1.1 *Given the functional*

$$E_{\text{hist eq}}(I) \equiv 2 \int_\Omega \left(I(x) - \frac{1}{2} \right)^2 dx \; - \; \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)|\,dxdy \tag{3.6}$$

if $I^* = argmin_I\, E_{\text{hist eq}}(I)$, then I^* has equalized histogram, i.e. $H(I^*(x)) = I^*(x)$ for all $x \in \Omega$.

Moreover, if I_0 is the original image, then the initial value problem for the gradient descent

$$\begin{cases} \partial_t I = -\delta E_{\text{hist eq}}(I) \\ I(0) = I_0 \end{cases}$$

has only one solution, where t is the evolution parameter of the iterative gradient descent scheme and the symbol δ represents the first variation of the functional.

$E_{\text{hist eq}}(I)$ will be called *Caselles-Sapiro functional* from now on. The entire proof of the theorem can be found in [107]. While the details of the proof of uniqueness are not inherent here, it is worthwhile reproducing the computation of the argmin of $E_{\text{hist eq}}(I)$. In fact, the tricks used in the computation of the Euler-Lagrange equations show explicitly the link with histogram equalization, which is far from being intuitive at a first sight.

First of all, let us start by recalling a classical lemma of variational calculus.

Lemma 3.1.1 *Given the two functionals*

$$E_1(I) = \int_\Omega \psi(I(x))\,dx, \quad E_2(I) = \iint_{\Omega^2} \phi(I(x), I(y))\,dxdy, \tag{3.7}$$

where ψ is a differentiable function defined on the codomain of I and ϕ is a differentiable function defined on the 2-th Cartesian power of the codomain of I, then their first variations are, respectively:

$$\delta E_1(I, J) = \int_\Omega \left.\frac{\partial \psi}{\partial I}\right|_{I(x)} J(x)dx \equiv \int_\Omega \psi'(I(x))\,J(x)\,dx \tag{3.8}$$

and

$$\delta E_2(I, J) = \iint_{\Omega^2} \left(\left.\frac{\partial \phi}{\partial I}\right|_{I(x)} J(x) + \left.\frac{\partial \phi}{\partial I}\right|_{I(y)} J(y) \right) dxdy, \tag{3.9}$$

J being a perturbation of I.

Then, notice that, by linearity, we can compute the first variation of the two terms appearing in $E_{\text{hist eq}}(I)$ and then add the results. For that, let us write:

$$D_{\frac{1}{2}}(I) = 2\int_\Omega \left(I(x) - \frac{1}{2}\right)^2 dx; \tag{3.10}$$

$$C(I) = \frac{1}{|\Omega|} \iint_{\Omega^2} |I(x) - I(y)|\,dxdy. \tag{3.11}$$

By virtue of formula (3.8), we have

$$\delta D_{\frac{1}{2}}(I, J) = \int_\Omega 4\left(I(x) - \frac{1}{2}\right) J(x)\,dx, \tag{3.12}$$

and thanks to formula (3.9), we have

$$\begin{aligned} \delta C(I, J) &= \frac{1}{|\Omega|} \iint_{\Omega^2} \left[\text{sign}(I(x) - I(y))J(x) - \text{sign}(I(x) - I(y))J(y)\right] dxdy \\ &= \frac{1}{|\Omega|} \left\{ \iint_{\Omega^2} \text{sign}(I(x) - I(y))J(x)\,dxdy - \iint_{\Omega^2} \text{sign}(I(x) - I(y))J(y)\,dxdy \right\}. \end{aligned} \tag{3.13}$$

Now, interchanging the role of the 'mute' variables x and y in the second integral of the last step, we have that

$$\frac{1}{|\Omega|}\iint_{\Omega^2} \text{sign}(I(x)-I(y))J(y)\,dxdy = \frac{1}{|\Omega|}\iint_{\Omega^2} \text{sign}(I(y)-I(x))J(x)\,dydx \tag{3.14}$$

but then, using the oddness of the sign function,

$$\frac{1}{|\Omega|}\iint_{\Omega^2} \text{sign}(I(x)-I(y))J(y)\,dxdy = -\frac{1}{|\Omega|}\iint_{\Omega^2} \text{sign}(I(x)-I(y))J(x)\,dydx. \tag{3.15}$$

Hence, we can write

$$\begin{aligned}\delta C(I,J) &= \frac{1}{|\Omega|}\left\{\iint_{\Omega^2} \text{sign}(I(x)-I(y))J(x)\,dydx \;+ \iint_{\Omega^2} \text{sign}(I(x)-I(y))J(x)\,dydx\right\}\\ &= \frac{2}{|\Omega|}\iint_{\Omega^2} \text{sign}(I(x)-I(y))J(x)\,dydx\end{aligned} \tag{3.16}$$

that can be conveniently rearranged as follows

$$\delta C(I,J) = \int_\Omega \left(\frac{2}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy\right) J(x)\,dx. \tag{3.17}$$

Now, since $\delta E_{\text{hist eq}}(I,J) = \delta D_{\frac{1}{2}}(I,J) - \delta C(I,J)$, by using formulas (3.12) and (3.17) we have

$$\delta E_{\text{hist eq}}(I,J) = \int_\Omega 4\left(I(x)-\frac{1}{2}\right)J(x)\,dx - \int_\Omega \left(\frac{2}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy\right)J(x)\,dx \tag{3.18}$$

i.e.

$$\delta E_{\text{hist eq}}(I,J) = \int_\Omega \left[4\left(I(x)-\frac{1}{2}\right) - \frac{2}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy\right]J(x)\,dx. \tag{3.19}$$

The stationary condition $\delta E_{\text{hist eq}}(I,J) = 0$, $\forall J$, implies that the expression in the square bracket must be zero, i.e.

$$\delta E_{\text{hist eq}}(I,J) = 0 \quad\Longleftrightarrow\quad 4\left(I(x)-\frac{1}{2}\right) - \frac{2}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy = 0, \tag{3.20}$$

so that the Euler-Lagrange equation relative to the energy functional $E_{\text{hist eq}}$ is the following implicit equation

$$2\left(I(x)-\frac{1}{2}\right) - \frac{1}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy = 0, \tag{3.21}$$

that can be suitably re-written as

$$\frac{1}{|\Omega|}\int_\Omega \text{sign}(I(x)-I(y))\,dy = 2I(x)-1. \tag{3.22}$$

Now, using the identity $\text{sign}(t) \equiv 2\text{sign}^+(t) - 1$, we can express the left-hand side of the Euler-Lagrange equation as

$$\begin{aligned}\frac{1}{|\Omega|}\int_\Omega (2\text{sign}^+(I(x)-I(y)) - 1)\,dy &= \frac{2}{|\Omega|}\int_\Omega \text{sign}^+(I(x)-I(y))\,dy - \frac{\int_\Omega dy}{|\Omega|}\\ &= 2H(I(x)) - 1,\end{aligned} \tag{3.23}$$

where we have used the fact that $\frac{1}{|\Omega|}\int_\Omega \text{sign}^+(I(x)-I(y))\,dy$ is the spatial version of the cumulative histogram $H(I(x))$, as noticed in eq. (3.5).

Thus, the Euler-Lagrange eq. (3.22) is equivalent to $2H(I(x))-1 = 2I(x)-1$, i.e. to $H(I(x)) = I(x)$, but then

$$\delta E_{\text{hist eq}}(I,J) = 0 \quad \Longleftrightarrow \quad H(I(x)) = I(x),\ \forall x \in \Omega, \tag{3.24}$$

which means that the image function I which satisfies the Euler-Lagrange equations of the functional $E_{\text{hist eq}}(I)$ has an equalized histogram, as it had to be proven.

3.2 Interpretation of Caselles-Sapiro's functional for histogram equalization

To interpretation of the energy functional $E_{\text{hist eq}}(I)$ in terms of image features is crucial for the following chapters. As remarked before, we can write $E_{\text{hist eq}}(I) = D_{\frac{1}{2}}(I) - C(I)$. Thus, the minimization of $E_{\text{hist eq}}(I) = D_{\frac{1}{2}}(I) - C(I)$ is achieved through the *minimization of* $D(I)$ and the *maximization of* $C(I)$.

The meaning of the two functional terms is the following:

- $D_{\frac{1}{2}}(I)$ is called *global quadratic dispersion* term around the middle gray level 1/2 and it is minimized when $I(x) \equiv 1/2$ for all $x \in \Omega$, i.e. the minimization of this term tends to turn I into an uniform gray image;

- $C(I)$ is called *global contrast* term and its maximization amplifies the global contrast of the image I, expressed by the set of absolute differences $|I(x) - I(y)|$.

Thus, the argmin of the Caselles-Sapiro functional is the image corresponding to the optimal balance between *two opposite effects*: on one side the minimization of $D_{\frac{1}{2}}(I)$ tends to *set all the levels to the average gray* 1/2 but, on the other side, the minimization of $-C(I)$ tends to *spread the intensity levels apart, as far as possible from each other*. So, the intrinsical meaning of Theorem 3.1.1 is that *the equilibrium among two conflicting actions, dispersion control around the middle gray and contrast enhancement, induces histogram equalization.*

One practical consequence of this result is that, applying for example the gradient descent technique to minimize $E_{\text{hist eq}}(I)$, one can stop the minimization process before reaching the complete equalization, thus realizing a *partial equalization* that can nonetheless be useful to avoid the typical over-enhancement of low-key images (see [40] for more details).

However, for the purposes of color image processing, the most important consequences of Theorem 3.1.1 are theoretical: in fact, as we will see in the next chapter, we can modify the functional $E_{\text{hist eq}}(I)$ in such a way that the basic balance principle of histogram equalization is preserved, but we can change the analytical form of the terms $D_{\frac{1}{2}}(I)$ and $C(I)$ taking inspiration from human visual perceptual features. The argmin image I^* of the modified functional is a color-corrected version of the original image driven by perceptual properties of the HVS.

Chapter 4

Perceptually-inspired variational models for color enhancement in the RGB space

Now that the basic properties of the HVS and the variational interpretation of histogram equalization have been introduced, it is possible to describe the variational framework for perceptually-inspired color image enhancement proposed in [77] and its consequences.

Before entering in the mathematical details, I consider worthwhile explaining the main similarities and differences between histogram equalization and HVS properties that brought us to the formulation of this model.

Let us start with similarities. First of all, after the light intensity reduction performed by the eye elements, the visual signal is transduced by photoreceptors to an electric signal centered around the semi-saturation level. This can be thought as the HVS process equivalent to the adjustment around the average gray level 1/2 in histogram equalization.

Secondly, the local contrast enhancement performed by the HVS opposes to the shrinking of light dynamic range, analogously to what happens in histogram equalization.

Regarding differences, the most important one to underline is that the human way to handle contrast is local, i.e. driven by the spatial context. Then, color constancy and Weber-Fechner's law are not taken into account in histogram equalization.

So, the aim of the work [77], and its related papers [15, 14, 92], was to modify the Caselles-Sapiro functional in order to make the new energy consistent with the HVS properties recalled before. It will be seen that, even working in the RGB space, the output images of the modified functional are qualitatively much better that those of histogram equalization.

Of course, in the literature, one can find several other proposals to modify histogram equalization which have comparable or even more efficient color correction properties than ours. The aim of our work was mainly theoretical, i.e. we wanted to test to what extent we can solely rely on a model of the HVS, and not on some ad-hoc mathematical and/or algorithmic requirements, to avoid, or at least strongly diminish, the problems related with RGB histogram equalization (and, in general separate waveband processing, as in multispectral imaging).

As it will be seen later, the framework that we have built is general enough to be used adapted also to other applications.

4.1 Beyond the Caselles-Sapiro model: Modification of histogram equalization functional to approach visual properties

The main problem in the modification of the Caselles-Sapiro functional is the determination of a contrast enhancement term which complies with the HVS features. Once determined it, the selection of the adjustment term can be done through dimensional coherence, as it will be explained in more detail in this section.

4.1.1 A contrast term coherent with HVS properties

Let us thus consider the problem of building a suitable a contrast measure $\overline{c}(a,b)$ between two gray levels $a,b>0$ (to avoid some singular cases, we shall assume that intensity image values are always positive). We require the contrast function $\overline{c}:(0,+\infty)\times(0,+\infty)\to\mathbb{R}$ to be continuous, symmetric in (a,b), i.e. $\overline{c}(a,b)=\overline{c}(b,a)$, increasing when $\min(a,b)$ decreases or $\max(a,b)$ increases.

Basic examples of contrast measures are given by $\overline{c}=|a-b|\equiv\max(a,b)-\min(a,b)$ or $\overline{c}(a,b)=\max(a,b)/\min(a,b)$.

Since our purpose is to enhance contrast by *minimizing* an energy, we define an *inverse contrast function* $c(a,b)$, still continuous and symmetric in (a,b), but decreasing when $\min(a,b)$ decreases or $\max(a,b)$ increases. Notice that, if $\overline{c}(a,b)$ is a contrast measure, then $c(a,b)=-\overline{c}(a,b)$ or $c(a,b)=1/\overline{c}(a,b)$ is an inverse contrast measure, so that basic examples of inverse contrast are: $c(a,b)=\min(a,b)-\max(a,b)$ or $c(a,b)=\frac{\min(a,b)}{\max(a,b)}$.

Let us now introduce a weighting function to localize the contrast computation. Let $w:\Omega\times\Omega\to\mathbb{R}^+$ be a positive symmetric kernel, i.e. such that $w(x,y)=w(y,x)>0$, for all $x,y\in\Omega$, that measures the mutual influence between the pixels x,y. The symmetry requirement is motivated by the fact that the mutual chromatic induction is independent on the order of the two pixel considered. Usually, we assume that $w(x,y)$ is a function of the Euclidean distance $\|x-y\|$ between the two points. We shall assume that *the kernel is normalized*, i.e. that

$$\int_\Omega w(x,y)=1\qquad\forall x\in\Omega. \tag{4.1}$$

Given an inverse contrast function $c(a,b)$ and a positive symmetric kernel $w(x,y)$, we define a contrast energy term by

$$C_w(I)=\iint_{\Omega^2} w(x,y)\,c(I(x),I(y))\,dxdy. \tag{4.2}$$

Thanks to the symmetry assumption, we may write $c(a,b)=\tilde{c}(\min(a,b),\max(a,b))$ for some function $\tilde{c}$ (indeed well defined by this identity). Notice that $\tilde{c}$ is non-decreasing in the first argument and non-increasing in the second one. The symmetry hypothesis is not restrictive, in fact, if the inverse contrast measure $c(a,b)$ were not symmetric, we could write it as the sum $c(a,b)=c_s(a,b)+c_{as}(a,b)$ where $c_s(a,b)$ and $c_{as}(a,b)$ are symmetric and anti-symmetric respectively. Since the sum $\iint_{\Omega^2} w(x,y)\,c_{as}(I(x),I(y))\,dxdy=0$, then the only remaining term is $\iint_{\Omega^2} w(x,y)\,c_s(I(x),I(y))\,dxdy$, hence we may assume that $c(a,b)$ is symmetric in (a,b).

In order to find out which properties the contrast term should satisfy, let us observe that, thanks to color constancy, an overall change in illumination, measured by the generic quantity $\lambda > 0$, does not affect the visual sensation. This requires the contrast function c to be homogeneous, recalling that c is homogeneous of degree $n \in \mathbb{Z}$ if

$$c(\lambda a, \lambda b) = \lambda^n \, c(a, b) \quad \forall \lambda, a, b \in (0, +\infty), \tag{4.3}$$

where a and b are synthetic representations of $I(x)$ and $I(y)$.

Of course, if $n = 0$, c automatically disregards the presence of λ, but we can say more: since λ can take any positive value, if we set $\lambda = 1/b$, we may write equation (4.3) as:

$$c(a, b) = b^n \, c\left(\frac{a}{b}, 1\right) \quad \forall a, b \in (0, +\infty), \tag{4.4}$$

so, when $n = 0$, $b^n = 1$ and c results as a function of the ratio a/b which intrinsically disregards overall changes in light intensity. If $n > 0$, then λ has a global influence and could be removed performing a suitable normalization (for instance, dividing by the n-th power of the highest intensity level).

We summarize these observations by saying that we assume that the inverse contrast function $c(a, b)$ is homogeneous.

Thanks to the arguments presented so far, we have that inverse contrast functions which are homogeneous of degree $n = 0$ are those that can be written as a monotone non-decreasing function of $\min(I(x), I(y))/\max(I(x), I(y))$.

Finally, let us consider Weber-Fechner's law: if the intensity of the uniform background in Weber's experiment is I_0, and the supra-threshold stimulus is I_1, then we can re-write the Weber ratio $\frac{I_1 - I_0}{I_0}$ as $I_1/I_0 - 1$, i.e. the sensation of intensity contrast is a function of I_1/I_0.

This reason motivates us to say that $c(a, b)$ is a *generalized Weber-Fechner contrast function* if c is an *inverse contrast* function which can be written as a non-decreasing function of $\min(a, b)/\max(a, b)$.

Even if Weber-Fechner's law is not valid for all intensity levels, it remains correct for a large dynamic range, thus it seems natural to require c to be a generalized Weber-Fechner contrast function.

Explicit examples of energy functionals complying with our set of assumptions on the contrast function are:

$$C_w^{\mathrm{id}}(I) := \frac{1}{4} \iint_{\Omega^2} w(x, y) \, \frac{\min(I(x), I(y))}{\max(I(x), I(y))} \, dxdy \tag{4.5}$$

$$C_w^{\log}(I) := \frac{1}{4} \iint_{\Omega^2} w(x, y) \, \log\left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))}\right) dxdy \tag{4.6}$$

$$C_w^{-\mathcal{M}}(I) := -\frac{1}{4} \iint_{\Omega^2} w(x, y) \, \mathcal{M}\left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))}\right) dxdy \tag{4.7}$$

where

$$\mathcal{M}\left(\frac{\min(I(x), I(y))}{\max(I(x), I(y))}\right) := \frac{1 - \frac{\min(I(x), I(y))}{\max(I(x), I(y))}}{1 + \frac{\min(I(x), I(y))}{\max(I(x), I(y))}} \equiv \frac{|I(x) - I(y)|}{I(x) + I(y)}, \tag{4.8}$$

is the well known Michelson's definition of contrast [69].

The upper symbol in the above definitions of C_w simply specifies the monotone function applied on the basic contrast variable $t := \frac{\min(I(x),I(y))}{\max(I(x),I(y))}$, i.e. identity $\mathrm{id}(t) = t$, logarithm $\log t$, and minus the Michelson's contrast function $-\mathcal{M}(t) = -\frac{1-t}{1+t}$, respectively.

Notice that the function $t = \min(I(x), I(y))/\max(I(x), I(y))$ is minimized when the minimum takes the smallest possible value and the maximum takes the largest possible one, which corresponds to a contrast stretching. Thus, minimizing an increasing function of the variable t, will produce a contrast enhancement. To refer to any one of them we use the notation $C_w^{\varphi}(I)$, where $\varphi = \mathrm{id}, \log, -\mathcal{M}$.

4.1.2 Entropic adjustment term

Let us now consider the term that should control the dispersion. As suggested previously, we intend it as an adjustment term to the initial given image I_0 and to the average illumination value, represented by the quantity μ. Thus, we define two dispersion functions: $d_1(I(x), I_0(x))$ to measure the separation between $I(x)$ and $I_0(x)$, and $d_2(I(x), \mu)$ which measures the separation from the value μ.

Both d_1 and d_2 are continuous functions $d_{1,2} : \mathbb{R}^2 \to \mathbb{R}^+$ such that $d_{1,2}(a, a) = 0$ for any $a \in \mathbb{R}$, and $d_{1,2}(a, b) > 0$ if $a \neq b$.

We write $d_{I_0,\mu}(I(x)) = d_1(I(x), I_0(x)) + d_2(I(x), \mu)$, and the dispersion energy term as

$$D(I) = \int_\Omega d_{I_0,\mu}(I(x))\, dx. \tag{4.9}$$

In principle, to measure the dispersion of I with respect to I_0 or $1/2$, any distance function can be used. The simplest example would be a quadratic distance

$$D^q_{\alpha,\beta}(I) := \frac{\alpha}{2}\int_\Omega \left(I(x) - \frac{1}{2}\right)^2 dx + \frac{\beta}{2}\int_\Omega (I(x) - I_0(x))^2\, dx, \qquad \alpha, \beta > 0. \tag{4.10}$$

However, given that contrast terms are expressed as homogeneous functions of degree 0, the variational derivatives are homogeneous functions of degree -1. We search for functions able to maintain dimensional coherence with the contrast term.

A good candidate for this is the *entropic dispersion* term [5], i.e.

$$D_{\alpha,\beta}(I) := \alpha \int_\Omega \left(\mu \log \frac{\mu}{I(x)} - (\mu - I(x))\right) dx + \beta \int_\Omega \left(I_0(x) \log \frac{I_0(x)}{I(x)} - (I_0(x) - I(x))\right) dx, \tag{4.11}$$

where $\alpha, \beta > 0$, which is based on the relative entropy distance [5] between I and μ (the first term) and between I_0 and I(the second term).

Notice that, if $a > 0$ and $f(s) = a \log \frac{a}{s} - (a - s)$, $s \in (0, 1]$, then $\frac{df}{ds}(s) = 1 - \frac{a}{s}$ and $\frac{d^2 f}{ds^2}(s) = \frac{a}{s^2} > 0$, $\forall s$. So, $f(s)$ has a global minimum in $s = a$. In particular, this holds when $a = I_0(x)$ or $a = \mu$. Given the statistical interpretation of entropy, we can say that *minimizing $D_{\alpha,\beta}(I)$ amounts to minimizing the disorder of intensity levels around μ and around the original data $I_0(x)$*. Thus, $D_{\alpha,\beta}(I)$ accomplishes the required tasks of an adjustment term.

The minimization of $E^{\varphi}_{w,\alpha,\beta,\mu}(I) = C^{\varphi}_w(I) + D_{\alpha,\beta,\mu}$, corresponds to a trade-off between two opponent mechanisms: on one hand we have entropic control of dispersion around μ and around the original data, on the other hand we have local contrast enhancement invariant

with respect to global illumination changes. These energies are called perceptually-inspired, or perceptual functionals for short.

In the following section we show the explicit minimization of the perceptual functionals.

4.2 Minimization of perceptual functionals

The existence of a minimum in the discrete framework can be guaranteed for a quite general class of energy functionals. Let us define, for $\rho > 0$, $\mathcal{F}_\rho := \{I : \Omega \to [0,1],\ I(x) \geq \rho\ \forall x \in \Omega\}$. We minimize the energy E in the class $\mathcal{F}_\rho$ to avoid singularities in the entropic dispersion term or when we use $\varphi = \log$.

In [77] the following result has been proven.

Theorem 4.2.1 *Let $c : (0,1] \times (0,1] \to \mathbb{R}$ be a continuous function. Let $E(I) := D_{\alpha,\beta}(I) + C_w(I)$ where $C_w(I)$ is given in (4.2). There is a minimum of $E(I)$ in the class of functions $\mathcal{F}_\rho$. An analogous result holds if we use the quadratic dispersion term $D^q_{\alpha,\beta}(I)$.*

Notice that if the energy E is differentiable the argmin I^* satisfies $\delta E(I^*) = 0$. Before computing the Euler-Lagrange equations, we notice that the contrast terms $C^\varphi_w(I)$, $\varphi = \mathrm{id}, \log, -\mathcal{M}$ are not convex and the basic function $t := \frac{\min(a,b)}{\max(a,b)}$ is not differentiable.

In fact, we may write $\min(a,b) = \frac{1}{2}(a+b-|a-b|)$, $\max(a,b) = \frac{1}{2}(a+b+|a-b|)$, for any $a, b \in \mathbb{R}$. The non-differentiability comes from the absolute value $A(z) = |z|$, $z \in \mathbb{R}$.

To be able to use a gradient descent approach for the minimization, we must regularize the basic variable t. We notice that $A'(z) = 1$ if $z > 0$, $A'(z) = -1$ if $z < 0$ and A is not differentiable at $z = 0$. But all the values $s \in [-1,1]$ are *subtangents* of $A(z)$ at $z = 0$, that is, $A(z) - A(0) \geq s(z - 0)$ for any $z \in \mathbb{R}$.

Thus we may write $A'(z) = \mathrm{sign}(z)$, where

$$\mathrm{sign}(z) = \begin{cases} 1 & \text{if } z > 0 \\ [-1,1] & \text{if } z = 0 \\ -1 & \text{if } z < 0 \end{cases} . \tag{4.12}$$

We define $\mathrm{sign}_0(z)$ as in (4.12), but with the particular choice 0 when $z = 0$.

It is useful to introduce the following definition. Given $\varepsilon > 0$, we say that $A_\varepsilon(z)$ is a '*nice regularization*' of $A(z)$, if $A_\varepsilon(z) \geq 0$ is convex, differentiable with continuous derivative, $A_\varepsilon(0) = 0$, $A_\varepsilon(-z) = A_\varepsilon(z)$, and

- (i) $A_\varepsilon(z) \leq |z|$ for any $z \in \mathbb{R}$ and $A_\varepsilon(z) = |z| + Q_{1,\varepsilon}(z)$, where $Q_{1,\varepsilon}(z) \to 0$ as $\varepsilon \to 0$, uniformly in $z \in [-1,1]$;
- (ii) Let us denote $s_\varepsilon(z) = A'_\varepsilon(z)$. Then $|s_\varepsilon(z)| \leq 1$ for any $z \in [-1,1]$, $s_\varepsilon(z) \to \mathrm{sign}_0(z)$ as $\varepsilon \to 0$ for any $z \in \mathbb{R}$, and $Q_{2,\varepsilon}(z) := A_\varepsilon(z) - z s_\varepsilon(z) \to 0$ as $\varepsilon \to 0$, uniformly in $z \in [-1,1]$.

Nice regularization of $A(z)$ actually exist:

- *Example* 1): $A_\varepsilon(z) = \sqrt{\varepsilon^2 + |z|^2} - \varepsilon$, in this case $s_\varepsilon(z) = \frac{z}{\sqrt{\varepsilon^2+|z|^2}}$, $Q_{1,\varepsilon}(z) = O(\varepsilon)$ and $Q_{2,\varepsilon}(z) := A_\varepsilon(z) - z s_\varepsilon(z) = O(\varepsilon)$;

- *Example* 2): $A_\varepsilon(z) = z\frac{\arctan(z/\varepsilon)}{\arctan(1/\varepsilon)} - \frac{\varepsilon}{2\arctan(1/\varepsilon)}\log(1+\frac{z^2}{\varepsilon^2})$, in this case $s_\varepsilon(z) = \frac{\arctan(z/\varepsilon)}{\arctan(1/\varepsilon)}$, $Q_{1,\varepsilon}(z) = O(\varepsilon\log(1/\varepsilon))$ and $Q_{2,\varepsilon}(z) = O(\varepsilon\log(1/\varepsilon))$, uniformly in $z \in [-1,1]$.

$O(F(\varepsilon))$ is any expression satisfying $|O(F(\varepsilon))| \le CF(\varepsilon)$ for some constant $C > 0$ and $\varepsilon > 0$ small enough. In both cases $s_\varepsilon(z) \to \mathrm{sign}_0(z)$ as $\varepsilon \to 0$ for any $z \in \mathbb{R}$, see [77].

Now let us assume that $A_\varepsilon(z)$ is a nice regularization of $A(z)$. We set

$$\min_\varepsilon(a,b) = \frac{1}{2}(a+b-A_\varepsilon(a-b)), \qquad \max_\varepsilon(a,b) = \frac{1}{2}(a+b+A_\varepsilon(a-b)). \tag{4.13}$$

We define the regularized version of the perceptual functionals as follows:

$$C^{\mathrm{id}}_{w,\varepsilon}(I) := \frac{1}{4}\iint_{\Omega^2} w(x,y)\,\frac{\min_\varepsilon(I(x),I(y))}{\max_\varepsilon(I(x),I(y))}\,dxdy; \tag{4.14}$$

$$C^{\log}_{w,\varepsilon}(I) := \frac{1}{4}\iint_{\Omega^2} w(x,y)\,\log\left(\frac{\min_\varepsilon(I(x),I(y))}{\max_\varepsilon(I(x),I(y))}\right)\,dxdy; \tag{4.15}$$

$$C^{-\mathcal{M}}_{w,\varepsilon}(I) := -\frac{1}{4}\iint_{\Omega^2} w(x,y)\,\frac{A_\varepsilon(I(x)-I(y))}{I(x)+I(y)}\,dxdy. \tag{4.16}$$

In [77] it has been proven the following result.

Theorem 4.2.2 *Assume that $A_\varepsilon(z)$ is a nice regularization of $A(z)$.*

(i) The first variation of $C^{\mathrm{id}}_{w,\varepsilon}(I)$ is:

$$\begin{aligned}\delta C^{\mathrm{id}}_{w,\varepsilon}(I) &= -\tfrac{1}{2}\textstyle\int_\Omega w(x,y)\frac{I(y)}{\max_\varepsilon(I(x),I(y))^2}s_\varepsilon(I(x)-I(y))\,dy + S_\varepsilon \\ &= -\tfrac{1}{2}\textstyle\int_\Omega w(x,y)\frac{I(y)}{\max(I(x),I(y))^2}s_\varepsilon(I(x)-I(y))\,dy + S'_\varepsilon,\end{aligned} \tag{4.17}$$

where $S_\varepsilon, S'_\varepsilon = O(Q_{1,\varepsilon}(I(x)-I(y)) + Q_{2,\varepsilon}(I(x)-I(y)))$. Notice that in the first formula we have max *and in the second one we have* $\max_\varepsilon$*;*
(ii) The first variation of $C^{\log}_{w,\varepsilon}(I)$ is:

$$\delta C^{\log}_{w,\varepsilon}(I) = -\frac{1}{2}\int_\Omega w(x,y)\frac{1}{I(x)}s_\varepsilon(I(x)-I(y))\,dy + S_\varepsilon, \tag{4.18}$$

where $S_\varepsilon = O(Q_{1,\varepsilon}(I(x)-I(y)) + Q_{2,\varepsilon}(I(x)-I(y)))$;
(iii) The first variation of $C^{-\mathcal{M}}_{w,\varepsilon}(I)$ is:

$$\delta C^{-\mathcal{M}}_{w,\varepsilon}(I) = -\int_\Omega w(x,y)\frac{I(y)}{(I(x)+I(y))^2}s_\varepsilon(I(x)-I(y))\,dy + S_\varepsilon, \tag{4.19}$$

where $S_\varepsilon = O(Q_{2,\varepsilon}(I(x)-I(y)))$.
In all cases $Q_{1,\varepsilon}(z), Q_{2,\varepsilon}(z) \to 0$ as $\varepsilon \to 0$, uniformly in $z \in [-1,1]$.

Thus we know that $S_\varepsilon = O(\varepsilon)$ if $A_\varepsilon(z)$ is given by Example 1), and $S_\varepsilon = O(\varepsilon \log 1/\varepsilon)$ if $A_\varepsilon(z)$ is given by Example 2). These are the cases of interest for us in the experiments.

Notice that, letting $\varepsilon \to 0$ we have that $\delta C^{\varphi}_{w,\varepsilon}(I) \to \delta C^{\varphi}_{w,0}(I)$, where

$$\begin{aligned}
\delta C^{\mathrm{id}}_{w,0}(I) &= -\frac{1}{2}\int_\Omega w(x,y)\frac{I(y)}{\max(I(x),I(y))^2}\,\mathrm{sign}_0(I(x)-I(y))\,dy \\
&= -\frac{1}{2}\left(\int_{\{y\in\Omega\,:\,I(x)>I(y)\}} w(x,y)\frac{I(y)}{I(x)^2} - \int_{\{y\in\Omega\,:\,I(x)<I(y)\}} w(x,y)\frac{1}{I(y)}\right);
\end{aligned}$$

$$\begin{aligned}
\delta C^{\log}_{w,0}(I) &= -\frac{1}{2}\int_\Omega w(x,y)\frac{1}{I(x)}\,\mathrm{sign}_0(I(x)-I(y))\,dy \\
&= -\frac{1}{2}\left(\int_{\{y\in\Omega\,:\,I(x)>I(y)\}} w(x,y)\frac{1}{I(x)} - \int_{\{y\in\Omega\,:\,I(x)<I(y)\}} w(x,y)\frac{1}{I(x)}\right);
\end{aligned}$$

$$\begin{aligned}
\delta C^{-\mathcal{M}}_{w,0}(I) &= -\int_{y\in\Omega} w(x,y)\frac{I(y)}{(I(x)+I(y))^2}\,\mathrm{sign}_0(I(x)-I(y))\,dy \\
&= -\int_{\{y\in\Omega\,:\,I(x)>I(y)\}} w(x,y)\frac{I(y)}{(I(x)+I(y))^2} + \int_{\{y\in\Omega\,:\,I(x)<I(y)\}} w(x,y)\frac{I(y)}{(I(x)+I(y))^2}.
\end{aligned}$$

Now, by direct computation, we have that the derivative of the entropic dispersion term is:

$$\delta D_{\alpha,\beta}(I) = \alpha\left(1-\frac{\mu}{I(x)}\right) + \beta\left(1-\frac{I_0(x)}{I(x)}\right). \tag{4.20}$$

This expression has a degree of homogeneity -1 with respect to $I(x)$, the same as the variation of the three contrast terms $C^{\varphi}_{w}(I)$.

Assume that $\alpha, \beta > 0$ are fixed. If $E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}(I) = D_{\alpha,\beta,\mu}(I) + C^{\varphi}_{w,\varepsilon}(I)$, $\varepsilon = \mathrm{id}, \log, -\mathcal{M}$, then by linearity of the variational derivative, we have $\delta E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}(I) = \delta D_{\alpha,\beta,\mu}(I) + \delta C^{\varepsilon}_{w,\varphi}(I)$.

The argmin of $E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}(I)$ satisfies $\delta E^{\varphi}_{w,\varepsilon}(I) = 0$. To search for the argmin we use a semi-implicit discrete gradient descent strategy with respect to $\log I$. The continuous gradient descent equation is

$$\partial_t \log I = -\delta E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}(I), \tag{4.21}$$

being t the evolution parameter. Since $\partial_t \log I = \frac{1}{I}\partial_t I$, we have

$$\partial_t I = -I\delta E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}(I). \tag{4.22}$$

Using the gradient descent in $\log I$ leads to (4.22), which is related to a gradient descent approach which uses *the relative entropy as a metric*, instead of the usual quadratic distance (see [5]).

Let us now discretize our scheme: choosing a finite evolution step $\Delta t > 0$ and setting $I^k(x) = I_{k\Delta t}(x)$, $k = 0, 1, 2, \ldots$, being $I^0(x)$ the original image, thanks to (4.20), we can write the semi-implicit discretization of (4.22) as

$$\frac{I^{k+1}(x) - I^k(x)}{\Delta t} = \alpha\left(\frac{1}{2} - I^{k+1}(x)\right) + \beta\left(I_0(x) - I^{k+1}(x)\right) - I^k(x)\delta C^{\varphi}_{w,\varepsilon}(I^k). \tag{4.23}$$

Now, considering the explicit expressions of $\delta C^{\varphi}_{w,\varepsilon}(I^k)$, neglecting their second terms containing S_ε (see Proposition 4.2.2 (i),(ii),(iii)) and performing some easy algebraic manipulations, we find the equation

$$I^{k+1}(x) = \frac{I^k(x) + \Delta t \left(\alpha\mu + \beta I_0(x) + \frac{1}{2} R^{\varphi}_{\varepsilon,I^k}(x)\right)}{1 + \Delta t(\alpha+\beta)}, \tag{4.24}$$

where the function $R^{\varphi}_{\varepsilon,I^k}(x)$ assumes three different forms for $\varphi = \mathrm{id}$, log, $-\mathcal{M}$, precisely

$$R^{\mathrm{id}}_{\varepsilon,I^k}(x) := -2I^k \delta C^{\mathrm{id}}_{w,\varepsilon}(I^k) = \int_\Omega w(x,y) \frac{I^k(x)I^k(y)}{\max_\varepsilon(I^k(x), I^k(y))^2} s_\varepsilon(I^k(x) - I^k(y))\, dy. \tag{4.25}$$

$$R^{\log}_{\varepsilon,I^k}(x) := -2I^k \delta C^{\log}_{w,\varepsilon}(I^k) = \int_\Omega w(x,y)\, s_\varepsilon(I^k(x) - I^k(y))\, dy. \tag{4.26}$$

$$R^{\mathcal{M}}_{\varepsilon,I^k}(x) := -2I^k \delta C^{-\mathcal{M}}_{w,\varepsilon}(I^k) = \int_\Omega w(x,y)\, \frac{2I^k(x)I^k(y)}{(I^k(x) + I^k(y))^2} s_\varepsilon(I^k(x) - I^k(y))\, dy. \tag{4.27}$$

The discrete equations corresponding to the limit $\varepsilon \to 0$ are:

$$R^{\mathrm{id}}_{0,I^k}(x) = \sum_{\{y\in\Omega\,:\, I^k(x) > I^k(y)\}} w(x,y) \frac{I^k(y)}{I^k(x)} - \sum_{\{y\in\Omega\,:\, I^k(x) < I^k(y)\}} w(x,y) \frac{I^k(x)}{I^k(y)}; \tag{4.28}$$

$$R^{\log}_{0,I^k}(x) = \sum_{y\in\Omega} w(x,y)\, \mathrm{sign}_0(I^k(x) - I^k(y)); \tag{4.29}$$

$$R^{-\mathcal{M}}_{0,I^k}(x) = \sum_{\Omega} w(x,y)\, \frac{2I^k(x)I^k(y)}{(I^k(x) + I^k(y))^2}\, \mathrm{sign}_0(I^k(x) - I^k(y)). \tag{4.30}$$

4.2.1 Stability of the iterative semi-implicit gradient-descent scheme

We assume here that $I_0 : \Omega \to [\rho, 1]$ where $\rho > 0$ is a minimum value for the initial image. For us $\rho = 1/255$, and this means that we assume that I_0 does not take the value 0. In [77] we have proven the following statement about the stability of the argmin computation for the perceptual functionals.

Theorem 4.2.3 *Assume that $\alpha \geq \frac{1}{1-2\rho} > 0$. Then, $\forall k \geq 1$:*

- $-1 \leq R^{\varphi}_{\varepsilon,I^k}(x) \leq 1\ \forall x \in \Omega$;
- $\rho \leq I^k(x) \leq 1\ \forall x \in \Omega$;

- $\forall p \in [1, \infty]$ *:*

$$\|I^{k+1} - I^k\|_p \leq \frac{1 + \Delta t \left(\frac{1}{\rho} + m_\varepsilon\right)}{1 + \Delta t(\alpha + \beta)} \|I^k - I^{k-1}\|_p,$$

where $m_\varepsilon := \max_{z \in [-1,1]} |s'_\varepsilon(z)|$.

As a consequence, if $\alpha + \beta > \frac{1}{\rho} + m_\varepsilon$*, then the iterative scheme given in (4.24) is convergent to the unique function* I^* *satisfying*

$$\alpha(I^* - \mu) + \beta(I^* - I_0) - \frac{1}{2} R^{\varphi}_{\varepsilon, I^*} = 0. \tag{4.31}$$

As usual, the $\|\cdot\|_p$ norm of a vector $v = (v_i)_{i=1}^n \in \mathbb{R}^n$, $n \geq 1$, is defined by $\|v\|_p := \left(\sum_{i=1}^n |v_i|^p\right)^{1/p}$ if $p \in [1, \infty)$, and $\|v\|_\infty := \max_{i=1,\ldots,n} |v_i|$ if $p = \infty$.

Notice that (4.31) is essentially (and not exactly, due to our regularization of the basic contrast variable) the Euler-Lagrange equation corresponding to the energy $E^{\varphi}_{w,\varepsilon,\alpha,\beta,\mu}$.

Notice also that $\alpha + \beta > \frac{1}{\rho} + m_\varepsilon$ is not a reasonable condition in practice. The reason is that $1/\rho = 255$ and $m_\varepsilon \approx 1/\varepsilon$ for the particular nice regularization of $A(z)$ given in the examples 1),2) given in subsection 4.2. Then the values of α and/or β are too big and produce a strong attachment to the initial data and/or the value 1/2; in this case we do not have enough enhancement power. The values of α and β used in practice are much smaller, $\alpha = 255/253$ and $\beta = 1$, but enough to guarantee stability and convergence for all the images that we have considered.

4.2.2 A general strategy for the reduction of computational complexity

To better appreciate the simplicity of the idea behind the computational complexity reduction, it is worthwhile omitting the superscript k to avoid a cumbersome notation.

The computational complexity relative the general form of semi-implicit gradient descent equation (4.24) is $\mathcal{O}(N)$, N being the total number of image pixels, for every $x \in \Omega$, this means that the computational complexity for the whole images is $\mathcal{O}(N^2)$. Of course, this large cost is due to the term $R^{\varphi}_I(x)$.

In fact, $R^{\varphi}_I(x)$ have the same analytical structure for every choice of φ, i.e.

$$R^{\varphi}_{\varepsilon,I}(x) = \int_\Omega w(x,y)\, r^{\varphi}_\varepsilon(I(x), I(y))\, dy, \tag{4.32}$$

where $r^{\varphi}_\varepsilon(I(x), I(y))$ is specified in the equations (4.25), (4.26), (4.27).

To reduce the complexity and obtain an algorithm that can be used in a reasonable time, at least three strategies can be followed.

The first consists in localizing the computation in a circular window centered in the pixel x under analysis and discarding all the pixels that lie outside this window, using the argument that the kernel w takes values very close to 0 when the distance between pixel pairs becomes large enough (of course this strategy implies a re-normalization of w in the circular sub-image). The second consists in replacing the kernel by a sparse sampling structure, for instance the local random sprays technique proposed in [95].

The third strategy is the one that we have pursued and it is based on a polynomial approximations of the function $r_\varepsilon^\varphi(I(x), I(y))$, as we have first proposed in [15] and then further refined in [77].

Let p be a generic polynomial of order n of the variables $I(x), I(y)$ and define

$$\tilde{r}_{n,\varepsilon}^\varphi(I(x), I(y)) = \text{argmin}_p \, \|p - r_\varepsilon^\varphi\|_2, \tag{4.33}$$

i.e. $\tilde{r}_n$ is the polynomial of order n with *minimal quadratic distance* with respect to the function φ.

It is convenient to write $\tilde{r}_{n,\varepsilon}^\varphi$ as follows $\tilde{r}_{n,\varepsilon}^\varphi(I(x), I(y)) = \sum_{j=0}^n f_j^{\varepsilon,\varphi}(I(x))I(y)^j$, where $f_j^{\varepsilon,\varphi}(I(x)) = \sum_{k=0}^{n-j} p_{k,j}^{\varepsilon,\varphi} I(x)^k$.

The numerical coefficients $p_{k,j}^{\varepsilon,\varphi}$ depend on the approximation order n, but we will not make explicit this dependence for the sake of a more readable mathematical notation.

Introducing this expression of $\tilde{r}_{n,\varepsilon}^\varphi$ in (4.32) instead of r_ε^φ and noticing that $f_j^{\varepsilon,\varphi}$ does not depend on y, we get

$$\tilde{R}_{n,\varepsilon,\varphi}^\varphi(x) = \sum_{j=0}^n f_j^{\varepsilon,\varphi}(I(x)) \int_\Omega w(\|x-y\|)I(y)^j dy, \tag{4.34}$$

but $\int_\Omega w(\|x-y\|)I(y)^j dy = (w * I^j)(x)$, that is the convolution between the kernel w and the j-th power of the image I, hence

$$\tilde{R}_{n,\varepsilon,\varphi}^\varphi(x) = \sum_{j=0}^n \sum_{k=0}^{n-j} p_{k,j}^{\varepsilon,\varphi} I(x)^k (w * I^j)(x). \tag{4.35}$$

The n convolutions $w * I^j$ can be *pre-computed* through the FFT (Fast Fourier Transform), which has computational complexity $\mathcal{O}(N \log N)$, thus drastically decreasing the computational time.

The degree n of the polynomial is a parameter that controls the precision of the polynomial approximation. In our tests, we have found that $n = 9$ is a good trade-off between time cost and approximation precision.

4.2.3 Results

In this section are presented some results of the algorithms derived by the model just described, with the three different choices of φ =id,log,$-\mathcal{M}$, for a more complete analysis of the results with respect to variations of parameters see [77].

The results presented here have been obtained by using the implementation [34], publicly available on the website IPOL, with the following parameters: $\alpha = \beta = 1.2$, $\varepsilon = 1/20$, w: Gaussian kernel with standard deviation equal to 1/5 of the length of the image diagonal.

Let us start by showing in Figure 4.1 how the low-key image presented in Figure 1.3 is handled by the variational algorithms versus the von Kries algorithm (normalization over the maximum intensity in the three chromatic channels separately) and the automatic color enhancement performed by photoshop. It can be seen that the perceptual-inspired color enhancement avoids the color shifting of the von Kries method and it is able to show many more details than the automatic photoshop algorithm. Also recall from Figure 1.3 that histogram equalization has a dramatic effect on this image, which is absent in our perceptually-inspired modification.

The intra-variations of the variational algorithm with respect to different choices of φ are perceptually small, the major difference being that, in general, $\varphi = \log$ corresponds to the brightest image, but with less saturated colors with respect to the other two choices.

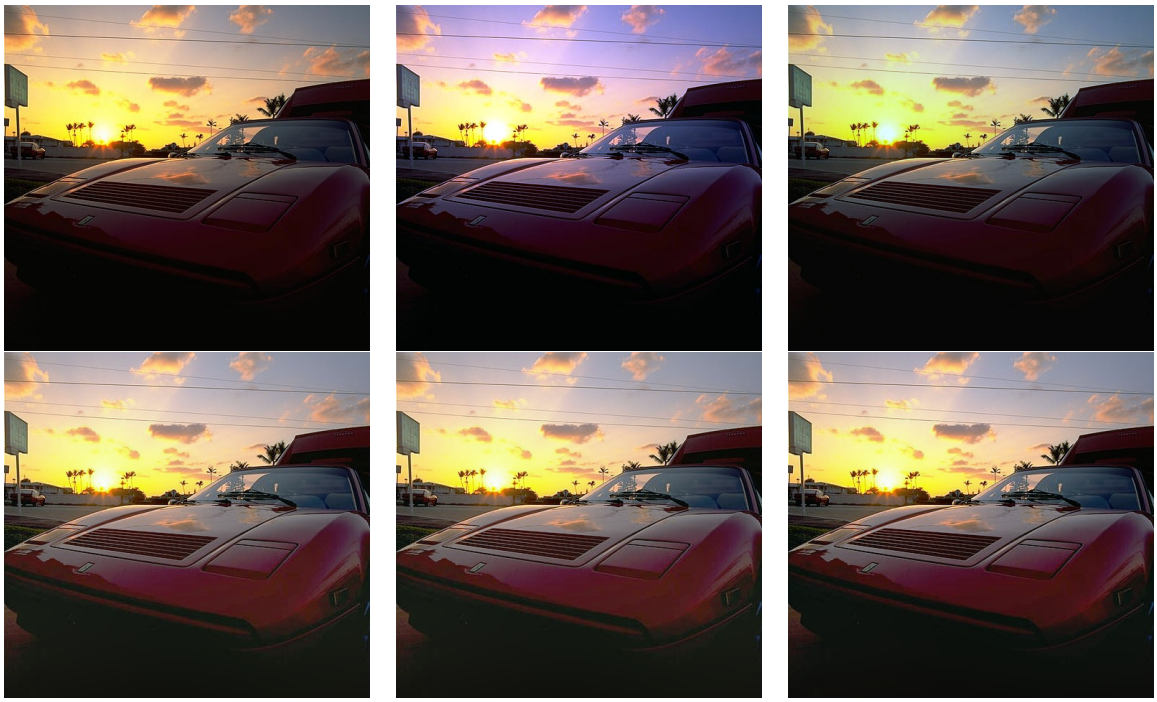


Figure 4.1: *First row from left to right*: original image, result of the global von Kries algorithms and result of the automatic color correction performed by Photoshop. *Second row from left to right*: result of the variational algorithms with $\varphi =$id,log,$-\mathcal{M}$.

In Figure 4.2 we can seen the effect of the variational algorithm with $\varphi =$id on an image with haze and the difference with respect to histogram equalization.

Finally, in Figure 4.3 we can seen the behavior of the perceptually-inspired variational algorithms with respect to images with color cast and the comparison with von Kries and histogram equalization methods. Notice that the von Kries algorithm is not able to completely remove color cast and that histogram equalization creates a dramatic color shift.

I would like to underline again that the main aim of the research presented in this chapter is not to build to the most efficient color correction algorithm ever, but to show that, by carefully modifying the histogram equalization functional guided by the HVS features, it is possible to avoid the typical problems related with large color shift and excessive amplification of contrast even working in the three separated chromatic channels.

The techniques developed here may be used in applications where a parallel elaboration is needed and they can also be extended to more than three wavebands, as it is the case in hyperspectral imaging.

Figure 4.2: *Row-wise from left to right*: original image, results of the variational algorithms with φ =id, of the global von Kries algorithms and of histogram equalization.

4.3 Embedding existing perceptually-inspired color correction models in the variational framework

One of the main theoretical achievements of [77] is the proof that the perceptual variational framework can embed and link together in a clear way two of the most used perceptually-inspired color correction algorithms: the famous Retinex of Land and McCann [62] and ACE, automatic color equalization [102].

The embedding of ACE is simple to discuss: if we choose $\varphi = \log$, $\mu = 1/2$, $\alpha = 1$ and $\beta = 0$ in eq. (4.24), then, as proven in [15], we obtain, as a particular instance, the ACE algorithm, which has been developed starting from the so-called 'gray-world' hypothesis introduced by Buchsbaum in [17], which says that the average reflectance in a natural image is achromatic.

The embedding of Retinex has proven to be more difficult, in fact, the first algorithmic description provided by Land and McCann in [62] seems distant from the variational framework just discussed. However, thanks to the mathematical analysis of Retinex developed in [93, 95, 96], in the paper [14] it has been proven that φ =id, $\alpha = 1$ and $\beta = 0$, corresponds to a symmetrized continuous version of Retinex.

The symmetrization is needed because, as it has been proven in [93], the original Retinex algorithm always increases pixel intensities, which, of course, is an unwanted effect since over-exposed images cannot be properly handled. Report all the details about Retinex and its embedding in the framework presented in this chapter would take too long, all the details can be found in the quoted papers.

Figure 4.3: *Row-wise from left to right*: original image, results of the variational algorithms with φ =id, of the global von Kries algorithms and of histogram equalization.

4.4 Perceptual enhancement in the wavelet domain

Let us conclude this chapter with the wavelet version of the variational model presented in section 4.1.

We start with a very brief introduction to wavelets.

Let us start by very briefly recalling the basic information about wavelet theory in one dimension, then we will extend the discussion to 2-D wavelets, the main reference for all the results quoted hereafter is [64]. A 1-D (mother) wavelet $\psi \in L^2(\mathbb{R})$ is a unit norm and null-mean function. Of course this is possible only if ψ oscillates, but, unlike infinite waves, wavelets can have compact support. The ψ-wavelet transform $W_\psi f$ of $f \in L^2(\mathbb{R})$ in the point ξ at the scale s is given by the inner product $W_\psi f(\xi, s) = \int_{\mathbb{R}} f(x) \frac{1}{\sqrt{s}} \overline{\psi}\left(\frac{x-\xi}{s}\right) dx$. $W_\psi f$ gives a 'measure of similarity' between f and ψ around the point ξ at the scale s. So, if a signal is constant or do not vary 'too much' in the support of a wavelet, then its wavelet transform will be zero or very small, this is how wavelets provide a multiscale information about the *local contrast* of a signal.

The set $\{\psi_{j,k}\}_{(j,k)\in\mathbb{Z}^2} \subset L^2(\mathbb{R})$ given by $\psi_{j,k}(x) \equiv \frac{1}{\sqrt{2^j}} \psi\left(\frac{x-2^j k}{2^j}\right)$ is a complete orthonormal system of $L^2(\mathbb{R})$. Moreover, $L^2(\mathbb{R})$ can be recovered by the closure of the union of a sequence of nested closed subspaces $V_j \subset V_{j-1}$ with suitable properties (see Mallat's book [64] for more details). The orthogonal projections of $f \in L^2(\mathbb{R})$ onto V_j and V_{j-1} give the approximation of f at the scales 2^j and 2^{j-1}, respectively. The 2^j-approximation is coarser and the missing details with respect to the finer 2^{j-1}-approximation are contained in the orthogonal complement W_j of V_j in V_{j-1}: $V_{j-1} = V_j \oplus W_j$. W_j is called the j-th *detail space* and it can be proven that

the orthogonal projection of f on W_j is given by $P_{W_j} f = \sum_{k\in\mathbb{Z}} \langle f, \psi_{j,k}\rangle \psi_{j,k} \equiv \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}$. The coefficients $d_{j,k}$ are called *detail coefficients of f at the scale 2^j*. Fine-scales detail coefficients at fine scale are sparse, in fact, they are non-null only when the support of $\psi_{j,k}$ intersects a high contrast zone, i.e. around sharp edges.

Finally, let us recall that every wavelet ψ is related to a mirror filter h and to a function ϕ, called *scale function*, through the following equation that involves their Fourier transforms: $\hat{\psi}(2\omega) = \frac{1}{\sqrt{2}} e^{-i\omega} \hat{h}^* (\omega + \pi)\, \hat{\phi}(\omega)$, see [111] for a complete and detailed description of how to generate wavelets using the filter design methodology. ϕ appears in the orthogonal projection of a signal f onto the approximation space V_j, in fact it can be proven that $P_{V_j} f = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j,k}\rangle \phi_{j,k} \equiv \sum_{k\in\mathbb{Z}} a_{j,k}\phi_{j,k}$, where $\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{x-2^j k}{2^j}\right)$ and $a_{j,k}$ are called *approximation coefficients* at the scale 2^j. It follows that $P_{V_{j-1}} f = P_{V_j} f + P_{W_j} f = \sum_{k\in\mathbb{Z}} a_{j,k}\phi_{j,k} + \sum_{k\in\mathbb{Z}} d_{j,k}\psi_{j,k}$.

In practical applications one is interested in a multiresolution analysis between two fixed scales 2^L and 2^J, $L, J \in \mathbb{Z}$, $L < J$. In this case $V_{J-1} = V_J \oplus W_J$, $V_{J-2} = V_{J-1} \oplus W_{J-1} = V_J \oplus W_J \oplus W_{J-1}$ and so on, thus $V_L = V_J \oplus \bigoplus_{2^J \geq 2^j \geq 2^{L+1}} W_j$. For this reason, following [64], we say that a discrete orthogonal wavelet multiresolution representation of a 1-dimensional signal f between two fixed scales 2^L and 2^J, $L, J \in \mathbb{Z}$, $L < J$, is given by the collection of detail coefficients $\{d_{j,k}\}$ at all scales, completed by the approximation coefficients at the coarser scale, i.e. $\{a_{J,k}\}$.

When we deal with 2-D signals, as images, we have to consider a multiresolution analysis of $L^2(\mathbb{R}^2)$. Multidimensional wavelet bases can be generated with tensor products of separable basis functions defined along each dimension. In this case, an orthogonal wavelet multiresolution representation between two scales 2^L and 2^J, $L, J \in \mathbb{Z}$, $L < J$, is given by three sets of detail coefficients $\{d^H_{j,k}, d^V_{j,k}, d^D_{j,k}\}$ at all scales, which correspond to the *horizontal, vertical and diagonal detail coefficients*, respectively, completed by the approximation coefficients at the coarser scale, i.e. $\{a_{J,k}\}$.

As seen in this chapter, the perceptual color correction proposed in [77] is performed through a local contrast enhancement balanced by the action of a adjustment term around the average value plus a conservative term that avoids over-enhancement. Wavelet detail coefficients are related to local contrast, let us see how this fact has been used in [92] to provide a wavelet-based framework for perceptually-inspired color correction.

The general scheme of the algorithm is the following:

1. We consider the coarsest approximation coefficients $\{a_{J,k},\, k \in \Omega\}$, and we modify them to implement adaptation to the average gray level in the wavelet domain (according to eq. (4.36) of Section 4.4.1);

2. We fix these new approximation coefficients and use them along with $\{d^\ell_{J,k},\, k \in \Omega\}$, $\ell = H, V, D$, to modify the horizontal, vertical and diagonal detail coefficients according to the equations of Proposition 4.4.1 of Section 4.4.2. This will implement local contrast enhancement in the wavelet domain at the scale 2^J;

3. We then pass to the scale 2^{J-1} and we compute the approximation coefficients by summing the approximation and detail coefficients just computed at the scale 2^J;

4. We fix these approximation coefficients and we repeat step 2 at the scale 2^{J-1};

5. We iterate this scheme until reaching the finest scale, linearly scaling the minimum value to 0 and the maximum to 1.

The equations quoted in the scheme above will be presented and analyzed in full detail in the two next subsections, but before doing that it is convenient to make some assumptions that will greatly simplify the exposition.

First of all, since changing the sign of a wavelet coefficient can result in drastic modifications of an image, we will modify only the *absolute value* of the wavelet coefficients, restoring the original sign at the end of the computation.

Moreover, in every scale 2^j, we will deal only with coefficient magnitudes bigger than a fixed threshold $T_j > 0$, leaving the other coefficients unchanged to avoid intensification of noise. Thus we will deal with *positive, bounded and finite* sequences of wavelet coefficient magnitudes.

To simplify the notation, we will avoid the superscript ℓ in the detail coefficients, by making the implicit assumption that the operations are repeated on the horizontal (H), vertical (V) and diagonal (D) detail coefficients.

The variability of the scale coefficient 2^j will be confined within the scales 2^L and 2^J, $L, J \in \mathbb{Z}$, $L < J$, 2^L being the finest and 2^J the coarsest. Typically $J - L$ ranges between 2 and 10, depending on the image dimension. Finally, with the notation $k \in \Omega$, we will implicitly assume a column-wise ordering of Ω, the spatial support of the image, so that Ω can be seen as a finite subset of $\mathbb{Z}$.

4.4.1 Adjustment to the average value in the wavelet domain

Let us first consider the effect of adaptation to the average level. If we were dealing with Fourier transforms, the average image intensity value μ would be represented by the zero-order Fourier coefficient, in the wavelet domain there is no such direct correspondence. However, since the coefficients $\{a_{j,k},\ k \in \Omega\}$ represent the image approximation at the scale 2^j, a natural analogue of the average value μ in the wavelet framework at the scale 2^j is represented by $\overline{a}_j \equiv \frac{1}{|\Omega|}\sum_{k\in\Omega} a_{j,k}$, i.e. the average approximation coefficient.

We also stress that we only need to modify the approximation coefficients of the coarsest scale, since this modification will be propagated to finer scales. Thus, we implement the adaptation to the average value at the coarser scale balanced by the adjustment to the original values through this convex linear combination

$$a_{J,k} \equiv \alpha \overline{a}_J + (1-\alpha) a^0_{J,k}, \tag{4.36}$$

where $\{a^0_{J,k},\ k \in \Omega\}$ is the original sequence of approximation coefficients at the scale 2^J and $\alpha \in [0,1]$ is a suitable weight coefficient. The bigger α, the stronger the adjustment to the average value $\overline{a}_J$, and viceversa.

4.4.2 Local contrast enhancement in the wavelet domain

Let us now consider local contrast enhancement. We have remarked that the two most important features of the contrast functional $E_{\mu,I_0,w,\varphi}(I)$ are the fact that it enhances contrast *locally* and in an *illuminant-independent* way. In order to maintain these characteristics also in the wavelet domain, in [92] the following local contrast enhancement functional at the scale 2^j has been proposed:

$$\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}(\{d_{j,k}\}) \equiv \sum_{k\in\Omega} w_j\, \varphi\left(\frac{a_{j,k}}{d_{j,k}}\right), \qquad 2^J \geq 2^j \geq 2^{L+1}, \tag{4.37}$$

where w_j are positive coefficients that permit to differentiate the contrast enhancement action depending on the scale 2^j and $\varphi : [0, \infty) \to [0, \infty)$ is a differentiable monotonically increasing function such that $\varphi(r) \to +\infty$ as $r \to \infty$.

We stress that the approximation coefficients are passed to the functional $\mathcal{C}$ as fixed parameters in every scale, from the coarsest to the finest. In particular, the approximation coefficients used in the coarsest scale are those defined by eq. (4.36).

$\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}$ is minimized when the ratio between the approximation and detail coefficients decreases, but since the approximation coefficients are fixed at each scale, the minimization of $\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}$ corresponds to an intensification of the coefficients $d_{j,k}$. This implies a local and multi-scale contrast enhancement of the image. The locality depends both on the mother wavelet chosen (because different mother wavelets have different shape) and on the scale 2^j at which one operates: the finer the scale, the more local is contrast enhancement, and viceversa.

Furthermore, the invariance with respect to global illumination changes, in the sense of the von Kries model [120], is guaranteed by the fact that also the functional $\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}$ is *homogenous of degree 0*, i.e. so the transformation $I \mapsto \lambda I$, $\lambda > 0$, will have no effects on $\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}$ and its Euler-Lagrange equations.

In order to prevent an excessive magnification of the original detail coefficients, whose absolute value is denoted with $d^0_{j,k}$, a conservative term should be introduced. To maintain dimensional coherence with $\mathcal{C}_{w_j,\varphi,\{a_{j,k}\}}$, an entropic dispersion functional is a suitable choice:

$$\mathcal{D}_{d^0_{j,k}}(\{d_{j,k}\}) \equiv \sum_{k\in\Omega} \left[d^0_{j,k} \log \frac{d^0_{j,k}}{d_{j,k}} - \left(d^0_{j,k} - d_{j,k}\right) \right], \qquad 2^J \geq 2^j \geq 2^{L+1}. \tag{4.38}$$

Combining these two effects one can define the energy functional that realizes local contrast enhancement as $\mathcal{E}_{w_j,\varphi,\{a_{j,k}\},d^0_{j,k}} = \mathcal{C}_{w_j,\varphi,\{a_{j,k}\}} + \mathcal{D}_{d^0_{j,k}}$, i.e.

$$\mathcal{E}_{w_j,\varphi,\{a_{j,k}\},d^0_{j,k}} \equiv \sum_{k\in\Omega} \left[w_j \varphi\left(\frac{a_{j,k}}{d_{j,k}}\right) + d^0_{j,k} \log \frac{d^0_{j,k}}{d_{j,k}} - \left(d^0_{j,k} - d_{j,k}\right) \right], \tag{4.39}$$

with $2^J \geq 2^j \geq 2^{L+1}$. The following theorem ensures the existence of a minimum of $\mathcal{E}_{w_j,\varphi,\{a_{j,k}\},d^0_{j,k}}$ and determines the corresponding Euler-Lagrange equations. Its proof can be found in [92].

Theorem 4.4.1 *There exists a minimum of the functional $\mathcal{E}_{w_j,\varphi,\{a_{j,k}\},d^0_{j,k}}$. Moreover, the Euler-Lagrange equations for the detail coefficients are:*

$$\frac{\partial \mathcal{E}_{w_j,\varphi,\{a_{j,k}\},d^0_{j,k}}}{\partial\{d_{j,k}\}}(d_{j,k}) = 0 \iff d_{j,k} = d^0_{j,k} + w_j \varphi'\left(\frac{a_{j,k}}{d_{j,k}}\right) \frac{a_{j,k}}{d_{j,k}}, \tag{4.40}$$

where φ' denotes the derivative of φ. In particular, when $\varphi \equiv id$,

$$\frac{\partial \mathcal{E}_{w_j,\{a_{j,k}\},d^0_{j,k}}}{\partial\{d_{j,k}\}}(d_{j,k}) = 0 \iff d_{j,k} = d^0_{j,k} + w_j \frac{a_{j,k}}{d_{j,k}}. \tag{4.41}$$

In Theorem 4.4.1 the role of the identity has been highlighted because it is the easiest choice for φ. Eq. (4.41) is an implicit equation that has to be solved using a numerical

method. In [92] Newton-Raphson's method has been used to find the zero of the function $F(d_{j,k}) \equiv d_{j,k} - d^0_{j,k} - w_j \frac{a_{j,k}}{d_{j,k}}$ by iteratively solving the equation

$$d^n_{j,k} = d^{n-1}_{j,k} - \frac{F(d^{n-1}_{j,k})}{F'(d^{n-1}_{j,k})}, \tag{4.42}$$

$n \geq 1$. Since the solution is not expected to differ too much from the original magnitude $d^0_{j,k}$, Newton-Raphson's algorithm is initialized with $d^0_{j,k}$.

A standard result guarantees the convergence of Newton-Raphson's algorithm as long as the initial condition $d^0_{j,k}$ is sufficiently near the solution, $F'(d^0_{j,k})$ is small enough, $F''(d_{j,k})$ varies smoothly and the inverse of $F''(d_{j,k})$ is bounded near the solution, see e.g. [21]. In particular, these conditions imply that one cannot take the weights w_j to be too big, otherwise Newton's algorithm can oscillate. In [91], it has been proven that if the identity function is substituted by the gamma function, then the stability of Newton-Raphson's algorithm in this setting is improved.

Notice that, since Newton-Raphson's method is initialized with the coefficients $d^0_{j,k}$, which are bounded from below, the solution of eq. (4.41) belongs to an open neighborhood of $d^0_{j,k}$ and the numerical method converges to a positive solution of eq. (4.41).

To stress the perceptual nature of the model, in Figure 4.4 it is reported a scan-line of the classical Mach-bands picture: the wavelet algorithm is able to reproduce the well-known undershoots and overshoots typical of the HVS behavior.

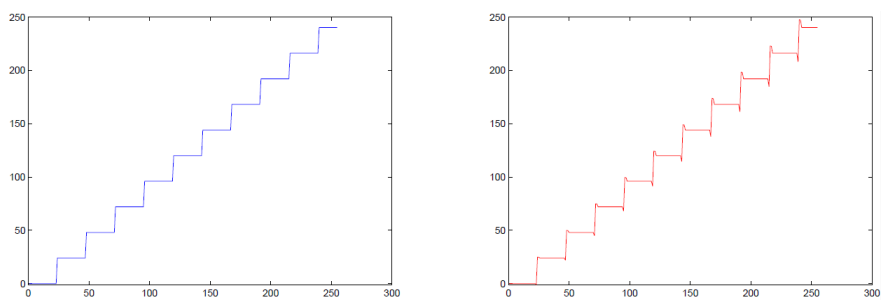


Figure 4.4: *From left to right*: a scan-line of the Mach-bands image before and after the wavelet algorithm, respectively. The undershoot and overshoot effects are typical HVS features that the wavelet model is able to reproduce. The wavelet algorithm was applied with following parameters: the mother wavelet is the Biorthogonal wavelet with 2 vanishing moments, the computation is performed over the maximum number of scales allowed for each image, $w_j = 0.5$, $T_j = \frac{\max_{k \in \Omega}\{d_{j,k}\}}{2.5}$ for each scale, and $\alpha = 0.1$.

The wavelet-based method just described can run in real time and it is as efficient as the spatially-based one. In Figure 4.5 the action of the wavelet algorithm can be seen on three images affected by distinct problems: under-exposure, color cast and over-exposure. As can be seen, the wavelet algorithm is able to perform a radiometric adjustment of the non-optimally exposed pictures and to strongly reduce the color cast, as proven in Figure 4.6.

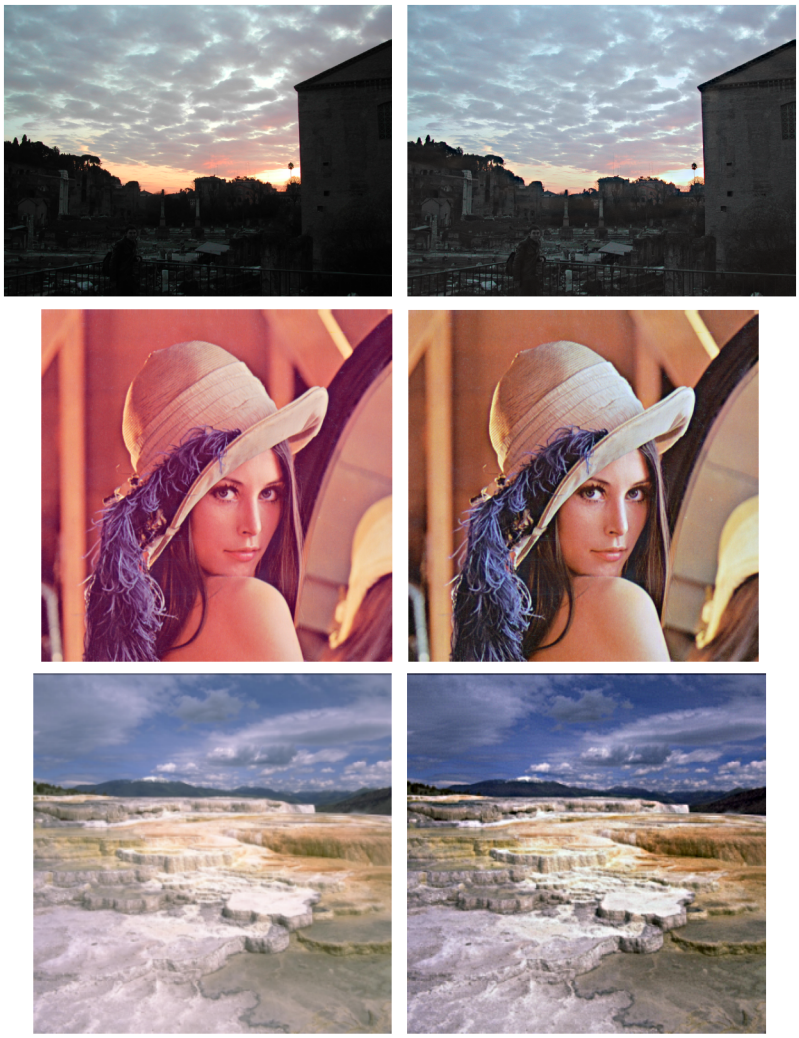


Figure 4.5: *Images on the left*: Original ones. *Images on the right*: enhanced versions after the wavelet algorithm: details appear in originally underexposed and overexposed areas, and the pink color cast in the 'Lena' image is removed. The filtering parameters are the following: the mother wavelet is the Daubechies wavelet with two vanishing moments, the computation is performed over the maximum number of scales allowed for each image, $w_j = 0.5$, and $T_j = \frac{\max_{k \in \mathfrak{I}}\{d_{j,k}\}}{10}$ for each scale 2^j.

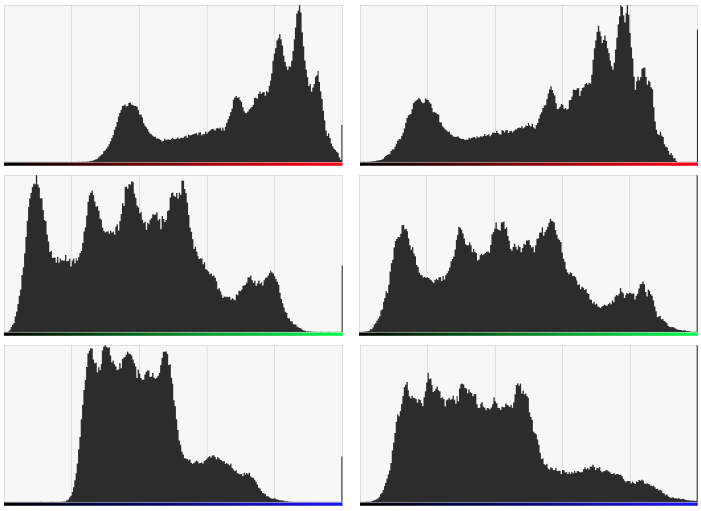


Figure 4.6: *Left column*: Originals RGB histograms of the image in the second row of figure 4.5. *Right column*: enhanced RGB histograms corresponding to the image in the second row of figure 4.5. Again, notice the histogram stretching without reaching a complete equalization, that would have been inappropriate.

Chapter 5

High dynamic range (HDR) imaging

Radiance in natural scenes can span several orders of magnitude, yet commercial cameras can capture only two orders of magnitude and reproduce Low Dynamic Range (LDR) photographs, often affected by saturated areas and loss of contrast and detail. During the last fifteen years, many techniques have been studied and proposed in order to expand the dynamic range of digital photographs to create the so-called High Dynamic Range (HDR) images, i.e. matrices whose entries are proportional to the actual radiance of the scene. For a thorough overview of the HDR imaging we refer the interested reader to [100].

HDR images have been widely used to capture scene illumination, but also for a realistic scene representation. That is to say, the wider dynamic range of HDR images allow to better capture details and color differences. This high precision is the reason why most photographers nowadays use cameras with a wider dynamic range than two orders of magnitude.

Two main problems remain to be solved in HDR imaging. Firstly, while the technique proposed by Debevec and Malick [23] can be considered as the 'de facto standard' for the creation of HDR images in *static conditions*, i.e. with perfectly still camera and without moving objects in the scene, no standard is available when these condition fail.

Secondly, HDR images cannot be entirely displayed or printed on the majority of commercial screens or printers, hence a so-called 'tone mapping' transformation is needed to properly reduce their range without losing details and respecting as much as possible the original color sensation.

In the following sections I will discuss my contributions in these two research fields.

5.1 Generation of HDR images in non-static conditions

We define the *radiance* $E(x)$ to be the electromagnetic power per unit of area and solid angle that reaches the pixel x of a camera sensor with spatial support $\Omega \subset \mathbb{R}^2$.

The *camera response function* f is the non-linear mapping $f : [0, +\infty) \to \{0, \ldots, 255\}$ that transforms the radiance $E(x)$ acquired in the time Δt into $I(x)$, the *intensity* value of the pixel x, i.e. $f(E(x)\Delta t) = I(x)$. f is assumed to be semi-monotonically increasing, i.e. its derivative is supposed to be non-negative, since the digital values stored by the camera remain constant or increase as the radiance increases.

In particular, the graph of f remains constant to 0 (black) for all the radiance values that

do not overcome the sensibility threshold of the camera sensor and to 255 (white) for all those radiance values that exceed its saturation limit.

For values that lie inside these two extremes, f is invertible and so we can use its inverse $f^{-1} : \{0, \ldots, 255\} \to [0, +\infty)$ to pass from image intensity values $I(x)$ to the corresponding radiance values $E(x)$: $f^{-1}(I(x)) = E(x)\Delta t$.

The camera response function can be built in controlled conditions by taking photos of a uniformly illuminated Macbeth chart with patches of known reflectance.

However, in many practical cases, the camera response function is not known, thus various techniques for recovering f or its inverse without a Macbeth chart have been developed.

These methods are called *chartless* and are all based on the same principle: since the camera sensor is limited by its sensibility threshold and saturation limit, in order to detect the whole scene radiance one has to take $N \geq 2$ shots with different time exposures Δt_j, $j = 1, \ldots, N$.

The next step consists in using the redundant information about the scene provided by the set of N digital images I_j to recover the inverse camera function f^{-1}. This is the only step that distinguishes the various algorithms proposed in the HDR imaging literature that we will briefly recall later.

Let us for the moment suppose that we have indeed computed f^{-1}, then we can construct the j-th *partial radiance* $E_j(x)$ as follows: $E_j(x) \equiv f^{-1}(I_j(x))\,/\Delta t_j\,,\ j = 1, \ldots, N$.

Notice that we use the adjective 'partial' when referring to the radiances E_j because they are built by applying f^{-1} to I_j, thus E_j cannot be considered a faithful representation of the entire radiance range, but only of the part corresponding to clearly visible details in the j-th image. Finally, in order to compute the set of final radiance values $\{E(x),\, x \in \Omega\}$ that will constitute the HDR image of the scene, one performs a weighted average of the partial radiances:

$$E(x) = \frac{\sum_{j=1}^{N} w(I_j(x))E_j(x)}{\sum_{j=1}^{N} w(I_j(x))}, \tag{5.1}$$

where the weights w take their maximum in 127, the center of the LDR dynamic range, and decrease as the intensity values approach 0 or 255, when the information provided by I_j is imprecise. If one deals with *color images*, then this procedure must be repeated three times in order to recover three radiance functions relative to the red, green and blue spectral picks of the scene radiance.

The algorithms for HDR image generation in static conditions are practically distinguished only for how they recover inverse camera response function f^{-1}.

Mann and Picard [66, 65] were the first to address this problem, they postulated a gamma-like analytical expression for f^{-1} and solved a curve fitting problem to determine the most suitable parameters.

Mitsunaga and Nayar [70] improved Mann and Picard's work by assuming f^{-1} to be a polynomial and determining its coefficients through a regression model. Debevec and Malik [23] worked with logarithmic data and did not impose any restrictive analytic form to f^{-1}, yet they required it to be smooth by adding a penalization term proportional to the second derivative of f^{-1} to the following optimization problem: $\text{Min}\| \log f^{-1}(I_j(x)) - \log E(x) - \log \Delta t_j \|^2$.

Their method is fast and gives reliable results, for this reason it has become the de facto standard in the field of HDR imaging with static conditions.

Let us now consider non static conditions, i.e. images taken with hand held camera and/or of dynamic scenes which are, by far, the most common. In this case, a motion detection step must be added to register the set of input LDR images and a more careful fusion process must be considered to avoid artifacts, in particular those called 'ghosts', i.e. objects with a translucent appearance induced by an incoherent image fusion, and 'bleeding', i.e. the diffusion of an artificial color over a flat image region, see Figure 5.1.

Figure 5.1: From left to right: an example of ghost and bleeding artifact, respectively.

Let us begin with registration: Ward [124] proposes to use multi-scale techniques to register a set of Median Threshold Maps (MTB), which are a binarization of the images with respect to their median value.

Although this approach is independent of the exposure time, it depends on noise and histogram density around the median value. Grosch [44] bases a local registration method on Ward's MTBs and Jacobs et al. [52] proposed an improvement of Ward's work by estimating motion with an entropy based descriptor.

These methods give good results for the case of small movements generated by a hand held camera, but they tend to produce artifacts when dealing with large object motion, as e.g. people walking through the scene. Tomaszewska and Mantiuk [114] and Heo et al. [47] propose to compute a global homography using RANSAC with SIFT descriptors which are based on gradients.

Finally, Kang et al. [55] propose to use the camera function to boost the original images in order to facilitate the registration process.

Let us consider now the methods that focus mainly on a suitable improvement of the fusion step. As we have said before, if we assume a perfect correspondence among pixels in non-static conditions and perform a weighted average of radiances, then ghosts artifacts appear in areas where motion has occurred.

Many methods in the state of the art, e.g. [56], [47], [41], [44] and [52], focus on avoiding ghost formation by modifying eq. (5.1) in order to reduce the influence of pixels corresponding to moving objects in the process of intensity fusion to generate the HDR image.

These methods may be robust to pixel saturation and small misalignments, but the areas that appear only in one image will be *copied* while the other image areas will be *averaged* from different images, thus new boundaries could be created in the process. In order to deal with these artifacts, Gallo et al. [37] propose to create a vector field by copying patches of gradients from the best exposed areas that match a reference image.

Then they blend the borders of neighboring patches and integrate the vector field. The resulting images are ghost-free but artifacts appear on the patch borders and flat regions.

Finally, let us report that gradient-based methods can deal with the radiance differences among the image sequence but are sensitive to misalignments and can produce color bleeding as Eden et al. [29] have pointed out.

5.2 HDR generation based on gradient fusion

To avoid the problems discussed above, instead of synthesizing a new HDR image from the original sequence of LDR images, we will select just one of them to be the *reference* and then we will improve the associated radiance map by adding as many details as possible without introducing artifact. In doing this, a fundamental role will be played by gradient fusion, which has several advantages with respect to intensity fusion in this case.

The method that we propose has several steps in cascade, before describing in detail all the steps it is worthwhile to give a brief summary of the whole algorithm. Firstly, as done in the static scenario, we apply the inverse camera function to every LDR image of the bracketing of pictures taken with different time exposures, obtaining the partial radiance map for each one of them. These radiance maps are radiometrically aligned using the camera function with the exposure time of the reference image. These modified images are used to compute a dense correspondence field, using the optical flow algorithm of Chambolle and Pock [19], for every image with respect to the reference one. These correspondences are subsequently filtered using a refinement step: for every pair of images, we compensate the motion between them and compute the absolute value image difference. The histogram of this image difference is modeled as a mixture of Gaussians, which allows us to distinguish between correct and erroneous correspondences. Finally, we use the corrected correspondences to obtain a gradient field that we integrate by solving a Poisson equation. To avoid color bleeding and other color artifacts we set a randomly selected group of points as Dirichlet boundary conditions. The intensity values of these points are computed through the Debevec-Malik intensity fusion of the aligned pixels.

The first step of the algorithm has already been described in the previous section, we present and discuss the following steps in separated subsections.

5.2.1 Radiance Registration

As previously declared, the goal of our method is to increase the detail rendition of the reference radiance map in the areas where the original LDR reference image was over/under exposed. The aim is to obtain these missing details from other images of the bracketing without creating artifacts produced by merging motion pixels. Thus, we need to distinguish motion pixels that we do not want to fuse from details that appear in other images (but not in the reference one) that we actually want to fuse.

A registration algorithm based on intensity or geometry comparison cannot solve this problem by its own, thus we need to modify the radiance maps before operating the registration: we need a *pre-registration step* that modifies the radiance maps so that the registration process will match the most precise details, i.e. those coming from dark areas of overexposed images and from bright areas of underexposed images.

Our tests have shown that the more efficient radiometric modification for this scope is given by the application of the camera response function with the (fixed) time exposure of the reference image, to each radiance map, as also observed by [55]. Thus instead of I_j we work with the modified images $f(\Delta t_{\text{ref}} E_j(x))$. Notice that the reference image remains unchanged

while for the others the transformation amounts to a normalization of the exposure time. Notice that, since the range of f is $\{0, \ldots, 255\}$, after this process the radiance maps become a new set of LDR images. The advantage of this transformation is that the intensity levels become closer and that the image areas corresponding to the over/under-exposed values of the reference image also appear saturated in the other radiometrically aligned images.

After this pre-registration step, we can apply any optical flow method over the new set of LDR images. We used the method of Chambolle and Pock [19] which provides dense correspondences between pixels. Although the matchings are accurate, the method can produce errors on the boundaries of moving objects and the areas where objects disappear due to motion. Thus a refinement step is needed to check whether the correspondences are correct or not.

5.2.2 Refinement step

Let us assume that we have two radiometrically aligned images I_{ref}, I_j, $j = 1, \ldots, N$, without motion. The pixel-wise distance between them defines the difference image $\|I_{\text{ref}}(x) - I_j(x)\|$ that depends on their noise. If we model this noise as additive and Gaussian, then the difference image histogram will be highly populated close to zero. As the motion (or mismatches) among the images increases, more modes appear in this histogram. We therefore propose modeling this difference histogram as a set of Gaussians using a Gaussian Mixture Model (GMM) where the most probable Gaussian is associated to the properly matched pixels and the other Gaussians represent matchings that are not correct. Let the most probable Gaussian be characterized by a median μ and a standard deviation σ. We define as a mismatch those correspondences with a distance from μ given by $\alpha\sigma$, where α is a parameter of the algorithm. After the refinement step, a pixel of the reference radiance map can have a set of correspondences in pixels of the other images of the sequence or no correspondence at all if the pixel belongs to an object that appears only in the reference image.

5.2.3 Gradient Fusion

After the steps previously discussed we know which pixels can be fused without generating artifacts in the final HDR image. Here we examine the process that will add details to the reference radiance map from the others in the set. Let us begin by observing that in general we do not have correspondences between all pixels of the reference image and pixels in all the other images of the bracketing, thus an intensity fusion can create artificial edges, as can be seen in Figure 5.2.

Note that the people are moving along the sequence, thus, their intensity values are being copied from the reference image to the final image while the other areas are computed by a weighted average among the corresponding pixels. This difference generates new edges around the subjects in motion. In order to avoid this problem, we propose using gradient fusion techniques in the Log-scale. Let us set for this scope $\tilde{I}_j = \log_{10}(I_j)$, $j = 1, \ldots, N$.

Poisson Editing [83] allows modifying the features of an image without creating new edges or artifacts in the LDR domain. The idea is to copy a target gradient field in a region Ω' of the image domain which is then integrated by solving a Poisson equation. Thus, the problem is to properly choose the target gradient field. To do that, recall that our aim is to improve the detail rendition of the final HDR image, so we must give more importance to gradients with largest norm. As discussed in [86], if we consider the weighted second moment matrix

Figure 5.2: Detail of a result obtained by building radiance maps with an intensity fusion process, it can be seen that artificial edges appear (in particular around the moving people). Image (courtesy of O. Gallo) shown with HDRShop at a fixed time exposure.

$G_s(x)$:

$$G_s(x) = \begin{pmatrix} \sum_{j=1}^{N} \left(s_j(x) \frac{\partial \tilde{I}}{\partial x_1} \right)^2 & \sum_{j=1}^{N} s_j^2(x) \frac{\partial \tilde{I}}{\partial x_1} \frac{\partial \tilde{I}}{\partial x_2} \\ \sum_{j=1}^{N} s_j^2(x) \frac{\partial \tilde{I}}{\partial x_2} \frac{\partial \tilde{I}}{\partial x_1} & \sum_{j=1}^{N} \left(s_j(x) \frac{\partial \tilde{I}}{\partial x_2} \right)^2 \end{pmatrix} \tag{5.2}$$

where the partial derivatives are computed at the point $x = (x_1, x_2)$ and

$$s_j(x) = \frac{\|\nabla \tilde{I}_j(x)\|}{\sqrt{\sum_{i=1}^{N} \|\nabla \tilde{I}_i(x)\|^2}}, \tag{5.3}$$

then the predominant gradient direction is given by the largest eigenvector of $G_s(x)$ and the corresponding eigenvalue gives its norm. In this process the sign of the gradient is lost. In our experiments, we have seen that the best results are achieved when we restore the sign to be that of the gradient with maximum modulus.

We can now define the target vector field for Poisson editing as follows: given a pixel x in the reference image, from all gradients of the bracketing of images at x, we select the gradient with largest modulus. Let us denote it by $M(x)$. Then the target vector V in x is defined as:

$$V(x) = \sqrt{\lambda} \varepsilon(x) \theta, \tag{5.4}$$

where λ is the largest eigenvalue of $G_s(x)$, θ its associated eigenvector, and

$$\varepsilon(x) = \begin{cases} -1 & \text{if } \langle \theta, M(x) \rangle \leq 0; \\ 1 & \text{otherwise.} \end{cases}$$

We stress that $V(x)$ is not necessarily a conservative vector field and, as remarked by Tao et al. [113] this can lead to bleeding effects along strong edges. In order to stop the bleeding effect, we have introduced Dirichlet boundary conditions at a set of randomly chosen points of Ω taken from the set of points for which there exist reliable correspondences. For them we have set the intensity as the radiance obtained with the Debevec-Malik intensity fusion.

5.2.4 Results and comparisons

Now that the method has been described we can proceed to give some implementation detail. The optical flow method of Chambolle and Pock was applied to the luminance values of the radiometrically aligned images. For the refinement step we modeled the histogram of image differences as a GMM (Gaussian Mixture Model) with two components. A final step in the refinement process is introduced to avoid taking into account pixels on the boundaries: interpreting the mismatches as motion masks, we apply to these masks a dilation filter with a 6×6 structuring element [40]. In the fusion process we randomly select the 10% of the total image area as possible Dirichlet points keeping only those with matches. Finally, we solve the Poisson equation using the Conjugate Gradient method for each color channel independently.

Let us now discuss the results of our proposal. Since HDR images cannot be represented on a LDR display, to show the results obtained from our algorithm we will show snapshots of the image provided by the free HDRShop software, available online at www.hdrshop.com, at a given exposure time. The comparison with the state of the art will be done following the procedure used in other papers, that is, by comparing the tone mapped versions of the HDR images.

First of all we would like to start by showing the improvement obtained using our method. In Figure 5.3 we can see two original partial radiance maps compared to our results. The enhancement is specially obvious on the building wall (first row), where our method recovers texture and color, and also in the trees, which show much more details (second row).

Figure 5.3: Left: original reference radiance map in the HDRshop software. Right: enhanced image with the proposed imethod. The parameter α was set to 1.

The parameter α controls the flexibility in the selection of matches in the registration process. Our experiments have shown that a value of α that gives overall good performances is 0.5, for which we have found no artifacts in the final images. As we approach 1 we generate more details, but we might also create artifacts, as can be seen in Figure 5.4.

The selection of the reference image can also influence the appearance of ghosts or artifacts. We have in fact observed that when a moving pixel with a color similar to the background is located in an area saturated by the camera function, then the refinement step may fail to distinguish background from foreground and color artifacts may appear. We can see an example in Figure 5.5: in the first row we present two images of the bracketing that were used as two different reference images. The corresponding results are shown in the second row. In both cases the parameter value is $\alpha = 1$. Note that the result shown on the left hand side has an artifact on the puppet's leg. The reason is that the reference image is saturated in that area, so the refinement step cannot distinguish between the puppet's leg and the ball.

Finally, we show the comparison between our results and those of Gallo et al. [37] who also

Figure 5.4: From left to right: results of our algorithm obtained by setting $\alpha = 0.5$ (no artifacts) and $\alpha = 1$ (artifacts appear), respectively.

use a gradient fusion technique. Since they present their results using tone mapped versions of their images, we also used a tone mapped version of ours. In Figure 5.6 we can see that our method avoids the generation of geometry artifacts as well as bleeding.

5.3 Tone mapping

HDR radiances may span many orders of magnitude, while common displays and printers usually span up to two [1]. So, a further compression step, called *Tone Mapping* (TM), is required to properly visualize the information stored in HDR images.

Many tone mapping operators (TMOs) have been proposed in the literature; for a thorough review and analysis of the state of the art until 2005, we refer to [100].

Here we just want to give a very brief overview of the different schools of thought that have been proposed so far. There is a category of perceptually-basedTMOs which can be either spatially global or local. The former are, in general, very fast and do not introduce halos or artifacts, but their contrast rendition tends to be poor. The first global TMOs used Stevens' law [3] to achieve range compression and were proposed by Tumblin and Rushmeier [115], Chiu et al. [20], and Ward et al. [123].

Schlick [108] proposed a rational function very close to the Michaelis-Menten formula (2.1) showing improvements with respect to Stevens' law [40]. The global Michaelis-Menten formula has been exploited by Pattanaik et al. [80], Reinhard and Devlin [98], and more recently, by Kuang et al. [60]. More sophisticated vision models were also proposed, taking into account time adaptation such as Ferwerda's model [35] or Weber-Fechner's law, e.g., the methods by Ward et al. [123] and Ashikhmin [6].

Reinhard et al. [99] and Tamburrino et al. [112] introduced local TMOs that improved detail but created halos and artifacts next to edges. A TMO based on a modified version of the Retinex model of color vision was proposed by Jobson et al. [54], and the anchoring theory1 has been used by Krawczyk et al. [59] after a suitable subdivision of the original image into layers of similar luminance.

There is another category, that of gradient-based TMOs, which rely on the idea of shrinking large intensity gradients while preserving small fluctuations, corresponding to fine details.

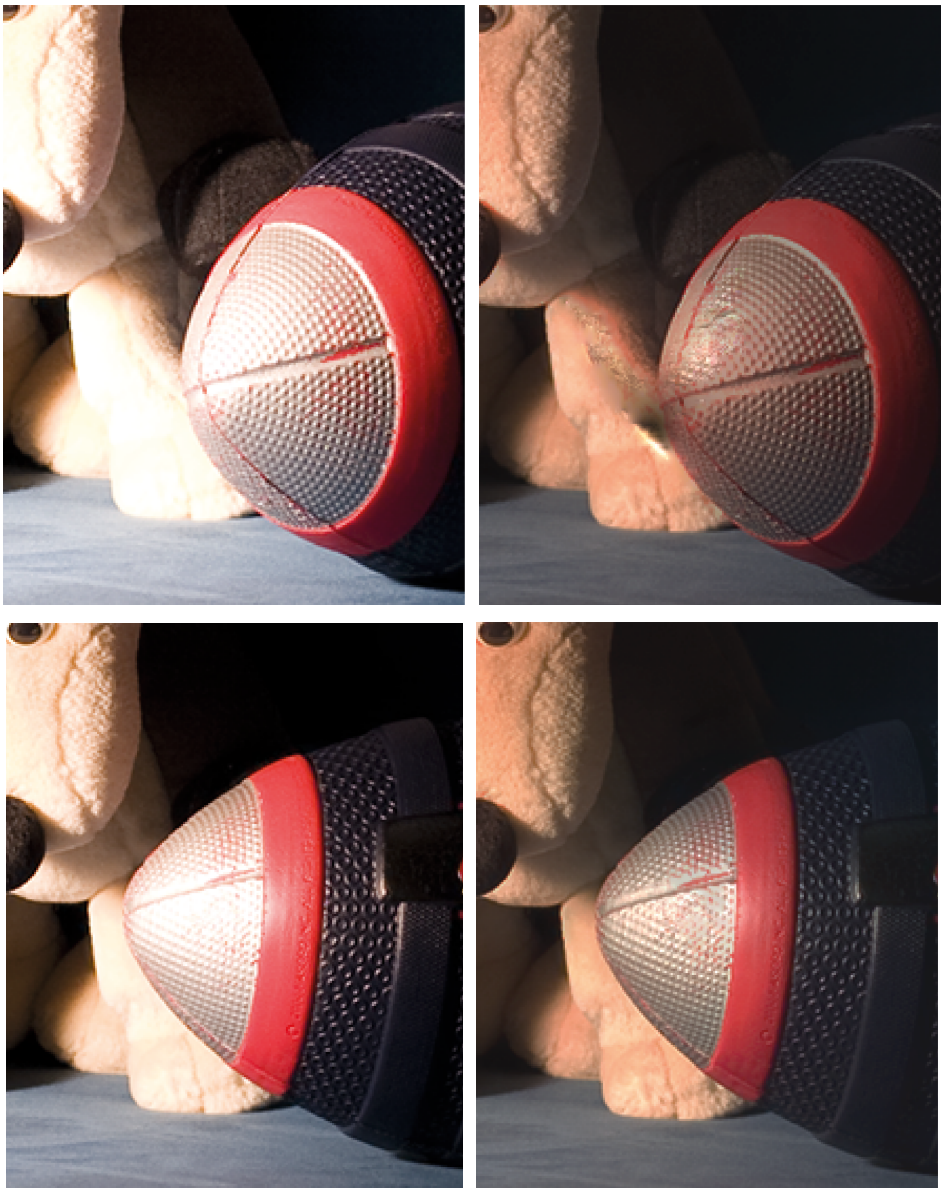


Figure 5.5: First column: two different original images taken as reference. Second column: output of our algorithm. It can be noticed an artifact in the left hand image due to the failure of the refinement step explained in the text.

Tumblin and Turk [116] used a hierarchical method based on a Partial Differential Equation (PDE) inspired by anisotropic diffusion.

Durand and Dorsey [28] and Kuang et al. [60] obtained improved results by using techniques inspired by bilateral filtering.

Fattal et al. [31] and Mantiuk et al. [67] proposed methods where spatially varying compression factors were used to implement suitable manipulations of the gradient field.

In [32], we followed the TM philosophy proposed by Ward et al. [123], which is that the main purpose of TM is to emulate as much as possible the perception of contrast and color

Figure 5.6: First row, from left to right: tone mapped results of our HDR generation model and that of Gallo et al., respectively. Second row: detail magnification of the pictures above. The color difference is due to the different tone mapping operators used. We stress that the artifact visible in the image of Gallo et al's is not produced by the tone mapping operator, but by their HDR formation model.

produced by the real-world scene. Lacking a perfect model of the Human Visual System (HVS), we concentrate instead on well-established facts of visual perception (experimental data on visual adaptation and contrast perception) to propose our TM operator.

While visual adaptation is considered a retinal process, local contrast enhancement (and color constancy) is thought to involve also higher stages of the visual cortex, see e.g. Hubel's book [49]. This suggests that the processes involved in these two phases are quite different; therefore, we have decided to mimic them (or rather their outputs) in two independent stages of our algorithm.

5.3.1 A two-stage tone mapping

Neuroscience experiments to measure visual adaptation have been performed using very simple, non natural images: brief pulses of light with intensity I were superimposed on a uniform background.

As already recalled, when a photoreceptor absorbs a photon, the electric potential of its membrane changes accordingly to the empirical law known as the Michaelis-Menten equation (2.1), that we replicate here for clarity:

$$r(I) = \frac{\Delta V}{\Delta V_{\max}} = \frac{I^{\gamma}}{I^{\gamma} + I_S^{\gamma}}, \tag{5.5}$$

as said in chapter 2, I_S is the light level at which the photoreceptor response is half maximal, called the semisaturation level, and which is usually associated with the level of adaptation.

Each type of cone is most sensitive over a particular waveband, thus the semisaturation constant must depend on the amount of light in this same waveband that reaches it, not on the luminance of the light source, as Boynton [16] and Wade and Wandell [121] point out.

The change of electric potential ΔV is the photoreceptor's physiological response to I, $V_{\max}$ is the highest difference of potential that can be generated, and γ is constant, measured as 0.74 for the rhesus monkey [117].

We should point out that this change in electric potential describes the light response of one single living photoreceptor isolated in a tiny glass pipette, and it is not the response of the cell embedded in the retina.

Adjacent photoreceptors strongly affect each other both chemically and electrically: none of the retina's approximately 125 million photoreceptor cells respond independently, and the brain receives no light meter-like signals from them. Instead, three layers of retinal neurons interconnect to construct 20 or more different kinds of signals that summarize local changes in retinal illumination, the receptive fields.

Little more than 1 million retinal ganglion cells send these change-summaries toward the brain by varying their firing rates (1-100 Hz). A bundle of their extremely long (2-3 cm) individual axons comprise the optic nerve bundle that leaves the back of each eyeball, even including response signals traveling from the brain back to the retina.

Equation (2.1) has been used extensively in the TM literature, as mentioned in the previous section. Although not been explicitly stated, the underlying assumption here is that the output of eq. (2.1), a ratio of voltage amplitudes of electrical responses (or the ratio of light intensities), which is a physical magnitude, is assumed to be correlated with the perceived brightness, which is a sensation, a perceptual magnitude.

To recap, each TMOs which use (2.1), or variations of it, take an HDR input intensity and give an LDR output trying to mimic the perceived brightness.

On the side of perception of intensity changes, we have Weber-Fechner's law of chapter 2, i.e. the logarithmic behavior of the sensation of light intensity: $s(I) = s_0 + k \log(n + I)$.

As pointed out in [100], this logarithmic transformation is probably the simplest procedure one might use as a TMO, and it works fairly well for medium-dynamic-range images. Nonetheless, this equation is questionable and still contested in the psychophysics community, where researchers disagree on the validity of accumulating and/or integrating JNDs to assess supra-threshold stimuli of any sort.

Now we have two equations, Michaelis-Menten's (MM) and Weber-Fechner's (WF), which are different but which are used (directly or with some modifications) under the assumption that they are both modeling the same thing, namely, the perceived brightness.

This obviously cannot be, so let us look at this matter in more detail. Michaelis-Menten (MM) and Weber-Fechner (WF) equations certainly do not match, in fact the two equations mathematically model results of markedly different experiments that measured very different visual phenomena. WF describes detectable-difference thresholds for small, steady-state stimuli; users adapt to a steady, uniform background, and experiments measure the tiniest steady difference we can distinguish from that background.

MM models electrical responses to flashed stimuli; it measures how well we can detect tiny differences in intensities far away from that of the adapting background (and measured before our adaptation level can change). Since MM does not match with WF, we cannot claim that a TMO based on the NR equation accurately models perception in the HVS.

But this would seem to be only an issue from a theoretical point of view because theTM-results obtained with these NR-like methods, based on modifications of (1), are quite good and usually better than those obtained with the logarithmic mapping corresponding to WF; their overall appearance seems closer to our own perception.

To try to understand this, let us recall that TMOs map an arbitrarily large radiance range into image values in a limited range, without loss of generality we may assume that the output range is the interval $[0, 1]$. The values given by MM are already in this interval, but for WF, it is necessary to apply a normalization step to the output.

The outputs of MM and WF are fairly similar for the first three or four orders of I, but then the output of MM tends to 1 while the output of WF increases indefinitely.

The fact that MM tends to 1 for high-radiance values is explained by the saturation property of photoreceptors: the activity of cones covers approximately four orders of magnitude around the semisaturation value, as it is explained by Valeton and van Norren [117], and for higher radiances the cones saturate.

The saturation of the cones is not considered in WF, and therefore the corresponding curve always increases. Taking all the above into account, we propose the following: since cones are indeed saturating, in order to model the perceptual response we must not use the curve obtained with steady stimuli fields because these fields do not cause saturation in the cones and we end up with the logarithmic curve, which increases indefinitely.

Instead, we must use the curve obtained with pulsating fields. This complies with WF in an operative range of four orders of magnitude, but for higher radiances increases very steeply because the cones saturate at those radiance values.

Dunn et al. [27], while talking about visual adaptation and saccades, say the following: 'retinal mechanisms [...] adjust sensitivity in the 200 ms intervals between saccades,' and 'As we make saccades to explore a visual scene, retinal neurons encounter a wide range of light intensities. Receptor and postreceptor adaptation permit the amplification required to see objects in shadows while avoiding saturation from the sky. The combination of these adaptive

mechanisms allows the visual system to encode details in a scene with greater fidelity than a standard camera at a single exposure setting'.

The implication would be that while viewing real-world scenes, as we frequently use saccadic movements to look for and track objects or simply to gather all of the details, the stimuli fields are not steady but pulsed. Hence, the curve that more accurately matches perception in real-world images would be the upper curve in Figure 5.7.

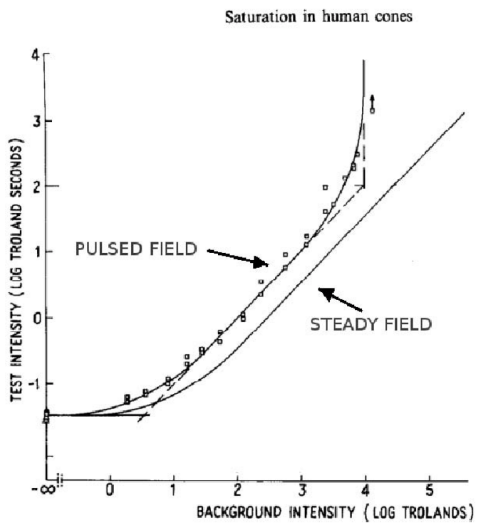


Figure 5.7: Cones saturate when the stimuli field is not steady but pulsed: perceptual curves of cones for pulsed and steady stimuli, figure adapted from Shevell [110].

In short, from Figure 5.7 there seems to be an inflexion point at the radiance value located approximately two orders of magnitude above the semisaturation radiance, which we will call I_M: for radiances below I_M, Weber-Fechner's law holds, so the perceptually valid tone mapping curve is the one given by WF; for radiances above I_M, the cones' responses start to saturate so the perceptually (and physiologically) valid tone mapping curve must have an horizontal asymptote, and for this we can use the curve given by the MM.

With this information, we can now introduce the first stage of our algorithm. We propose to use a TMcurve $\mathcal{C}(\mathcal{I})$ which combines the curves MM and WF in the aforementioned way:

$$\mathcal{C}(I) = \begin{cases} s_0 + k\log(I+n) & \text{if } I \leq I_M \\ \frac{L^\gamma}{L^\gamma + L_S^\gamma} & \text{if } I > I_M \end{cases} \tag{5.6}$$

where s_0 is chosen so as to ensure the continuity of c at the critical point I_M. the output of c will be normalized, i.e., linearly scaled so that it spans the interval $[0, 1]$.

Fig. (5.8) shows the TM results obtained using MM, WF and $\mathcal{C}(\mathcal{I})$ on the same HDR image. The left-hand image, obtained using MM, has an overall good contrast but the brightest regions appear overexposed (e.g., the areas near the windows or the trees); the middle image, obtained with the logarithmic mapping of WF, has overall poor contrast but the details of the brightest regions are not lost.

The right-hand result is the one obtained with the *first stage* of our proposed method, $\mathcal{C}(\mathcal{I})$, and we see that it combines the best characteristics of the other two images, showing overall good contrast and little or no loss of detail in the brightest areas.

Figure 5.8: Left: TM output using the Michaelis-Menten equation. Middle: TM output using a logarithmic mapping (Weber-Fechner's law). Right: TM output of the first stage of our method, eq. 5.6.

The *second stage* consists simply in the application of the perceptually-inspired variational algorithm for color correction discussed in chapter 4, which is associated to higher stages of visual perception and will produce a local enhancement of the first stage, also allowing the reduction of a possible color cast due to a non-neutral illuminant in the photographed scene.

All the implementation details of the two-stage tone mapper and the comparison with the state of the art TMOs are given in [32], here I just want to stress the fact that the proposed tone mapper has the best overall performance with respect one of the most vastly used quality measure [9].

Chapter 6

Statistics of natural images

The Human Visual System (HVS from now on) has evolved to optimize the elaboration and transmission of visual signals originating from natural scenes. Since the whole optimization process is complex and still not fully understood, here we will only deal with second-order redundancy reduction. Two kind of redundancies can be distinguished in the interaction between humans and natural scenes: firstly, in natural scenes nearby points are likely to send similar radiance information to the eyes (unless they lie in the proximity of a sharp edge), which implies a strong *spatial correlation*. Secondly, light signals are absorbed by the three L, M, S-type cones, whose sensitivity is not independent, as can be seen by looking at their spectral sensitivity functions $L(\lambda), M(\lambda), S(\lambda)$, depicted in Figure 6.1. This implies a strong *chromatic correlation*. When both effects are taken into account, one speaks about *spatio-chromatic correlation*.

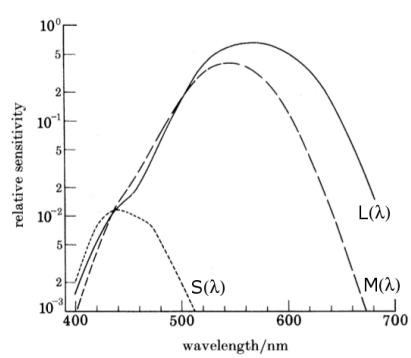


Figure 6.1: The Vos-Walraven cone sensitivity functions (adapted from [18], page 92).

Known results of second order statistics between pixel values are the *Fourier-like structure* of Principal Component Analysis (PCA), a result of *spatial stationarity*, and the *power-law decay* of the covariance, as a possible consequence of *scale-invariance*. Higher order statistics have also been largely investigated, for instance through wavelets or sparse coding.

On the other hand, several works have been concerned with chromatic redundancy in images, mostly through second order property and in connection with opponent color spaces.

However, the spatio-chromatic structure of color images has been less studied. One of the most striking known empirical observation is that the spatio-chromatic covariance matrices resemble a tensor product between a Fourier basis and color opponent channels, as pointed

out in section 6.1.

In this work, we focus on this statistical characteristic, both from a theoretical and an experimental perspective, proving that two simple assumptions on the nature of spatiochromatic covariance matrices of real-world images are enough to explain the reason for the appearance of the tensor product structure.

In order to have a better perspective on this result, we will first start by recalling the most relevant results of second order natural image statistics related to this work.

6.1 Brief state of the art in second order natural color image statistics

The literature about natural image statistics is vast and its exhaustive presentation is far beyond the scope of this paper. Here we will emphasize only the results from [18] and from [106], which are essential to understand our results.

Before describing these works, let us recall that when principal component analysis (PCA) is performed on small natural image patches, the basic features that result are Fourier descriptors, see for instance [75]). This fact is a consequence of spatial stationarity. More involved patch decompositions, relying on the minimization of redundancy, as in [7], or on sparse decompositions, as in [75], yield localized, band-pass and oriented filters resembling wavelet decompositions. Analogous elementary patches have been obtained with the use of Independent Component Analysis (ICA), see e.g. [51].

6.1.1 Chromatic redundancy in natural images

The first statistical information about chromatic redundancy has been experimentally obtained in [73] in the framework of color segmentation of RGB images. For each picture of a database of 8 RGB images, the authors computed the covariance matrix C of the distribution of the values of R, G and B at each pixel. If $\lambda_1, \lambda_2, \lambda_3$ are the eigenvalues of C, with $\lambda_1 \geq \lambda_2 \geq \lambda_3$, and if the column vectors $\mathbf{v}_i$, $i = 1, 2, 3$, are the corresponding eigenvectors, then the color features defined by the inner product $X_i = (R, G, B) \cdot \mathbf{v}_i$, are uncorrelated and orthogonal.

They found that the eigenvectors of the covariance matrix are approximately the following ones for each image of the database: $\mathbf{v}_1 = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)^t$, $\mathbf{v}_2 = \left(\frac{1}{2}, 0, -\frac{1}{2}\right)^t$, $\mathbf{v}_3 = \left(-\frac{1}{4}, \frac{1}{2}, -\frac{1}{4}\right)^t$. These vectors correspond to the three following uncorrelated color features: $X_1 = \frac{R+G+B}{3}$, $X_2 = \frac{R-B}{2}$, $X_3 = \frac{2G-(R+B)}{4}$.

This shows that the feature that corresponds to the largest variance is the luminance X_1 (or *achromatic channel*) and the other two features are described by the *opponent channels* X_2 (red-blue) and X_3 (green-violet).

[18] approached the problem of finding uncorrelated color features from a purely theoretical point of view. Following the already quoted ideas of Attneave, Barlow and MacKay, they analyzed the problem of an efficient post-retinal information transmission by performing a PCA on the LMS *cone activation values*. We shall now give a detailed presentation of this work, to which our contributions are closely related.

Buchsbaum and Gottschalk considered the abstract ensemble of all possible visual stimuli (radiances), i.e. $\mathcal{S} \equiv \{S(\lambda),\ \lambda \in \mathcal{L}\}$, where $\mathcal{L}$ is the spectrum of visible wavelengths. From a given representative $S(\lambda) \in \mathcal{S}$, a weighted integration of $S(\lambda)$ over the visual spectrum, with

weights given by the Vos-Walraven spectral sensitivity functions $L(\lambda), M(\lambda), S(\lambda)$ depicted in Figure 6.1, yields the three cone activation values $L = \int_{\mathcal{L}} S(\lambda)L(\lambda)\,d\lambda$, $M = \int_{\mathcal{L}} S(\lambda)M(\lambda)\,d\lambda$, $S = \int_{\mathcal{L}} S(\lambda)S(\lambda)\,d\lambda$.

Assuming that the stimulus $S(\lambda)$ (coming from a fixed point $\bar{x}$ of a scene) is a random variable, a convariance matrix can be build from the three random variables L, M, S. This matrix, called the *chromatic covariance matrix* is defined as:

$$C = \begin{bmatrix} C_{LL} & C_{LM} & C_{LS} \\ C_{ML} & C_{MM} & C_{MS} \\ C_{SL} & C_{SM} & C_{SS} \end{bmatrix}, \tag{6.1}$$

where $C_{LL} \equiv \mathbb{E}[L \cdot L] - (\mathbb{E}[L])^2$, $C_{LM} \equiv \mathbb{E}[L \cdot M] - \mathbb{E}[L]\mathbb{E}[M] = C_{ML}$, and so on, $\mathbb{E}$ being the expectation operator.

Let $K(\lambda, \mu) = \mathbb{E}[S(\lambda)S(\mu)] - \mathbb{E}[S(\lambda)] \cdot \mathbb{E}[S(\mu)]$ be the *covariance function*, then the entries of the covariance matrix can be written as $C_{LL} = \iint_{\mathcal{L}^2} K(\lambda, \mu)L(\lambda)L(\mu)\,d\lambda d\mu$, and so on.

To be able to perform explicit calculations, the analytical form of the covariance function $K(\lambda, \mu)$ must be specified. In the absence of a database of multispectral images, Buchsbaum and Gottschalk used abstract non-realistic data to compute $K(\lambda, \mu)$. They chose the easiest covariance function corresponding to visual stimuli maximally uncorrelated with respect to their energy at different wavelengths, i.e. $K(\lambda, \mu) = \delta(\lambda - \mu)$, δ being the Dirac distribution. As the authors observe, this condition is satisfied only if the ensemble $\mathcal{S}$ is made of monochromatic signals.

With this choice, the entries of the covariance matrix C are all positives and they can be written as $C_{LL} = \int_{\mathcal{L}} L^2(\lambda)\,d\lambda$, $C_{LM} = \int_{\mathcal{L}} L(\lambda)M(\lambda)\,d\lambda$, and so on. C is also real and symmetric, so it has three positive eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ with corresponding eigenvectors $\mathbf{v}_i$, $i = 1, 2, 3$. If W is the matrix whose columns are the eigenvectors of C, i.e. $W = [\mathbf{v}_1|\mathbf{v}_2|\mathbf{v}_3]$, then the diagonalization of C is given by $\Lambda = W^t C W = \mathrm{diag}(\lambda_1, \lambda_2, \lambda_3)$.

The eigenvector transformation of the cone excitation values L, M, S, in the special case of monochromatic stimuli, is then

$$\begin{pmatrix} A(\lambda) \\ P(\lambda) \\ Q(\lambda) \end{pmatrix} = W^t \begin{pmatrix} L(\lambda) \\ M(\lambda) \\ S(\lambda) \end{pmatrix}.$$

The transformed values A, P, Q are *uncorrelated* and their covariance matrix is Λ. A is the achromatic channel, while P and Q are associated to the opponent chromatic channels.

The *key point* in Buchsbaum and Gottschalk's theory is the application of *Perron-Frobenius theorem* (see e.g. [13] for more details), which assures that positive matrices, i.e. matrices whose entries are all strictly greater than zero, have one and only one eigenvector whose entries have all the positive sign, and this eigenvector corresponds to the largest eigenvalue, i.e. λ_1. So, only the transformed A channel will be a linear combination of the cone activation values L, M, S with positive coefficients, while the channels P and Q will show opponency. This is the theoretical reason underlying the evidence of post-retinal chromatic opponent behavior, following Buchsbaum and Gottschalk.

The monochromatic signal energy of the channels has the following property:

$$\int_{\mathcal{L}} A^2(\lambda)\,d\lambda : \int_{\mathcal{L}} P^2(\lambda)\,d\lambda : \int_{\mathcal{L}} Q^2(\lambda)\,d\lambda = \lambda_1 : \lambda_2 : \lambda_3. \tag{6.2}$$

The explicit form of the matrices C, W^t and Λ within Buchsbaum-Gottschalk's theory are the following:

$$C = \begin{pmatrix} 77.0622 & 38.6204 & 0.0649 \\ 38.6204 & 22.8099 & 0.0646 \\ 0.0649 & 0.0646 & 0.0151 \end{pmatrix}, \tag{6.3}$$

$$W^t = \begin{pmatrix} 0.887 & 0.461 & 0.0009 \\ -0.46 & 0.88 & 0.01 \\ 0.004 & -0.01 & 0.99 \end{pmatrix}, \tag{6.4}$$

$$\Lambda = \mathrm{diag}(97.2, 2.78, 0.015). \tag{6.5}$$

The large covariance values between L and M and the very small ones between these two channels and S are a direct consequence of the use of Vos-Walraven's sensitivity functions and the hypothesis $K(\lambda, \mu) = \delta(\lambda - \mu)$. In Section 6.3 we will see that if we compute C from a database of natural images, then the difference among covariance values is rather small.

Using the data obtained above, Buchsbaum and Gottschalk could write explicitly the transformation from (L, M, S) to (A, P, Q) as follows:

$$\begin{cases} A \simeq 0.887L + 0.461M \\ P \simeq -0.46L + 0.88M \\ Q = 0.004L - 0.01M + 0.99S, \end{cases}$$

the energy ratios among A, P and Q being $97.2 : 2.78 : 0.015$. Again, we observe that the unrealistic hypothesis of maximally uncorrelated visual signals implies that the achromatic channel accounts for the great majority of the energy transmitted and the blue channel has practically no influence in the computation of the achromatic stimulus.

6.1.2 Spatio-chromatic redundancy in natural images

The earliest attempts to build a computational model of spatiochromatic image coding in early vision are represented by [26] and [7]. However, the most influential paper in this field is [106], where Ruderman, Cronin and Chiao proposed a *patch-based* spatio-chromatic coding and tested Buchsbaum-Gottschalk's theory on a database of 12 multispectral natural images of *foliage*.

We now give a detailed account the experiments performed in [106]. First of all, the authors have shown that the scatterplots in the LM and LS planes of the L, M, S cone activations values (corresponding to 1000 pixels randomly selected in the database) show a high degree of correlation but also asymmetry.

The authors decided to study these data by first reducing their asymmetry: they modified the LMS values by taking their decimal logarithm and then they subtracted the average image value in the logarithmic domain. They obtained the so-called *Ruderman-Cronin-Chiao coordinates*, i.e. $\tilde{L} = \mathrm{Log}\, L - \langle \mathrm{Log}\, L \rangle$, $\tilde{M} = \mathrm{Log}\, M - \langle \mathrm{Log}\, M \rangle$ and $\tilde{S} = \mathrm{Log}\, S - \langle \mathrm{Log}\, S \rangle$. This transform is motivated with the fact that, following Weber-Fechner's law, uniform logarithmic changes in stimulus intensity tend to be equally perceptible, see [39]. Moreover, second-order statistics of log-transformed data is similar to that of linear images, see [105]. Instead, the motivation for the average substraction is to assess the data independently on the illumination level, analogously to a von Kries procedure (see [120]).

Following [106], if $\tilde{L}$, $\tilde{M}$, $\tilde{S}$, are the basis vectors in the logarithmically-transformed space, then the application of the PCA gives the following three principal axes:

$$\begin{cases} l = \frac{1}{\sqrt{3}}(\tilde{L} + \tilde{M} + \tilde{S}) \\ \alpha = \frac{1}{\sqrt{6}}(\tilde{L} + \tilde{M} - 2\tilde{S}) \\ \beta = \frac{1}{\sqrt{2}}(\tilde{L} - \tilde{M}). \end{cases} \tag{6.6}$$

The color space spanned by these three principal axes is called *$l\alpha\beta$ space*. The standard deviations of the l, α, β coordinates are $\sigma_l = 0.353$, $\sigma_\alpha = 0.0732$ and $\sigma_\beta = 0.00745$. Notice that there is an inversion in the importance of opponent channels with respect to [18]: here the $L - M$ channel has the lowest variance.

To study spatiochromatic decorrelated features, Ruderman, Cronin and Chiao considered 3×3 patches, with each pixel containing a 3-vector color information, so that every patch is converted in a vector with 27 components that they analyzed with the PCA. The principal axes of these small patches in the logarithmic space are depicted in Figure 6.2. It can be seen that the first principal axis shows fluctuations in the achromatic channel, followed by blue-yellow fluctuations in the α direction and red-green ones in the β direction.

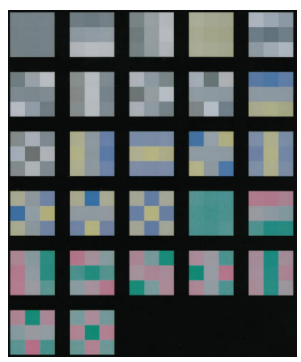


Figure 6.2: Principal axes of 3×3 patches arranged in order of decreasing eigenvalue, from left to right, top to bottom (adapted from [106], page 2041). The *color rendering* that the authors performed is the following: firstly, the R, G, B values are linearly related to the logarithmic $\tilde{L}, \tilde{M}, \tilde{S}$ values via these formulae $R = 128(\tilde{L} + 1)$, $G = 128(\tilde{M} + 1)$, $B = 128(\tilde{S} + 1)$. Notice that $\tilde{X} = 0$ if and only if $\mathrm{Log} X = \langle \mathrm{Log}\, X \rangle$, for $X = L, M, S$, so the previous relations set the average image values in the logarithmic domain to 128. Finally, the (R, G, B) values so obtained are linearly stretched to [0,255].

The spatial axes are largely symmetrical and can be represented by Fourier features, in line with the translation-invariance of natural images, as argued in [36]. It is important to stress that in Figure 6.2 no pixel within the patches appear other than the primary gray, blue-yellow or red-green colors, i.e. no mixing of l, α, β has been found in any 3×3 patch. These means that not only the single-pixel principal axes l, α, β, but also the spatially-dependent principal axes $l(x), \alpha(x), \beta(x)$, viewed as functions of the spatial coordinate x inside the patches, are decorrelated.

These results have been confirmed by [79] and, in Section 6.3, we will perform similar experiments on much larger databases.

6.2 Relationship between second order stationarity and the decorrelated spatiochromatic features of natural images

In this section we will analyze the consequence of second order stationarity in natural images on their decorrelated spatiochromatic features. For the sake of clarity, we will first start with the simplest case of gray-level images, where stationarity implies that the principal components are Fourier basis functions. We will then extend this result to the color case and show that a supplementary hypothesis on color covariance matrices yields principal components obtained as the tensor product between Fourier basis functions on the one hand, and achromatic plus opponent color coordinates on the other hand.

6.2.1 The gray-level case

Let I be a gray-level natural image of dimension $W \times H$, W being the width (number of columns) and H being the height (number of rows) of I.

If we denote the H rows of I as $r^0, \ldots, r^{H-1}$, then we can describe the position of each pixel of I row-wise as follows:

$$I = \{r_k^j;\ j = 0, \ldots, H-1,\ k = 0, \ldots, W-1\}, \tag{6.7}$$

j is the row index and k is the column index[1]. Each row $r^j = (r_0^j, \ldots, r_{W-1}^j)$ will be interpreted as a W-dimensional random vector and each component r_k^j as a random variable.

Let us define the *spatial covariance of the two random variables* r_k^j, $r_{k'}^{j'}$:

$$\mathrm{cov}(r_k^j, r_{k'}^{j'}) \equiv c_{k,k'}^{j,j'} = \mathbb{E}[r_k^j r_{k'}^{j'}] - \mathbb{E}[r_k^j]\mathbb{E}[r_{k'}^{j'}]. \tag{6.8}$$

Due to the symmetry of covariance we have $c_{k,k'}^{j,j'} = c_{k',k}^{j',j}$. Then, we can write the *spatial covariance matrix of the two random vectors* r^j, $r^{j'}$ as $\mathrm{cov}(r^j, r^{j'}) \equiv C^{j,j'}$, where $C^{j,j'}$ is the $W \times W$ matrix:

$$C^{j,j'} = \begin{bmatrix} c_{0,0}^{j,j'} & c_{0,1}^{j,j'} & \cdots & c_{0,W-1}^{j,j'} \\ c_{1,0}^{j,j'} & c_{1,1}^{j,j'} & \cdots & c_{1,W-1}^{j,j'} \\ \vdots & \vdots & \ddots & \vdots \\ c_{W-1,0}^{j,j'} & \cdots & \cdots & c_{W-1,W-1}^{j,j'} \end{bmatrix}. \tag{6.9}$$

Finally, the *spatial covariance matrix* C of the image I can be written as:

$$C = \begin{bmatrix} C^{0,0} & C^{0,1} & \cdots & C^{0,H-1} \\ C^{1,0} & C^{1,1} & \cdots & C^{1,H-1} \\ \vdots & \vdots & \ddots & \vdots \\ C^{H-1,0} & \cdots & \cdots & C^{H-1,H-1} \end{bmatrix}. \tag{6.10}$$

Notice that C is a $HW \times HW$ matrix because each sub-matrix $C^{j,j'}$ is a $W \times W$ matrix.

[1] To avoid cumbersome repetitions of the indexes variability, from now on, we will suppose that $j, j' \in \{0, \ldots H-1\}$ and $k, k' \in \{0, \ldots W-1\}$, unless otherwise specified.

Explicitly, we have that

$$C = \begin{bmatrix} \begin{bmatrix} c_{0,0}^{0,0} & \cdots & c_{0,W-1}^{0,0} \\ \vdots & \ddots & \vdots \\ c_{W-1,0}^{0,0} & \cdots & c_{W-1,W-1}^{0,0} \end{bmatrix} & \cdots & \begin{bmatrix} c_{0,0}^{0,H-1} & \cdots & c_{0,W-1}^{0,H-1} \\ \vdots & \ddots & \vdots \\ c_{W-1,0}^{0,H-1} & \cdots & c_{W-1,W-1}^{0,H-1} \end{bmatrix} \\ \vdots & \ddots & \vdots \\ \begin{bmatrix} c_{0,0}^{H-1,0} & \cdots & c_{0,W-1}^{H-1,0} \\ \vdots & \ddots & \vdots \\ c_{W-1,0}^{H,1} & \cdots & c_{W-1,W-1}^{H-1,0} \end{bmatrix} & \cdots & \begin{bmatrix} c_{0,0}^{H-1,H-1} & \cdots & c_{0,W-1}^{H-1,H-1} \\ \vdots & \ddots & \vdots \\ c_{W-1,0}^{H-1,H-1} & \cdots & c_{W-1,W-1}^{H-1,H-1} \end{bmatrix} \end{bmatrix} \tag{6.11}$$

Each sub-matrix $C^{j,j'}$ is a symmetric and positive definite real matrix.

Let us now introduce the hypothesis of spatial invariance in this context and and analyze its consequences.

6.2.2 Spatial invariance of the covariance

Hypothesis 1. From now on, the covariance of I is assumed to be invariant under translations of the row and column index: $c_{k,k'}^{j,j'} = c_{|k-k'|}^{|j-j'|}$.

Hypothesis 1 will be tested in Section 6.3. We notice that it is weaker than the typical definition of second order stationarity because here we do not assume the translation invariance of the mean.

Alongside this hypothesis, we add a technical requirement on the geometry of digital images which is implicitly assumed every time the Fourier transform is considered, i.e. we will consider a *symmetrized spatial domain with a toroidal distance*, which means that we will perform the identification $r_k^j = r_{k'}^{j'}$ when $j \equiv j' \pmod{H}$ and $k \equiv k' \pmod{W}$, i.e. every time there exist $a, b \in \mathbb{Z}$ such that $j' - j = aH$ and $k' - k = bW$.

As a covariance matrix, C is real, symmetric and positive-definite. Now, as a consequence of the previous hypotheses, the matrix C is also block-circulant with circulant blocks. Indeed, the $C^{j,j'}$ are *circulant matrices*, i.e. matrices where each row vector is rotated one element to the right relative to the preceding row vector[2]. If we use the convenient shorthand notation 'circ( )' to denote a circulant matrix, by specifying only the first row between the round brackets, then $C^{j,j'}$ can be written as follows:

$$C^{j,j'} = \operatorname{circ}\left(c_{0,0}^{j,j'}, c_{0,1}^{j,j'}, \ldots, c_{0,W-1}^{j,j'}\right). \tag{6.12}$$

Now, if we write $C^j \equiv C^{0,j}$, $j = 0, \ldots, H-1$ it is straightforward to see that the covariance matrix C is block-circulant and can be explicitly written as:

$$C = \operatorname{circ}\left(C^0, C^1, \ldots, C^{H-1}\right). \tag{6.13}$$

It is well known that an $n \times n$ circulant matrix has n eigenvalues corresponding to the DFT of the finite sequence given by the first row of the matrix itself, and its eigenvectors are the Fourier basis functions, see e.g. [42].

[2]This can be easily verified by noticing that $c_{k,k'}^{j,j'} = c_{k+1,k'+1}^{j,j'}$.

Let us apply this general result to the $W \times W$ circulant matrices C^j. The set of eigenvalue equations $C^j \mathbf{e}_m = \lambda^j_m \mathbf{e}_m$, $\lambda^j \in \mathbb{C}$ and $\mathbf{e} \in \mathbb{C}^W$, $m = 0, \ldots, W-1$, can be written as the following matrix equation $C^j E_W = \Lambda^j E_W$, where[3]:

$$\Lambda^j = \mathrm{diag}(\hat{c}^j_m;\ m = 0, \ldots, W-1), \quad \hat{c}^j_m = \sum_{k=0}^{W-1} c^j_k e^{-\frac{2\pi i m k}{W}}, \tag{6.14}$$

and

$$\begin{aligned} E_W &= [\mathbf{e}_0 | \mathbf{e}_1 | \cdots | \mathbf{e}_{W-1}] \\ &= \frac{1}{\sqrt{W}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & e^{-\frac{2\pi i}{W}} & \cdots & e^{-\frac{2\pi i (W-1)}{W}} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-\frac{2\pi i (W-1)}{W}} & \cdots & e^{-\frac{2\pi i (W-1)^2}{W}} \end{bmatrix}. \end{aligned} \tag{6.15}$$

The following remark will help us understanding how to extend the previous diagonalization procedure to the whole matrix C.

Remark 1. Let $M = \mathrm{circ}(M^0, \ldots, M^{H-1})$ be a block-circulant matrix and let us assume that the blocks M^j can be diagonalized on the same basis B. If we write $E_H = [\mathbf{e}_0 | \mathbf{e}_1 | \cdots | \mathbf{e}_{H-1}]$, with the vectors $\mathbf{e}_j$ defined as in eq. (6.15) for all $j = 0, \ldots, H-1$, then it can be verified by direct computation that $E_H \otimes B$ is a basis of eigenvectors of M, where $\otimes$ denotes the Kronecker product.

In the case of our spatial covariance matrix C, all the submatrices C^j have the same basis of eigenvectors E_W, thus the result stated in the previous remark can be directly applied on the matric C to guarantee that $E_H \otimes E_W = [\mathbf{e}_{m,l}]$, where

$$\mathbf{e}_{m,l} = \frac{\left(1, e^{-2\pi i \left(\frac{m}{W} + \frac{l}{H}\right)}, \ldots, e^{-2\pi i \left(\frac{m(W-1)}{W} + \frac{l(H-1)}{H}\right)}\right)^t}{\sqrt{HW}}, \tag{6.16}$$

for $m = 0, \ldots, W-1$, and $l = 0, \ldots, H-1$ provides a basis of eigenvectors for the matrix C.

Actually, due to the symmetry of covariance matrices, the complex parts of the exponentials involving the sinus function cancel out (see [42]) and so the 2D cosine Fourier basis also constitute a basis of eigenvectors of C.

6.2.3 The color case

Let $\mathbf{u} : \Omega \to [0, 255]^3$ be an RGB image function, where Ω is the spatial domain, and, for all $(j, k) \in \Omega$, $\mathbf{u}(j, k) = (R(j, k), G(j, k), B(j, k))$ is the vector whose components are the red, green and blue intensity values of the pixel defined by the coordinates (j, k).

We define the *spatiochromatic covariance matrix among two pixels of position* (j, k) *and* (j', k'), $c^{j,j'}_{k,k'}(R, G, B)$, by extending eq. (6.8) as follows:

$$\begin{bmatrix} C_{RR}(j, j', k, k') & C_{RG}(j, j', k, k') & C_{RB}(j, j', k, k') \\ C_{GR}(j, j', k, k') & C_{GG}(j, j', k, k') & C_{GB}(j, j', k, k') \\ C_{BR}(j, j', k, k') & C_{BG}(j, j', k, k') & C_{BB}(j, j', k, k') \end{bmatrix} \tag{6.17}$$

In the particular case defined by $j' = j$ and $k' = k$, we will call $c^{j,j'}_{k,k'}(R, G, B)$ '*chromatic autocovariance*' and denote it simply as $c^0(R, G, B)$. Notice that the matrix analyzed in [18] is the chromatic autocovariance of LMS values.

[3]We have used the simplified notation $c^j_m \equiv c^{0,j}_{0,m}$ to denote the matrix element of position m in the first row of the matrix C^j, $m = 0, \ldots, W-1$.

We then define the *spatiochromatic covariance matrix* $C^{j,j'}(R,G,B)$ *among the two random vectors* r^j, $r^{j'}$ given by the j-th and j'-the rows of the spatial support of $\mathbf{u}$ by extending eq. (6.9) as follows:

$$\begin{bmatrix} c^{j,j'}_{0,0}(R,G,B) & c^{j,j'}_{0,1}(R,G,B) & \cdots & c^{j,j'}_{0,W-1}(R,G,B) \\ c^{j,j'}_{1,0}(R,G,B) & c^{j,j'}_{1,1}(R,G,B) & \cdots & c^{j,j'}_{1,W-1}(R,G,B) \\ \vdots & \vdots & \ddots & \vdots \\ c^{j,j'}_{W-1,0}(R,G,B) & \cdots & \cdots & c^{j,j'}_{W-1,W-1}(R,G,B) \end{bmatrix}. \tag{6.18}$$

Finally, we define the *spatiochromatic covariance matrix* $C(R,G,B)$ *of the RGB image* u by extending eq. (6.10) to the $3HW \times 3HW$ matrix defined in this way:

$$\begin{bmatrix} C^{0,0}(R,G,B) & C^{0,1}(R,G,B) & \cdots & C^{0,H-1}(R,G,B) \\ C^{1,0}(R,G,B) & C^{1,1}(R,G,B) & \cdots & C^{1,H-1}(R,G,B) \\ \vdots & \vdots & \ddots & \vdots \\ C^{H-1,0}(R,G,B) & \cdots & \cdots & C^{H-1,H-1}(R,G,B) \end{bmatrix}. \tag{6.19}$$

Now, supposing that all the elements of the matrices (6.17) are positive, thanks to the Perron-Frobenius theorem we can assure that each of these $c^{j,j'}_{k,k'}(R,G,B)$ matrices has a basis of eigenvectors that can be written as a triad of achromatic plus opponent chromatic channels. If we further *assume that the matrices (6.17) can be diagonalized on the same basis of eigenvectors* (A,P,Q), then, thanks to Remark 1, we have that the eigenvectors of the spatiochromatic covariance matrix $C(R,G,B)$ can be written as the following Kronecker product:

$$(A,P,Q) \otimes \mathbf{e}_{m,l} \in \mathbb{R}^{3HW}, \tag{6.20}$$

which is precisely the type of eigenvectors that have been exhibited experimentally in [104]. A standard result of linear algebra guarantees that a set of matrices can be diagonalized on the same basis of eigenvectors if and only if they commute[4]. Thanks to the hypothesis of translation invariance of covariance, this is verified if and only if the generic covariance matrix $c^{j,j'}_{k,k'}(R,G,B)$ commutes with the chromatic autocovariance matrix $c^0(R,G,B)$.

It is convenient to resume all the hypotheses made and results obtained so far in the following proposition.

Theorem 6.2.1 *Let* $\boldsymbol{u} : \Omega \to [0,255]^3$ *be an RGB image function, with a periodized spatial domain* Ω*, and suppose that*

1. *The spatiochromatic covariance matrices matrices* $c^{j,j'}_{k,k'}(R,G,B)$ *defined in (6.17) depend only on the distances* $|j-j'|$, $|k-k'|$, *i.e. the covariance of* $\boldsymbol{u}$ *is stationary;*
2. *All matrices* $c^{j,j'}_{k,k'}(R,G,B)$ *are positive, i.e. their elements are strictly greater than 0;*
3. *The following commutation property holds:*

$$[c^0(R,G,B), c^{j,j'}_{k,k'}(R,G,B)] = 0, \quad \forall (j,k),(j',k') \in \Omega. \tag{6.21}$$

Then, the eigenvectors of the spatiochromatic covariance matrix $C(R,G,B)$ *defined in (6.19) can be written as the Kronecker product* $(A,P,Q) \otimes \boldsymbol{e}_{m,l}$*, where* (A,P,Q) *is the achromatic plus opponent color channels triad and* $\boldsymbol{e}_{m,l}$ *is the 2D cosine Fourier basis.*

proposition

Proposition 6.2.1 defines a mathematical framework where the empirical result shown in Figure 6.2 can be formalized and understood in terms of statistical properties of natural images. In the following section we will test this framework with the help of two large databases of RGB images.

[4]We recall that, given two generic matrices A and B for which the products AB and BA is well defined, $[A,B] \equiv AB - BA$ is called the 'commutator' between them. Of course A and B commute if and only if $[A,B] = 0$.

6.3 Validations on natural image databases

The validity of Proposition 6.2.1 is constrained by some hypotheses. In this section we present the tests that we have performed to check their validity.

To perform our numerical experiences we have selected two databases that are best suited for different scopes. The first one is an excerpt from the database described in [46], which consists of 2.3 million of 1024×768 copyright-free RGB images taken from the popular website Flickr. The images of the database have been randomly ordered to reduce as much as possible the scene content bias. The advantage of this first database is its large number of images, which enabled us to check that the stability of our results.

The second database is made of personal RAW photographs of 7000 natural scenes. Each 4-neighborhood of pixels in a raw image contains two pixels corresponding to the R and B channels and two pixels corresponding to the G channel. We demosaicked each RAW image to build a subsampled RGB image simply by keeping unaltered the R and B information and averaging the G channel. The advantage of this second database is that RAW images are free from post-processing operations such as gamma correction, white balance or compression, thus, modulo camera noise, they provide a much better approximation of irradiance than the images of the first database.

6.3.1 Computation of the covariance matrices

Since the test that we have performed are related to covariance matrices, it is logic to start describing how these matrices are obtained from the images of the databases. First of all, in order to simplify their computation, the expectation of the empirical average image of the ensemble has been subtracted to all images.

We first concentrated on the evaluation of the chromatic autocovariance $c^0(R,G,B)$. Even if the simplest way to compute the covariance is via the Fourier transform, which implicitly assumes periodicity and may lead to biases, we chose the alternative strategy that we are going to describe in detail hereafter. First of all, the expectation operator $\mathbb{E}$ involved in its computation has been approximated by randomly selecting a pixel in N different images of the database, storing its RGB values in three N-dimensional row vectors v_μ, $\mu \in \{R,G,B\}$, and then estimating the elements of the chromatic autocovariance matrices as follows: $C_{RR} = v_R {v_R}^t/N$, $C_{RG} = v_R {v_G}^t/N$, and so on. Notice that it is possible to compute the covariances in this way because of the initial subtraction of the average image.

We then turned our attention to the spatiochromatic covariance matrices $c^{j,j'}_{k,k'}(R,G,B)$ with $j \neq j'$ and $k \neq k'$. To simplify the notation, from now on we will write $c^{j,j'}_{k,k'}(R,G,B) \equiv c^d(R,G,B)$, where $d = \sqrt{(j-j')^2 + (k-k')^2}$. We compute $c^d(R,G,B)$ by randomly selecting a different point $(j,k) \in \Omega$ in each image of the packet and considering its four neighbors $(j-d,k)$, $(j+d,k)$, $(j,k-d)$, $(j,k+d)$, we then count how many of these neighbors actually fall in the image domain Ω and we create the vectors v_μ and w_μ, $\mu \in \{R,G,B\}$: in w_μ we store the R,G,B values of the neighbors of (j,k) that fall in Ω, while in v_μ we store the R,G,B values of the central pixel (j,k) repeated as many times as the length of v_μ. We iterate the procedure for all the images of the packet and we concatenate the values of the random pixels and their neighbors in the vectors v_μ and w_μ, respectively. The estimation of the matrix elements of $c^d(R,G,B)$ can be done as follows: $C^d_{RR} = v_R {w_R}^t/L$, $C^d_{RG} = v_R {w_G}^t/L$, and so on, where here L denotes the common length of v_μ and w_μ.

6.3.2 Stability of the covariance computation with respect to the number of images and the image content

As previously stated, the very large Flickr database allows us checking the stability of the covariance matrices computation. We will now introduce the details of the stability tests. If we fix a threshold $\varepsilon = 10^{-D}$, $D \in \mathbb{N}$, then we consider the estimation of $c^0(R,G,B)$ *ε-stable* when the relative error defined by $\|c^0_{N+1}(R,G,B) - c^0_N(R,G,B)\| / \|c^0_{N+1}(R,G,B)\|$ is smaller than ε, where $c^0_N(R,G,B)$ is

the estimation of $c^0(R,G,B)$ obtained with N images and $\| \ \|$ is the 2-norm. Due to the law of large numbers, we expect the relative error to decrease proportionally to $1/N$. This is confirmed by our experiments, as can be seen in Figure 6.3. A good trade-off between precision and computational time required to perform the experiments is given by $N = 10^5$, which guarantees a 10^{-4}-stable estimation of the covariance.

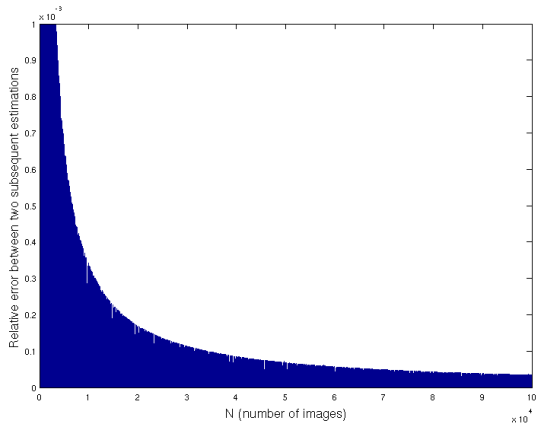


Figure 6.3: Relative error decay in the estimation of the chromatic autocovariance matrix elements.

With this value of N fixed, we tested the stability of the computation with respect to the image content by selecting 10 different packets of N pictures and comparing the estimation of $c^0(R,G,B)$. Our tests have reported differences in the estimation of $c^0(R,G,B)$ of magnitude 10^{-4}, which is the same order as the stability error, this confirms that the bias induced by the image content is negligible.

When we computed the spatial distributions of the chromatic autocovariances we have found an almost perfect homogeneity, with only a small vertical gradient: covariances are slightly larger in the upper part of the images and smaller in lower parts.

6.3.3 The chromatic autocovariance matrix $c^0(R,G,B)$ and its eigenvalues and eigenvectors

The expressions of the chromatic autocovariance matrices relative to the Flickr and RAW databases, ${c^0}_{\text{Flickr}}(R,G,B)$ and $c^0_{\text{RAW}}(R,G,B)$, respectively, that we have obtained are the following:

$$c^0_{\text{Flickr}}(R,G,B) = \begin{bmatrix} 0.0719 & 0.0651 & 0.0612 \\ 0.0651 & 0.0713 & 0.0710 \\ 0.0612 & 0.0710 & 0.0851 \end{bmatrix}, \tag{6.22}$$

$$c^0_{\text{RAW}}(R,G,B) = \begin{bmatrix} 0.0022 & 0.0021 & 0.0021 \\ 0.0021 & 0.0021 & 0.0022 \\ 0.0021 & 0.0022 & 0.0024 \end{bmatrix}, \tag{6.23}$$

which confirm the positivity assumption on $c^0(R,G,B)$. Notice that the covariances observed on the RAW database are much smaller than those observed on the Flickr database. We believe that this is mostly due to the fact that the contrast of images posted on Flickr is often much higher than the contrast of unprocessed RAW images. Despite this difference, the eigenvectors of the previous matrices are very similar:

$$\begin{cases} A_{\text{Flickr}} = (0.5483, 0.5761, 0.6061) \quad \longleftrightarrow \quad \lambda_1 = 0.2080, \\ P_{\text{Flickr}} = (0.7179, 0.0474, -0.6945) \quad \longleftrightarrow \quad \lambda_2 = 0.0170, \\ Q_{\text{Flickr}} = (0.4289, -0.8160, 0.3876) \quad \longleftrightarrow \quad \lambda_3 = 0.0034. \end{cases} \tag{6.24}$$

and

$$\begin{cases} A_{\text{RAW}} = (0.5679, 0.5683, 0.5954) \quad \longleftrightarrow \quad \lambda_1 = 0.0065, \\ P_{\text{RAW}} = (0.7210, 0.0055, -0.6930) \quad \longleftrightarrow \quad \lambda_2 = 0.0002, \\ Q_{\text{RAW}} = (0.3971, -0.8228, 0.4066) \quad \longleftrightarrow \quad \lambda_3 = 7.8 \cdot 10^{-7}. \end{cases} \tag{6.25}$$

We can see that, using a database of real RGB images and not the idealized visual stimuli of Buchsbaum and Gottschalk, the blue channel not only appears in the achromatic direction A, but it is even its dominant component. Observe the similarity between the eigenvectors obtained with the Flickr and RAW database and those reported in [73], which where also obtained from RGB camera images.

6.3.4 The exponential decay of spatiochromatic covariance matrix elements

All the spatiochromatic matrices $c^d(R, G, B)$ that we have estimated turned out to be *positive* and their decay with respect to increasing values of d is reported in the linear, bi-logarithmic and semi-logarithmic scale in Figure 6.4 for the Flickr database and in Figure 6.5 for the RAW database.

Let us write the generic element of the matrix $c^d(R, G, B)$ as $c^d_{\mu\nu}$, $\mu, \nu \in \{R, G, B\}$. Notice that a power-law behavior for $c^d_{\mu\nu}$ would be represented by a linear relationship in the bi-logarithmic scale, i.e. $\log(c^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu} \log(d)$, which, in fact, is equivalent to $c^d_{\mu\nu} = e^{\alpha_{\mu\nu}} d^{\beta_{\mu\nu}}$, $\beta_{\mu\nu} < 0$. However, as can be seen in Figures 6.4, 6.5, the graphs in the bi-logarithmic scale show a significant deviation from a linear behavior from $d = 100$, these distance being expressed in pixels. This confirm in the color case the fact that *the power-law decay of the covariance is not valid for large pixels distances*, a fact already noticed for gray level images in [48].

Moreover, notice that the graphs of Figures 6.4, 6.5 in the semi-logarithmic scale show a linear decay for all the distances that we have tested (from 1 to 300 pixels). To quantify this behavior we have performed a linear fit: the value of the coefficient of determination R^2 (which expresses the percentage of empirical data variance that is described by the linear approximation) is greater than 0.98 for all curves.

A linear behavior in the semilogarithmic domain corresponds to an exponential decay: $\log(c^d_{\mu\nu}) = \alpha_{\mu\nu} + \beta_{\mu\nu} d$ is equivalent to $c^d_{\mu\nu} = e^{\alpha_{\mu\nu}} e^{\beta_{\mu\nu}\, d}$, $\beta_{\mu\nu} < 0$. Since $c^0_{\mu\nu} = e^{\alpha_{\mu\nu}}$, we can write

$$c^d_{\mu\nu} = c^0_{\mu\nu} e^{\beta_{\mu\nu}\, d}, \ \mu, \nu \in \{R, G, B\}, \tag{6.26}$$

where $c^0_{\mu\nu}$ is the generic element of the chromatic autocovariance matrix.

The value of the coefficients $\beta_{\mu\nu}$ are listened in Table 6.1. It can be seen that the spatiochromatic

Flickr Database	RAW Database
$\beta_{RR} = -0.0028$	$\beta_{RR} = -0.0023$
$\beta_{GG} = -0.0026$	$\beta_{GG} = -0.0021$
$\beta_{BB} = -0.0022$	$\beta_{BB} = -0.0020$
$\beta_{RG} = \beta_{GR} = -0.0028$	$\beta_{RG} = \beta_{GR} = -0.0022$
$\beta_{RB} = \beta_{BR} = -0.0028$	$\beta_{RB} = \beta_{BR} = -0.0022$
$\beta_{GB} = \beta_{BG} = -0.0025$	$\beta_{GB} = \beta_{BG} = -0.0021$

Table 6.1: Slopes of the straight lines which approximate the spatiochromatic covariance graphs in the semilogarithmic scale for the Flickr and the RAW databases.

covariance relative to the blue channel decreases less rapidly than that of the red and green channels. This may be explained by the fact that pictures in which the sky is present are characterized by large homogeneous areas dominated by the blue channel.

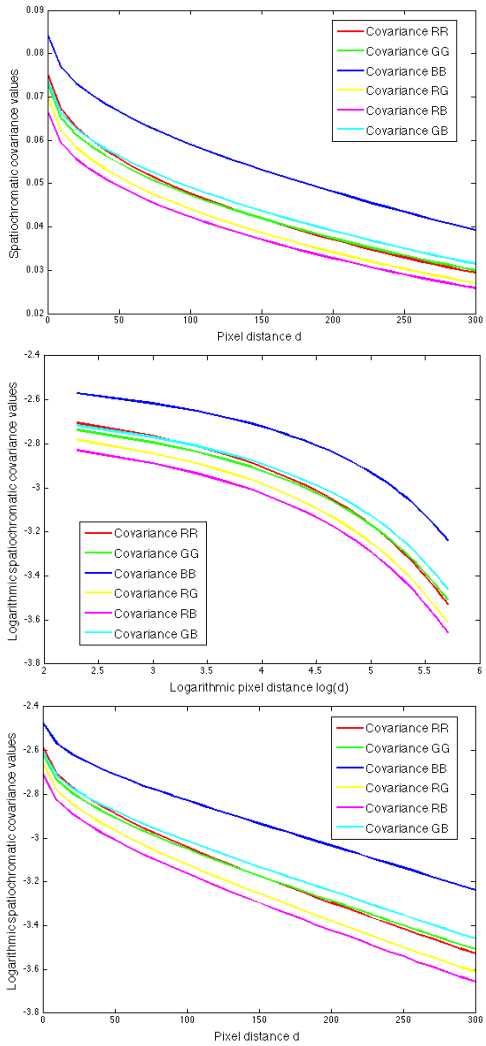


Figure 6.4: *Top to bottom.* Graphs of the six distinct spatiochromatic covariance matrix elements in the linear, (natural) bi-logarithmic and semi-logarithmic scale, respectively, as a function of the pixel distance d. The values were obtained using the Flickr database.

The explicit analytical expressions of $c^d(R, G, B)$ that we have managed to obtain are interesting for two reasons: firstly, they provide an accurate model for the covariance that corrects the power-law decay; secondly, they allow computing the commutators $[c^0(R, G, B), c^d(R, G, B)]$ for every distance $d > 0$. If the coefficients $\alpha_{\mu\nu}$ were all equal, then these commutators would be null matrices, however, the differences in the values of the exponentials make the matrix elements of the commutators different than zero. Figure 6.6 shows the absolute and normalized 2-norms of the commutators as a function of d. It can be seen that, for small values of d, the commutators can be considered approximately null, however, as d increases, they show deviations from the zero matrix, but they are still small.

A further and, in our personal opinion, more significant test that we have performed on RAW images is the following: we have transformed the initial data with the Michaelis-Menten eq. (2.1) to model the first stage of photoreceptors' output.

We then repeated the spatiochromatic covariance computations on this new transformed database and we have found again an exponential decay, but this time with exactly the same coefficients for all the chromatic combinations, i.e. $\beta_{\mu\nu} = -0.0033$, for all $\mu, \nu \in \{R, G, B\}$. This can be explained by

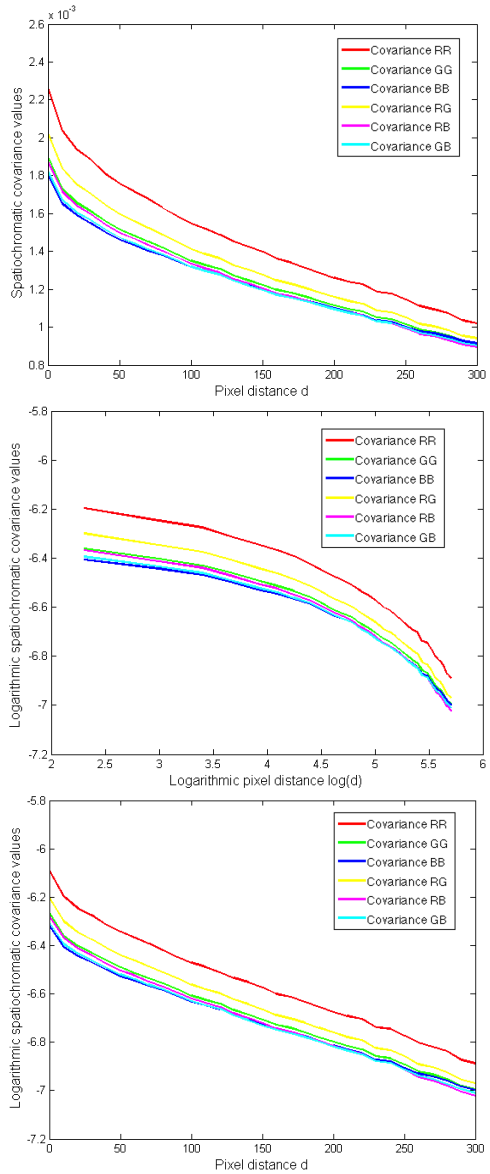


Figure 6.5: *Top to bottom.* Graphs of the six distinct spatiochromatic covariance matrix elements in the linear, (natural) bi-logarithmic and semi-logarithmic scale, respectively, as a function of the pixel distance d. The values were obtained using the RAW database.

the fact that the photochemical transduction process normalizes the dynamic range of the data to the interval $[0, 1]$ and sets the average value of each chromatic channel to $1/2$.

The by-product of photoreceptors' photochemical transduction is the rearranging of radiance values so that all the spatiochromatic covariance matrices commute perfectly. Up to our knowledge, this test has never been performed before.

We notice that the presence of the three average radiance values in the photoreceptors' photochemical transduction formula and the fact that they are normalized to the same value $1/2$ is crucial for the rearrangement of the slopes $\beta_{\mu\nu}$.

As already remarked, RAW data provide a good approximation of real physical irradiance, thus we believe that this test gives an interesting hint about the consequence of photochemical transduction

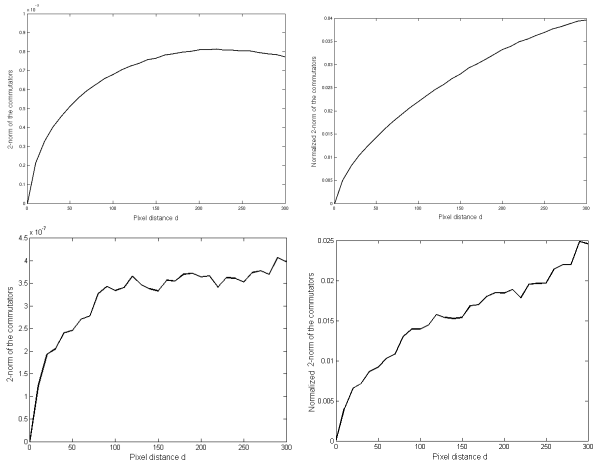


Figure 6.6: *Left*: Graph of the 2-norm of the commutators between the spatiochromatic covariance matrices as a function of pixel distance d. *Right*: Graph of the normalized 2-norm of the commutators, the normalization is done over the mean value of 2-norm of the product matrix performed from left to right and from right to left. *First row*: data obtained with the Flickr database. *Second row*: data obtained with the RAW database.

on covariance of natural visual stimuli. However, a better choice would be to perform this test on multispectral radiance values, in fact the irradiance acquired in the RAW data is integrated with respect to the RGB camera sensitivity functions, which are very different than the cone sensitivity curves. This is not yet possible due to the lack of a large database of natural multispectral images. I will come back on this point in chapter 8.

Chapter 7

Variational histogram transfer

This final chapter is devoted to a variational formulation of color grading, that is the color matching of two or more images. We will show that, by minimizing a suitable energy functional, images are transformed so that they achieve an intermediate common histogram.

Color grading is a main tool for color processing in the post-production industry in order to match colors of images which may have been taken at different times or with different illumination conditions, for flicker stabilization of video sequences. or for color balancing, a process that is mostly done manually.

On the other hand, numerical comparison of images in order to correlate them or extract mutual information is one of the main topics in computer vision, with applications in different contexts like depth or disparity computation in stereo images, or optical flow computation. Since images may be taken by different cameras or under different lighting conditions, their color normalization is a basic process previous to their comparison.

The main approach to color grading is based on the computation of a color transfer map. In the case of two color images, source and target, the problem is to find a mapping in the space of colors that transforms the source into an image with a color distribution as similar as possible to that of the target. We call this mapping the *transfer map*. In other cases, instead of a target image, one might want to specify a reference palette to be matched. The modified image should keep the geometry of the original one, only its color aspect should be modified. A similar problem can be formulated for more than two images.

In case of two color images, a transfer map can be computed to match simple statistical values, like mean and variance [97], the covariance matrix [1], or the whole color histogram [71, 72, 89, 90].

The matching of statistics can be done separately for each channel in the $l\alpha\beta$ color space, which has uncorrelated components, as in [97]. Another possibility is to compute an affine mapping [58, 87] to map the covariance matrices of two color images. The matching of color histograms is proposed in [45] and is done on the three channels independently. A color histogram equalization is performed in [85] by computing a transfer map using mesh deformations to fit the existing histogram to a uniform histogram.

The formulation of color histogram matching in terms of Monge-Kantorovich's mass transportation theory has been considered in [71, 24, 88]. In [71, 24] the authors compute a transport plan (also called transference plan, see [119]) which minimizes the total displacement cost between the two histograms that have to be matched. This cost is usually known under the names of Earth Mover Distance (EMD) or Wasserstein distance [119]. In the case of color images, this cost measures the total amount of color changes and the minimum cost is computed using linear programming [71, 24]. Recall that a transport plan $\pi = \{\pi_{\lambda,\mu} \in [0,\infty) : \lambda \in \{1,\dots,n\}, \mu \in \{1,\dots,m\}\}$ matching two histograms $\{h_0(\lambda)\}_{\lambda=1}^n$ and $\{h_1(\mu)\}_{\mu=1}^m$ is an assignment of the mass $h_0(\lambda)$ of bin λ to the bins of h_1 and viceversa. A transfer map T mapping h_0 to h_1, i.e., such that $h_0 \circ T^{-1} = h_1$, corresponds to the case in which $\pi_{\lambda,\mu} = \delta_{T(\lambda),\mu}$, where $\delta_{T(\lambda),\mu} = 1$ if $T(\lambda) = \mu$ and 0 otherwise.

This technique has been applied in [24] to *midway color histogram equalization* (see [25, 22] for the

case of gray level images), that is, to the problem of finding a midway histogram in the sense of EMD distance [119] and computing mappings that transform the original images into images that share this histogram. If we know the transfer map T mapping the histogram of image U_0 onto the histogram of U_1, then the intermediate histogram is associated to the transfer map $\lambda \mapsto \frac{\lambda+T(\lambda)}{2}$ where $\lambda \in \mathbb{R}$ for gray level images or in $\mathbb{R}^3$ for color images. As in [25], if the algorithm computes a transport plan π, one can define a transfer map $T(\lambda)$ as the barycenter of $(\pi_{\lambda,\mu})_\mu$.

As a practical method, assuming that the histograms of two given color images can be approximated by multivariate Gaussian distributions, in [88] Monge-Kantorovich's formulation is restricted to find a linear map that approximates the transfer of histograms. Although this method is applied in [88] to transfer the histogram of one image to another, it can also be used to define a midway histogram as in [24, 25]. As another efficient method, the transfer map in the color grading method proposed in [90] is computed by iteratively equalizing all principal axis marginals computed for every possible rotation of the histogram, until convergence. Then the method is refined by a filtering process, constrained by the original gradient field, to reduce possible grain artifacts [90].

Here we follow the idea of midway histogram equalization to map the color of two given images so that they share in as much as possible this histogram.

This permits to distribute the errors in matching histograms in the two images without giving privilege to one of them. Midway histogram equalization was studied by J. Delon in [24, 25] using Monge-Kantorovich mass transportation theory both for gray level (a case already considered in [22]) and color images.

But once the transfer map between the color histogram of two images is known, it can be used to map both images to share a common midway histogram.

Here we propose a new variational model to transform two or more color images in order to approach their color histograms as much as possible. To briefly describe the process, let us consider a set of color images $\{\vec{U}^k\}_{k=1}^m$: we want to transform them into new images $\{\vec{u}_k\}_{k=1}^m$ that approximately share an intermediate color histogram which is not specified in advance and, at the same time, we want to maintain the geometry of each $\vec{U}^k$. Thus, the energy has to combine the comparison of histograms with color and shape constraints (which can be codified by the gradient direction). The shape constraints can be related to the filtering process constrained by the original gradient field used in [90].

It would be interesting to address this problem from the point of view of Monge-Kantorovich mass transport theory [119]. However, when three or more images are involved the problem of computing its barycenter using the Wasserstein distance [119] becomes computationally intractable.

Up to our knowledge, there is no method in the literature that allows realizing such a process without defining the histogram beforehand.

7.1 A short review of midway histogram equalization

Midway histogram equalization for grey-scale and color images has been extensively studied by J. Delon [24, 25]. As she observes, if H_0, H_1 are the cumulative histograms of two original gray-scale images U_0 and U_1, respectively, then mapping them to have a cumulative histogram which is the direct average of the original ones, i.e. $\tilde{H} := (H_0 + H_1)/2$, is not always a suitable choice for midway equalization. An illuminating counterexample is given by the case when U_0 and U_1 are unimodal images with spread modes centered at the levels λ_0 and λ_1, respectively, with $\lambda_0 \neq \lambda_1$. Ideally, one would expect a midway image to present a single mode centered at $(\lambda_0 + \lambda_1)/2$, instead it is clear that the composition of U_0 with $\tilde{H}^{-1} \circ H_0$ spreads the gray levels of the image on two modes centered at λ_0 and λ_1 in order to make its histogram correspond as much as possible to $\tilde{H}$. Thus, it generates new structures that were not present before [24, 25]. Instead, the natural intermediate histogram $\bar{H}$ is given by

$$\bar{H} := \left(\frac{H_0^{-1} + H_1^{-1}}{2} \right)^{-1}. \tag{7.1}$$

Let $\psi_0 := \bar{H}^{-1} \circ H_0$ and $\psi_1 := \bar{H}^{-1} \circ H_1$ be the mappings of images U_0 and U_1 into images that share the midway histogram $\bar{H}$. Defining $T := H_1^{-1} \circ H_0$, we have

$$\psi_0 = \frac{I + H_1^{-1} \circ H_0}{2} = \frac{I+T}{2}, \qquad \psi_1 = \frac{I + H_0^{-1} \circ H_1}{2} = \frac{I+T^{-1}}{2} \tag{7.2}$$

and $\bar{H} = \left(\frac{I+T}{2}\right)^{-1} \circ H_0 = H_1 \circ \left(\frac{I+T^{-1}}{2}\right)^{-1}$, where I denotes the identity map. Observe that T maps U_0 into an image whose histogram is H_1. In [22] ψ_0, ψ_1 are computed using a dynamic programming algorithm.

Delon shows that these results can be embedded in a more abstract Monge-Kantorovich's framework which has the advantage to be easily extendible to any dimension, thus permitting to compute a midway equalization of color images. Let h_0, h_1 be the histograms of two generic images U_0, U_1, then h_0, h_1 can be seen as the Lebesgue densities of the probability measures $d\mu_0(x) = h_0(x)dx$, $d\mu_1(x) = h_1(x)dx$ describing the distribution of the image values. The search of an intermediate histogram between h_0 and h_1 is then equivalent to the search of an intermediate measure between μ_0 and μ_1. It can be proven that $\bar{H}$ can be seen as the cumulative distribution function of the middle point $\mu_{\frac{1}{2}}$ of the geodesic that connects μ_0 and μ_1 in the space of probability measures with respect to the quadratic transportation cost [119, 25].

As already observed, the discrete Monge-Kantorovich problem gives a transport plan that requires the solution of a linear programming problem [71, 24] to be computed. Let us mention that transport plans permit to split the mass of a bin of a given image histogram and distribute it between several bins of the other histogram. This may be considered a disadvantage. Moreover, the transport plan has the dimension of the product between the number of bins in both the original and the target image, this prevents from using fine quantization of the color space (although this may be unavoidable due to the insufficient quantization of color images). Finally, the extension to the case of more than two color images is not obvious. As we mentioned in the Introduction, in [88] the authors propose to restrict the computation of transfer maps to the case of affine maps. Even in this context, the extension of the method to the case of several images is not evident, although it is an interesting problem.

As stated in the Introduction, the variational method proposed in this paper approaches the color histogram of the given images without pre-computing it and can be applied to more than two images.

7.2 A variational approach to histogram transfer

To be coherent with the canonical notation in the state of the art about this topic, we will denote the images as $\vec{U} : \Omega \to [0,1]^N$, $\vec{U}(x) = (U_1(x), \dots, U_N(x))$. For color images, we have $N = 3$ and $U_k(x)$ is the intensity level of the pixel $x \in \Omega$ in the chromatic channel $k \in \{R, G, B\}$. $\Omega = \{1, \dots, W\} \times \{1, \dots, H\} \subset \mathbb{N}^2$, $W, H \geq 1$ being integers.

The histogram of the image $\vec{U}$ will be denoted by the density function $h_{\vec{U}} : [0,1]^N \mapsto [0,1]$, while $H_{\vec{U}} : [0,1]^N \mapsto [0,1]$ will represent the cumulative histogram of $\vec{U}$, coherently with the notations of chapter 3.

Thus, if $\vec{\lambda} = (\lambda_1, \dots, \lambda_N) \in [0,1]^N$ represents the generic N-dimensional intensity level, we have:

$$H_{\vec{U}}(\vec{\lambda}) = \int_0^{\lambda_1} \cdots \int_0^{\lambda_N} h_{\vec{U}}(\vec{\mu}) d\vec{\mu}. \tag{7.3}$$

Note that the cumulative histogram can be equivalently defined from the image $\vec{U}$ as:

$$H_{\vec{U}}(\vec{\lambda}) = \frac{1}{|\Omega|} |\{x \in \Omega, \text{ such that } U_1(x) \leq \lambda_1, \dots, U_N(x) \leq \lambda_N\}|, \tag{7.4}$$

where $|\quad|$ denotes the area operator.

Our aim is to propose a variational formulation for midway histogram equalization. Assume that we are given two color images $\vec{U} : \Omega_{\vec{U}} \to \mathbb{R}^3$ and $\vec{V} : \Omega_{\vec{V}} \to \mathbb{R}^3$ where $\Omega_{\vec{U}}, \Omega_{\vec{V}}$ denote the corresponding

spatial domains. We want to map the images $\vec{U}, \vec{V}$ onto new images $\vec{u}, \vec{v}$, respectively, so that $\vec{u}$ and $\vec{v}$ share, as much as possible, a common intermediate histogram between those of $\vec{U}$ and $\vec{V}$. Thus, the main ingredients of the energy functional that we propose are the distance between the two color histograms of $\vec{u}$ and $\vec{v}$ and the constraints that impose $\vec{u}$ and $\vec{v}$ to share the same geometric information as $\vec{U}$ and $\vec{V}$, respectively.

In order to define such an energy, we need to discuss first the variational formulation of histogram transfer based on the L^2 distance between two histograms. In order to make our presentation more comprehensible, we shall discuss first the case of gray level images (which have one-dimensional histograms) and compute the first variation of the corresponding energy. This will clarify the differences between this functional and the transport approach to midway histogram equalization and will also explain why we need to impose color and geometric constraints. After this, we consider the two-dimensional case, $N = 2$, presenting the energy functional and its first variation. The computation details and the formulae of the color case, $N = 3$, can be found in [78].

7.2.1 The L^2 distance between the cumulative histogram of a gray-scale image and a reference one and its first variation

Let $U : \Omega \to [0, 1]$ be a gray level image whose cumulative histogram is $H_U(\lambda) = \frac{1}{|\Omega|}|\{x \in \Omega : U(x) \leq \lambda\}|$, $\lambda \in [0, 1]$. Let $H_{\text{ref}}(\lambda)$ be a given reference cumulative histogram. In order to impose the cumulative histogram of U to be close to $H_{\text{ref}}(\lambda)$ we propose to minimize the following energy:

$$\mathcal{E}(U) = \frac{1}{2} \int_0^1 (H_U(\lambda) - H_{\text{ref}}(\lambda))^2 \, d\lambda. \tag{7.5}$$

Obviously, the minimum would be attained when $H_U(\lambda) = H_{\text{ref}}(\lambda)$. This energy term will be later used together with other (attachment to data and shape-preserving) terms in order to construct an energy for intermediate histogram equalization of a given set of images.

Let us compute the first variation of the energy (7.5). For that, observe that we may write

$$H_U(\lambda) = \frac{1}{|\Omega|} \sum_{x \in \Omega} \mathrm{h}_\lambda(U(x)), \tag{7.6}$$

where

$$\mathrm{h}_\lambda(r) = \chi_{[0,\lambda]}(r) \qquad \lambda, r \geq 0, \tag{7.7}$$

$\chi_{[0,\lambda]}$ being the characteristic function of the set $[0, \lambda]$. Hence, if $J : \Omega \to \mathbb{R}^+$ denotes a perturbation, we have

$$\begin{aligned} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} H_{U+\epsilon J}(\lambda) &= \frac{1}{|\Omega|} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \sum_{x \in \Omega} \mathrm{h}_\lambda(U(x) + \epsilon J(x)) \\ &= \frac{1}{|\Omega|} \sum_{x \in \Omega} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathrm{h}_\lambda(U(x) + \epsilon J(x)) \\ &= -\frac{1}{|\Omega|} \sum_{x \in \Omega} \delta(U(x) - \lambda) J(x) \,. \end{aligned} \tag{7.8}$$

Then

$$\begin{aligned} \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} \mathcal{E}(U + \epsilon J) &= -\frac{1}{|\Omega|} \int_0^1 (H_U(\lambda) - H_{\text{ref}}(\lambda)) \sum_{x \in \Omega} \delta(U(x) - \lambda) J(x) \, d\lambda \\ &= -\frac{1}{|\Omega|} \sum_{x \in \Omega} \int_0^1 (H_U(\lambda) - H_{\text{ref}}(\lambda)) \delta(U(x) - \lambda) J(x) \, d\lambda \\ &= -\frac{1}{|\Omega|} \sum_{x \in \Omega} (H_U(U(x)) - H_{\text{ref}}(U(x))) J(x) \,. \end{aligned} \tag{7.9}$$

That is

$$\nabla_U \mathcal{E}(U) = -\frac{1}{|\Omega|}(H_U(U) - H_{\text{ref}}(U)). \tag{7.10}$$

We remark that we may use a gradient descent algorithm to minimize the energy (7.5). Initializing with $U^0 = U$, for some positive evolution parameter step τ, the process is described by the equation

$$U^{k+1} = U^k - \tau \nabla_{U^k} \mathcal{E}(U^k). \tag{7.11}$$

If H_1, H_2 are two given reference cumulative histograms, the energy

$$\mathcal{E}_{1,2}(U) := \frac{1}{2}\int_0^1 (H_U(\lambda) - H_1(\lambda))^2\, d\lambda + \frac{1}{2}\int_0^1 (H_U(\lambda) - H_2(\lambda))^2\, d\lambda \tag{7.12}$$

expresses the distance of H_U to H_1 and H_2. Using (7.10), the Euler-Lagrange equation $\nabla_U \mathcal{E}_{1,2}(U) = 0$ can be written as

$$H_U(U(x)) = \frac{H_1(U(x)) + H_2(U(x))}{2} \qquad \forall x \in \Omega. \tag{7.13}$$

Notice that this equation has to be satisfied by the minimum U of (7.12) and it is not the definition of $H_U(U(x))$ in terms of the right hand side of (7.13). On the other hand, it does not determine U uniquely. Moreover, we remark that this energy is not a suitable choice for intermediate histogram equalization since it may generate new unwanted structures (level lines) in the image (see [24, 25] and our discussion in Section 7.1). Thus, we add a constraint in order to keep the geometry of the original image. Before going into details about this constraining, let us extend the above computations to the vector case.

7.2.2 The L^2 distance between the cumulative histogram of a vector-valued image and a reference one and its first variation

Let us consider the case of vector valued images $\vec{U} : \Omega \to [0,1]^N$. To simplify the presentation, let us study here the case $N = 2$, the case of color images with $N = 3$ is described in [78]. Thus, if $\vec{U} = (U_1, U_2)$ and $\vec{\lambda} = (\lambda_1, \lambda_2) \in [0,1]^2$, the cumulative histogram of U is defined by $H_{\vec{U}}(\vec{\lambda}) = \frac{1}{|\Omega|}|\{x : U_1(x) \le \lambda_1,\, U_2(x) \le \lambda_2\}|$. Let $H_{\text{ref}}(\vec{\lambda})$ be a reference cumulative histogram. We then define the transfer energy:

$$\mathcal{E}(\vec{U}) = \frac{1}{2}\int_0^1\int_0^1 (H_{\vec{U}}(\vec{\lambda}) - H_{\text{ref}}(\vec{\lambda}))^2\, d\vec{\lambda}. \tag{7.14}$$

If $\mathcal{H}^1$, $\mathcal{H}^2$ are defined by

$$\mathcal{H}^1(i_1, i_2) \equiv \int_0^{i_2} H(i_1, \lambda_2) d\lambda_2, \quad \mathcal{H}^2(i_1, i_2) \equiv \int_0^{i_1} H(\lambda_1, i_2) d\lambda_1, \tag{7.15}$$

then, as proven in [78], the first variations of $\mathcal{E}(\vec{U})$ with respect to U_1, U_2 are

$$\begin{aligned}\nabla_{U_1}\mathcal{E}(\vec{U}) = &-\frac{1}{|\Omega|}\left[\mathcal{H}^1_{\vec{U}}(U_1(x), 1) - \mathcal{H}^1_{\text{ref}}(U_1(x), 1)\right] \\ &+\frac{1}{|\Omega|}\left[\mathcal{H}^1_{\vec{U}}(U_1(x), U_2(x)) - \mathcal{H}^1_{\text{ref}}(U_1(x), U_2(x))\right],\end{aligned} \tag{7.16}$$

$$\begin{aligned}\nabla_{U_2}\mathcal{E}(\vec{U}) = &-\frac{1}{|\Omega|}\left[\mathcal{H}^2_{\vec{U}}(1, U_2(x)) - \mathcal{H}^2_{\text{ref}}(1, U_2(x))\right] \\ &+\frac{1}{|\Omega|}\left[\mathcal{H}^2_{\vec{U}}(U_1(x), U_2(x)) - \mathcal{H}^2_{\text{ref}}(U_1(x), U_2(x))\right].\end{aligned} \tag{7.17}$$

7.2.3 Symmetrization of the cumulation of histogram

Let us mention that, by definition of cumulative histogram, the process described above has an important limitation. Let us explain it in the most simple case when $N = 1$. Indeed, note that for any grey image $U : \Omega \mapsto [0,1]$, we have $H_U(1) = \frac{1}{|\Omega|}|\{x : U(x) \leq 1\}| = 1$. As a consequence, when computing the gradient value of white pixels of color value 1, we will always find $H_U(1) - H_{\text{ref}}(1) = 1 - 1 = 0$, so white pixels never change their intensity value. To remove this limitation, we need to consider H^*, the histogram defined as $H^*_U(\lambda) = |\{x : U(x) \geq \lambda\}|$, $\lambda \in [0,1]$. The energy (7.5) should then be completed with the term $\frac{1}{2}\int_0^1 (H^*_U(\lambda) - H^*_{\text{ref}}(\lambda))^2\, d\lambda$, and the symmetrized energy will be

$$\mathcal{E}(U) = \frac{1}{2}\int_0^1 \left[(H_U(\lambda) - H_{\text{ref}}(\lambda))^2 + (H^*_U(\lambda) - H^*_{\text{ref}}(\lambda))^2\right] d\lambda, \tag{7.18}$$

allowing white pixels to move their value. Obviously, by symmetry, this new term will not affect the black pixels, as $H^*_U(0) = 1$ for any image $I : \Omega \to [0,1]$. For the N-dimensional case, $N > 1$, we have

$$\mathcal{E}(\vec{U}) = \frac{1}{2}\underbrace{\int_0^1 \cdots \int_0^1}_{N \text{ times}} \left[(H_{\vec{U}}(\vec{\lambda}) - H_{\text{ref}}(\vec{\lambda}))^2 + (H^*_{\vec{U}}(\vec{\lambda}) - H^*_{\text{ref}}(\vec{\lambda}))^2\right] d\vec{\lambda}. \tag{7.19}$$

It is clear that the gradient energy formulae previously written still hold for the symmetrized energy, modulo the introduction of the symmetric terms depending on H^*.

7.3 Intermediate histogram equalization of a set of images

As reviewed in section 7.1, histogram equalization is mainly addressed from the point of view of mass transfer [25, 90, 88], where the unknown is the transfer function $\varphi : [0,1]^N \mapsto [0,1]^N$ mapping a variable histogram to a reference one. Considering the problem of equalizing (at least) two different images with respect to their midway histogram, mass transfer methods require an *a priori* computation of the midway histogram. The purpose of this section is to show how can we make different images evolve simultaneously so that their histograms get naturally as close as possible to an unspecified intermediate histogram.

7.3.1 An energy term to approach the histogram of two images

Suppose that we are given two color images $\vec{U} : \Omega_{\vec{U}} \to [0,1]^3$ and $\vec{V} : \Omega_{\vec{V}} \to [0,1]^3$, $\vec{U} = (U_1, U_2, U_3)$, $\vec{V} = (V_1, V_2, V_3)$, and we want to map them into two images with a common histogram as much as possible. We can then try to construct two color images $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$ related to $\vec{U}$, $\vec{V}$, respectively, and defined on the spatial domains $\Omega_{\vec{U}}$ and $\Omega_{\vec{V}}$, so that they minimize the histogram-based energy $\mathcal{E}_H$:

$$\mathcal{E}_H(\vec{u}, \vec{v}) = \frac{1}{2}\int_0^1\int_0^1\int_0^1 (H_{\vec{u}}(\vec{\lambda}) - H_{\vec{v}}(\vec{\lambda}))^2\, d\vec{\lambda} + \frac{1}{2}\int_0^1\int_0^1\int_0^1 (H^*_{\vec{u}}(\vec{\lambda}) - H^*_{\vec{v}}(\vec{\lambda}))^2\, d\vec{\lambda}. \tag{7.20}$$

That is, $\vec{u}, \vec{v}$ should have similar histograms and be relatively close to the original images $\vec{U}$, $\vec{V}$.

Note that, even if the final histograms of $\vec{u}$ and $\vec{v}$ are equal, we cannot guarantee that this is the true midway histogram of $\vec{U}$ and $\vec{V}$ (in the sense described in Section 7.1). If we consider for instance two unimodal images $\vec{U} = \lambda_0$ and $\vec{V} = \lambda_1$, all pixels of each image are initially equal and they will always share a common value all along the minimization process (since in this case, the gradient (7.10) is the same for the whole image). Then, we could obtain any unimodal image $\vec{u} = \vec{v} = s$, $s \in [0,1]$, as a minimum of energy (7.20), while we would prefer finding the particular value $s \equiv \frac{\lambda_0+\lambda_1}{2}$. This value can be obtained if we add a data attachment term to enforce $\vec{u}, \vec{v}$ to be close to $\vec{U}$, $\vec{V}$, respectively. Thus, we propose to complement the energy term (7.20) with data attachment terms constraining the color and geometric information of $\vec{u}, \vec{v}$ to be close to those of $\vec{U}$, $\vec{V}$.

7.3.2 Attachment to data and to the image geometry

Contrary to the methods based on histogram transfer, as our variables are the images, we can directly add some color and geometrical restrictions on the image evolution.

Constraining the colors In order to enforce $\vec{u}$ and $\vec{v}$ to remain as close as possible to the original images $\vec{U}$ and $\vec{V}$ we define the following *attachment to data energy term* $\mathcal{E}_D$:

$$\mathcal{E}_D(\vec{u},\vec{v}) = \frac{\alpha_{\vec{U}}}{2}\sum_{x\in\Omega_{\vec{U}}} \|\vec{U}(x)-\vec{u}(x)\|^2 + \frac{\alpha_{\vec{V}}}{2}\sum_{x\in\Omega_{\vec{V}}} \|\vec{V}(x)-\vec{v}(x)\|^2\,, \tag{7.21}$$

where $\alpha_{\vec{U}}$ and $\alpha_{\vec{V}}$ are positive real parameters and $\|\cdot\|$ denotes the Euclidian norm (a notation that will be used throughout the paper). With this new term, by setting $\alpha_{\vec{U}} = \alpha_{\vec{V}}$, we can constrain the final images $\vec{u}$ and $\vec{v}$ to have a similar distance to $\vec{U}$ and $\vec{V}$.

Let us go back to the unimodal toy example $\vec{U} = \lambda_0$ and $\vec{V} = \lambda_1$ and consider the energy $\mathcal{E}_{HD} = \mathcal{E}_H + \mathcal{E}_D$, with $\alpha_{\vec{U}} = \alpha_{\vec{V}} = \epsilon > 0$. In this case both solutions $\vec{u},\vec{v}$ are unimodal, say $\vec{u} = \lambda_0, \vec{v} = \lambda_1$ with $\lambda_0 < \lambda_1$. Then we can write the energy $\mathcal{E}_{HD}$ in terms of λ_0, λ_1 and check that when $\epsilon > 0$ is small enough (which amounts to say that the histogram term $\mathcal{E}_H$ is dominant in $\mathcal{E}_{HD}$) the minimum of $\mathcal{E}_{HD}$ is reached for the unimodal image $\vec{u} = \vec{v} = \frac{\lambda_0+\lambda_1}{2}$ whose histogram corresponds to the midway one[1].

In real applications, the energy $\mathcal{E}_H$ is usually not zero, and we do not reach the true midway histogram. Nevertheless, from our experiments, we can assert that the histogram and data energies (7.20) and (7.21) allow the histograms of $\vec{u}$ and $\vec{v}$ to approach naturally a histogram in a neighborhood of the midway histogram of $\vec{U}$ and $\vec{V}$.

Constraining the image geometry We can constrain the shapes of $\vec{u}$ and $\vec{v}$ to be those of the given images $\vec{U}$ and $\vec{V}$. To describe the shapes of U_i, V_i we use the normals to its level lines, that is, we use the vector fields $\theta(U_i) = \frac{\nabla U_i}{\|\nabla U_i\|}$, $\theta(V_i) = \frac{\nabla V_i}{\|\nabla V_i\|}$ [10] (the quotient must be set to 0 when the denominator is 0) where ∇ denotes here a discrete gradient operator. Then we say that the scalar image u_i shares the same shapes as U_i (in some sense, the same level lines) if $\theta(U_i)\cdot\theta(u_i) = 1$, i.e. when $\theta(U_i)\cdot\nabla u_i = \|\nabla u_i\|$ (see [10] for more details).

In order to constrain the shapes of $\vec{u}$ and $\vec{v}$ to be those of $\vec{U}$ and $\vec{V}$, respectively, we consider (together with the histogram and color terms) the following shape preserving energy $\mathcal{E}_S$

$$\mathcal{E}_S(\vec{u},\vec{v}) = \beta_{\vec{U}}\sum_{i=1}^{3}\sum_{\Omega_{\vec{U}}}\left(\|\nabla u_i\| - \theta(U)_i\cdot\nabla u_i\right) + \beta_{\vec{V}}\sum_{i=1}^{3}\sum_{\Omega_{\vec{V}}}\left(\|\nabla v_i\| - \theta(V)_i\cdot\nabla v_i\right), \tag{7.22}$$

where $\beta_{\vec{U}}$ and $\beta_{\vec{V}}$ are positive parameters. Imposing the channels of two color images $\vec{U}$ and $\vec{u}$ to share the same shapes (or level lines) is not a strong constraint on the color distributions.

Notice that (7.22) incorporates the constraint $\theta(U_i)\cdot\nabla u_i = \|\nabla u_i\|$ instead of $\theta(U_i)\cdot\theta(u_i) = 1$. Formally they are equivalent, but the second formulation would lead to a very nonlinear problem much more difficult to handle numerically. On the other hand, we avoid division by zero when the modulus of the gradient of u_i vanishes.

7.3.3 A model for intermediate histogram equalization of a set of images

Combining the histogram energy (7.20) with the color and geometry preserving terms of Section 7.3.2, we finally get the following energy functional for equalizing the color of two images $\vec{U}$ and $\vec{V}$:

$$\mathcal{E}_{\vec{U},\vec{V}}(\vec{u},\vec{v}) = \mathcal{E}_H(\vec{u},\vec{v}) + \mathcal{E}_D(\vec{u},\vec{v}) + \mathcal{E}_S(\vec{u},\vec{v}). \tag{7.23}$$

[1] This would not be the case by considering a L^1 attachment to data term $\mathcal{E}_D(\vec{u},\vec{v}) = \alpha_U||U-u||_1 + \alpha_V||V-v||_1$. Indeed, the energy $\mathcal{E}_{H,D}$ would have the same value for any unimodal image $\vec{u} = \vec{v} = s$, $s \in [\lambda_0, \lambda_1]$.

The energy can be extended to deal with the equalization of more than two images. Let us consider $\vec{U}^k$, $k = 1, \ldots, m$, a set of color images. We can then use the energy (7.23) and define the following problem

$$\min_{\vec{u}_1,\ldots,\vec{u}_m} \sum_{1\leq k<l\leq m} \mathcal{E}_{\vec{U}^k,\vec{U}^l}(\vec{u}_k, \vec{u}_l), \tag{7.24}$$

in order to normalize the color of the set of images $\vec{U}^k$.

The minimization of (7.23) is performed via a primal-dual algorithm, as detailed in [78].

7.4 Experiments

In all our experiments, we used the same set of parameters. Histograms are discretized into $M = 64$ bins in each dimension. Concerning the model parameters, we used $\alpha_{\vec{U}} = \alpha_{\vec{V}} = 0.1$ for the data term and $\beta_{\vec{U}} = \beta_{\vec{V}} = 0.1$ for the shape preserving term. The time steps for the primal-dual proximal point algorithm are fixed as $\tau = \tau_\sigma = 10^{-4}$. This small value of the time step parameter is due to the form of the gradient of the histogram-based energy term $\mathcal{E}_H$ in dimension 3. This gradient contains the sum of several cumulative histogram functions $\mathcal{H}^i$, $i = 1, 2, 3$. In the worst case, i.e. in case of a black image, the maximum value of cumulative histograms $\mathcal{H}^i$ can be of the order of $M^6/8$, while we are dealing with colors in the range of $[0, 1]$. Such a big value is never obtained in practice, but it shows that a very small value of the time step parameters is necessary to stabilize numerically the primal-dual proximal point algorithm.

The complexity of the present algorithm is $n_{it} \times O(|\Omega| + M^3)$ where n_{it} is the number of iterations, $|\Omega|$ denotes the number of pixels of the image domain M is the number of bins in each dimension, in our case $M = 64$. This value of M has been chosen as a tradeoff between the complexity of the algorithm and the quality of the results. The number of iterations is at most 100 but the algorithm stops automatically when the value of (7.25) between one iteration and the following is below a given threshold that we have fixed in 0.0005. As an example, on a standard PC (P4, 3GHz), it took 12,51 seconds in order to transform the first two images in the third row of Fig. 7.2, whose dimension is 765×560, to the corresponding midway histogram images.

7.4.1 Histogram transfer

We first apply our process with $\vec{v} = \vec{V}$ being fixed, so that only the image $\vec{u}$ can change its values. In this case, we recover the same problem as the one considered in [88]. As the code of [88] is available on the web[2], we have compared this method with ours in terms of the distance between histograms. If $h_{\vec{u}}$ and $h_{\vec{v}}$ are the histograms of images $\vec{u}$ and $\vec{v}$, respectively, both methods are compared by means of the quantity

$$\int_0^1 \int_0^1 \int_0^1 (h_{\vec{u}}(\vec{\lambda}) - h_{\vec{v}}(\vec{\lambda}))^2 d\lambda, \tag{7.25}$$

computed for the images $\vec{u}$ obtained using both our method and the one in [88]. We have discretized the integrals in (7.25) using the same number bins used to discretize the histograms. To be more robust to histogram quantization effects we computed (7.25) after a slight smoothing of $h_{\vec{u}}$ and $h_{\vec{v}}$.

To illustrate the performance of our algorithm, let us display some experiments which can be judged by visual comparison. Since we are mapping images to an intermediate histogram, it is convenient to use the same image with different illuminations. For this reason we used a data set composed of pairs of images showing the same painting which are taken by unknown cameras and with different illuminations. We considered paintings, as they involve strong color contrast, and are quite close to the examples of color grading shown in [88]. This data set is thus suitable for the application of the method presented in [88]. The original and final images obtained by both methods, the one in [88] and

[2] http://www.mee.tcd.ie/∼sigmedia/Research/ColourGrading

ours, on the paintings are shown in Figs. 7.1 and 7.2. As illustrated in table 7.1, both approaches give errors with the same magnitude in the sense of (7.25), the method of [88] being slightly numerically better in general. However, visually speaking, the two methods are practically indistinguishable. In Fig. 7.3, we also draw the cumulative histograms of each channel for the Dali images of Fig. 7.2 in order to show that the functional does its job and indeed minimizes the distance between cumulative histograms.

In table 7.1, we have also added the final errors provided by a simple affine map that transforms the histograms in order to minimize their mean square difference. As it can be seen, this transformation produces errors which are, in general, almost one order of magnitude bigger than ours or those of [88].

In Fig. 7.4, we show the result of our algorithm applied on two quite different images. In this case the visual judgement is less intuitive to perform, however we believe that the result is in line with what one would expect from a midway histogram equalization. In fact, the dominant hue is blue in the Dali painting and brown in the 'Horloge' image, but after the midway equalization some parts of the Dali painting turn brown, e.g the ground, and some areas of the Horloge image turn bluish, e.g. the sky behind the building structure. The numerical comparison of the error (7.25) between the original and the final images confirms our claim

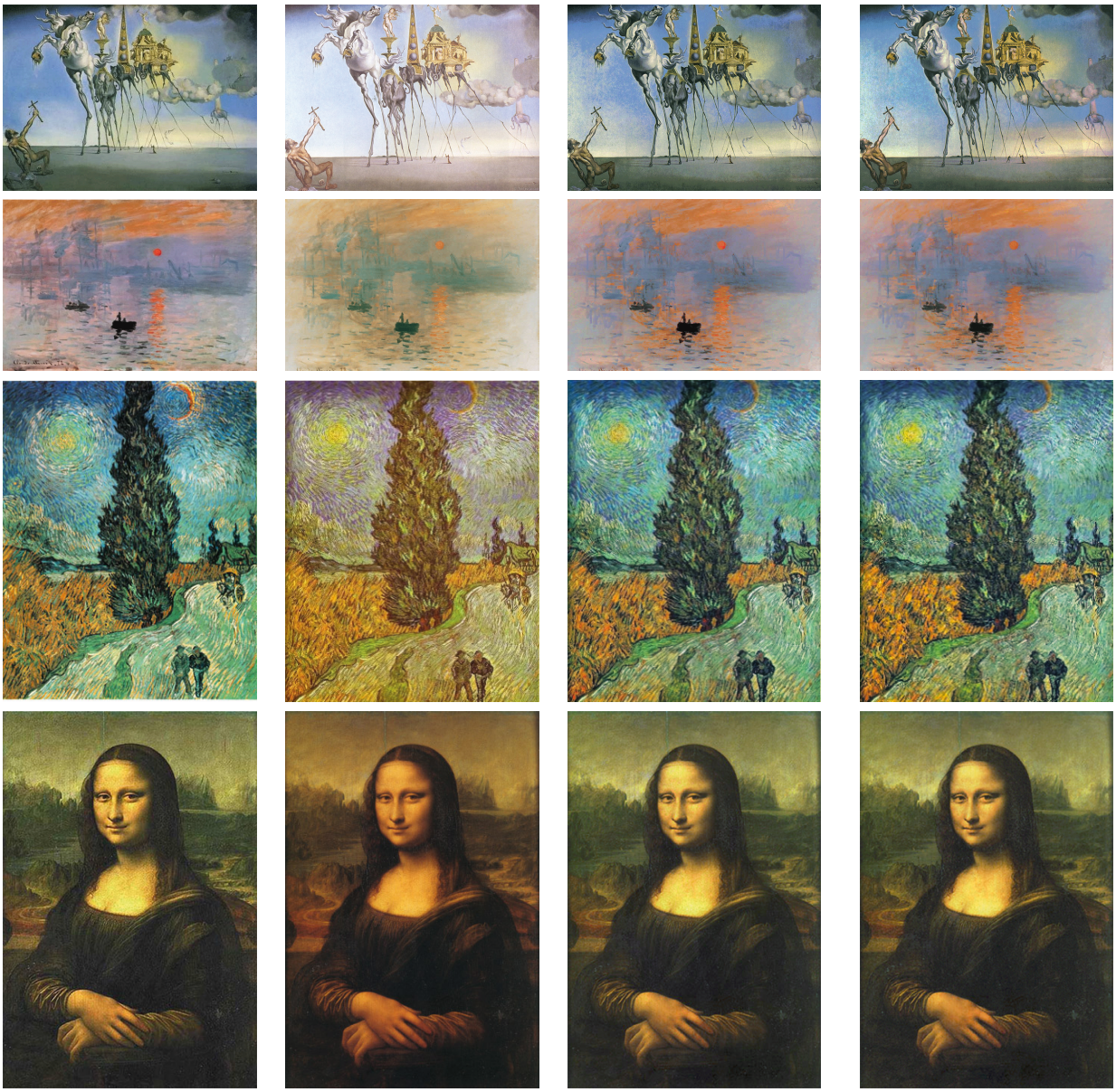
Reference image $\vec{V}$

Original image $\vec{U}$

Result $\vec{u}$ with [88]

Result $\vec{u}$ with our method

Figure 7.1: Comparison between our method and the one of [88] for the painting of Géricault: The Raft of the Medusa.



Reference image $\vec{V}$ | Original image $\vec{U}$ | Result $\vec{u}$ with [88] | Result $\vec{u}$ with our method

Figure 7.2: Comparison with the method of [88] for our data set. First row: Dali, Tentation. Second row: Monet, Impression Soleil Levant. Third row: Van Gogh: Cypress against a starry sky. Fourth row: Da Vinci, Gioconda.

7.4.2 Intermediate histogram equalization of a set of images

Let us consider $\vec{U}^k$, $k = 1, \ldots, m$, a set of color images representing the same scene with different orientations or illuminations. We solve the problem (7.24) in order to normalize the color of the set of images $\vec{U}^k$. As already commented, we underline that, to our knowledge, there is no method in the literature that allows realizing such a process without defining a midway histogram beforehand.

In [71, 24], when $m = 2$, the problem has been considered from the point of view of Monge-

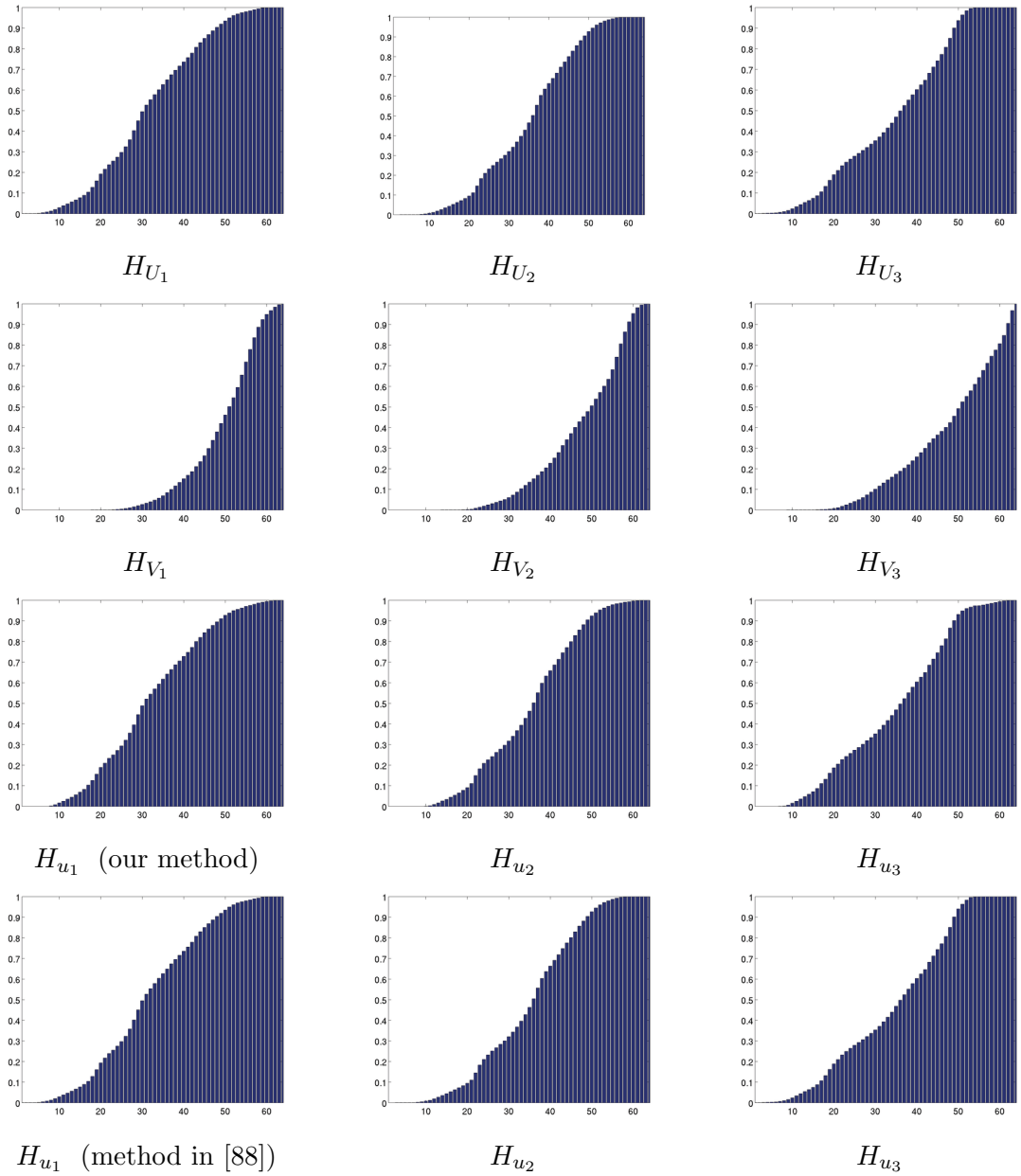


Figure 7.3: Cumulative histograms corresponding to the experiment on the Meduse image shown in Fig. 7.1. First line: Reference Red, Green and Blue cumulative histograms. Second line: Original Red, Green and Blue cumulative histograms. Third line : Final Red, Green and Blue cumulative histograms with our method. Fourth line: Final Green Red, Green and Blue cumulative histograms with the method of [88].

Kantorovich mass transportation theory leading to a linear programming problem for the computation of the transport plan transferring both histograms. Transport plans permit to split the mass of a bin of one of the histograms and distribute it between several bins in the other one. In [88], assuming that the histograms of two given color images can be approximated by multivariate Gaussian distributions, the Monge-Kantorovich formulation is restricted to find a linear map that approximates the transfer of

Data	Initial error	Final error [88]	Final error [Our method]	Final error [Affine map]
Medusa	0.001576	0.000149	0.000215	0.001375
Joconde	0.010981	0.001281	0.002050	0.009073
Tentation	0.001766	0.000223	0.000415	0.001337
Rising Sun	0.001134	0.000188	0.000115	0.001124
Cypress	0.000163	0.000025	0.000077	0.000134

Table 7.1: Quantitative comparison between the method of [88] (third column), our variational method (fourth column) and the simple affine transformation discussed in the text (fifth column). The errors are all computed with the formula (7.25).

Figure 7.4: First column: original images (the error (7.25) between them is: 0.006472). Second column: results of our algorithm (the error (7.25) between them is: 0.002040).

histograms. The color grading method proposed in [90] is based on the computation of a transfer map and could be adapted to compute a midway histogram equalization (see [25]) in the case of two images. The transfer map is computed by iteratively equalizing all principal axis marginals computed for every possible rotation of the histogram, until convergence. Then the method is refined by a filtering process to reduce possible grain artifacts [90]. As mentioned in Section 7.3.2, this process is related to our geometry attachment term. Thus our method, although different in spirit, has common points with [90]. On the other hand, our method can easily deal with more than two images.

In Figs. 7.5 and 7.6 we show some examples of intermediate histogram equalization for sets of three images.

Figure 7.5: Lighthouse sequence. First line: Original images. Second line: Final images.

$\vec{U}^1$ $\vec{U}^2$ $\vec{U}^3$

$\vec{u}^1$ $\vec{u}^2$ $\vec{u}^3$

Figure 7.6: Gros Horloge sequence. First line: Original images. Second line: Final images.

Chapter 8

Future perspectives

The research that I have developed so far can be divided in the four main categories:

1. *Perceptually-inspired color correction models*: mathematical analysis of the original Retinex algorithm and construction of a variational framework for Retinex-like color enhancement methods;
2. *HDR imaging* and *Tone mapping*: construction of high dynamic range images in non-static conditions and suitable reduction of the dynamic range to fit into the canonical low dynamic range $\{0, 1, \ldots, 255\}$;
3. *Statistics of natural images in color*: explanation of the appearance of 2D Fourier-like fluctuations in the achromatic plus color-opponent channels in terms of second order stationarity of covariance in natural images;
4. *Histogram transportation*: conception of a functional energy able to transfer the histograms of a set of images while respecting their geometry, without the need of an a-priori specification of a midway histogram.

As underlined throughout this document, the topics above are strongly interconnected, the glue among them being color; this should not come as a surprise due to its inherently interdisciplinary nature.

In line with this observation, I am willing to continue and even strengthen the multidisciplinary nature of my research by enlarging the present collaboration with mathematicians, physicists, engineers and computer scientists, to include metrologists, biologists, neuroscientists, cognitive psychologists, artists and art-restorers.

In the following sections I will provide a sketch of the research that I am willing to pursue in the future.

8.1 Collaboration with cognitive psychologists and psycho-physicists

I consider collaboration with psychologists and psycho-physicists strategical for the development of the research conducted so far. I will detail the collaboration in separated subsections.

8.1.1 Variational interpretation of context-driven effects in higher-level cognition

The parameters of the energy functionals presented in chapter 4 are not precisely determined. We were able to bound their variability on the basis of mathematical properties as algorithm convergence or stability, however perceptual experiments could help determine the optimal value of those parameters with respect to human vision fidelity.

A particularly important parameter is the weight function w appearing in the contrast term eq. ()4.2). Some psychophysical experiments, see e.g. [122] and [103], have paved the way for a quantitative measurement of the local induction weight w.

However, these experiments are far from being exhaustive, for this reason we have considered w as a normalized Gaussian with tunable standard deviation σ, i.e. $w(\|x-y\|) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\|x-y\|^2}{2\sigma^2}\right)$. Of course, the larger the value of σ, the less local is the contrast representation, and vice-versa.

I am currently working on an extension of [103] in [43], a work-in-progress paper where I am also analyzing the possibility of interpreting higher-level cognition context-driven effects with the variational framework developed for color perception.

8.1.2 Development of non-reference quality measures based on human visual properties

The non-reference quality measure assessment of a natural image is still an important open problem in image processing, see e.g. [81]. The specification 'no reference' indicates that there is no ground truth to compare with.

In these cases, quality measures are built mainly by taking advantage of statistical data coming from a more or less large number of subjective experiments.

The knowledge about the interaction of the human visual system with light stimuli coming from a natural scene is often essential to devise these subjective tests. Contrast perception plays a major role in the way humans interact with natural scenes.

In the last 8 years I have studied contrast perception in several different ways and I am now interested in developing quality measures starting from the findings that I have contributed to put in evidence.

In particular, it is quite remarkable that the contrast term in the wavelet-based functional described in section 4.4.2, obtained through a mathematical analogy with the spatially-based functional described in section 4.1, can be identified with the so-called Peli's contrast, proposed by E. Peli [82] on the basis of completely different argumentations.

Unfortunately, when Peli's contrast is used alone to provide contrast measures in HDR images, it fails to be coherent with what expected from a perceptual contrast measure.

I am currently working on suitable extensions of Peli's contrast measures with collaborators in the multimédia group of Télécom-ParisTech.

8.1.3 Analysis of further HVS features in terms of the variational framework

The variational framework discussed in chapter 4 shows that there is only one class of functionals coherent with color constancy, local contrast enhancement, adjustment to the average luminance and Weber-Fechner's law simultaneously.

Even if it's true that those are the most important HVS features with respect to natural images, there are other properties that have not yet been considered in the variational framework since they are mostly related to artificial images or regular pattern.

A particularly intriguing feature is assimilation [76], displayed e.g. by the well-known White's and Bezold's effects, see Figure 8.1.

Two distinctive features of assimilation is that it is evidently opposed to contrast enhancement and it is related with geometric patterns (which, although not very frequent in natural images, are common in many webpages).

A non-trivial modification of the energy functionals discussed in chapter 4 is needed to integrate assimilation effects.

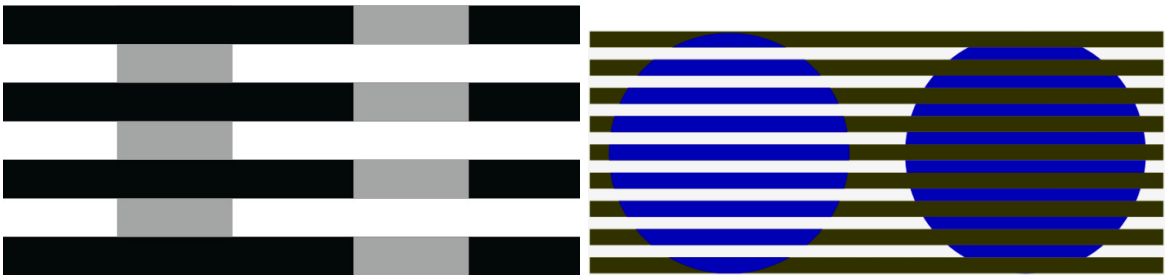


Figure 8.1: *Left*: White's effect. The gray rectangles in the two columns have exactly the same intensity, but those on the right are perceived as brighter, they are *assimilated* to the white rows, while those on the left are perceived as darker, they are *assimilated* to the black rows. *Right*: Bezold's effect. A sort of colored version of White's effect: the blue rows seem very different in brightness and quite different in hue, however they have exactly the same RGB triplet.

8.1.4 Color naming, metamer mismatching and color constancy

The world color survey [12] showed the existence of four *pure colors* which are consistently singled out in practically all human society, i.e. red, green, blue and yellow. They are considered pure because they are perceived as unmixed, unique, in the sense that they do not show tint of any other color, and for this reason they are deemed worth to be given a precise name in every human culture.

Philipona and O'Regan proposed in [84] a linear model to explain the existence of four unique colors. They claimed the existence of a linear operator $A : \mathbb{R}^3 \to \mathbb{R}^3$ which maps the accessible information (i.e. the retinal photon catches) about the illumination to the accessible information about light reflected from a surface.

They showed that surfaces with unique hues are characterized by a *singular behavior* of A: one of its eigenvalues is significantly different than the other two, in the sense that it is either much larger or smaller than the others.

They were able to define a so-called *singularity index* which match with surprising precision with the data of the world color survey.

In spite of this very interesting achievement, there are still a certain number of open questions. First of all, in theory, the linear operator A should depend on the surface characteristics, however, tests have shown that for the great majority of surfaces, one can use the same operator A. Why that happens? Why some surfaces violate the invariance of A? Is it possible to characterize them?

Secondly, the singularity index definition is quite involved and not directly related to colorimetric properties of surfaces, is it possible to define new singularity indexes closely related to, e.g. chromatic attributed as saturation, chromaticity and so on? A first attempt in this direction has been made in [118], but more research is needed.

Thirdly, the singularity index shows a very impressive stability with respect to changes of illumination. It is interesting to give a mathematical explanation for this behavior and to compare it to the predictions of the so-called metamer mismatching theory [63].

This goes towards a better comprehension of both human and digital color constancy.

8.2 Differential geometry of the space of perceived colors

In the extraordinary paper [101], H.L. Resnikoff deduced the Riemannian geometry of the space of perceived colors $\mathcal{P}$ and the Stiles metric on it via a Lie group-theoretical approach reinforcing Schrödingers axioms of color perception [125] with two original ones.

Probably due to the large variety of advanced mathematical concepts used by Resnikoff (e.g. Lie group and algebras and their representations, differential geometry and Jordan algebras), his paper

remained in a sort of oblivion.

I deem of great importance to reconsider Resnikoff's ideas and renew them in light of modern knowledge about color perception and the results described in this document.

To give an example: one of the most important drawback of Resnikoff's model was the lack of spatial interaction among points. As discussed in many chapters, color vision is highly context-dependent and the surround has a strong influence of color perception.

Thus, it seems natural to claim that the geometry of the space of perceived colors cannot be static, instead it should be dynamic and vary with respect to the spatial context. This is analogous to what happens in general relativity, where the curvature of space-time is modified by the presence of mass and energy.

As we did for the construction of the variational framework of chapter 4, I plan to use HVS features as a selection principle to find the tensor equations for the Riemannian metric on $\mathcal{P}$. The psycho-physical experiments on induction that I described in 8.1 will give fundamental data to finely tune the metric's parameter, in particular the induction weight.

8.3 Future research in statistics of natural images and multi-spectral images

I am interested in extending the work [94] about the statistics of natural images in color in essentially two ways, which I describe in separated subsections.

8.3.1 Generation of a large unbiased database of multispectral and HDR images

First of all, the numerical results that we have obtained cannot be considered wholly exhaustive to infer properties of the HVS. Even though RAW images give a nice approximation of radiance, they are still not sufficient to represent the real light information which constitutes the physical input for sight.

Such a representation is given by *multispectral images* $\mathcal{M}$, i.e. radiance matrices measured with a narrow spectral sensitivity curve. In practice, such matrices are associated with light energy with a wavelength confined in a few nm, so that the cones activation values for each pixel x, $L(x), M(x), S(x)$ can be computed simply by a weighted integration of the multispectral values over the visual spectrum Λ with weights given by the spectral sensitivity functions $\bar{l}, \bar{m}, \bar{s}$ of human cones, see Figure 2.2. Thus, for example, $L(x) = \int_\Lambda \mathcal{M}(x,\lambda)\bar{l}(\lambda)\, d\lambda$, and analogously for $M(x)$ and $S(x)$, where $\mathcal{M}(x,\lambda)$ is the value of the multispectral image in the pixel x at the wavelength λ.

The tests performed in [94] show that only starting from an order of 10^4 images the error in the estimation of the second order statistical features becomes negligible, this poses a serious problem when dealing with multispectral images. In fact, the only multispectral or high dynamic range images databases available have an order of 10^1 elements, and there is a strong bias in the semantic content of the pictures.

The main reason why large databases of multispectral or high dynamic range images are not available is related to the ghosting problem already mentioned in chapter 5. In fact, while HDR pictures need multiple time exposures to be generated, multispectral images need a time exposure long enough to allow the action of multiple filters associated with different wavelengths.

A new kind of multispectral sensors with a single sensor composed with a mosaic of filters has been recently put on the market, however, even this kind of hardware cannot be considered the solution to the above mentioned problem for two reasons. Firstly, such a sensor needs a demosaicking procedure to generate a multispectral image, this procedure must be carefully designed to avoid the introduction of typical artifacts as aliasing or elimination of high frequency components. Secondly, after demosaicking, the final multispectral image resolution is too small to be used to extract statistical information about covariance or contrast between distant pixels.

The generation of a large databases of multispectral and HDR images is thus far from being a simple, yet time consuming, picture acquisition process, instead, it involves a careful mathematical analysis and the development of new image processing techniques to avoid the above mentioned problems which are likely to distort all measures obtained with such a database.

Patch-based methods as those used in [2] could be an interesting starting point in this sense.

8.3.2 Analysis of the effects of HVS transformations on the statistical results

In my opinion, one of the most interesting achievement of [94] is the discovery that, after the application of the photo-electrical transduction represented by the Michaelis-Menten equation (2.1), the exponential decay speed of spatial covariance becomes independent of the chromatic channel and, as a consequence, spatiocovariant matrices commute.

This property, along with second order stationarity, is needed to prove the existence of a tensor product basis that supports the hypothesis of a separate elaboration of spatial and chromatic information of the visual signal.

The fact that the Michaelis-Menten equation (2.1) has such an important role in photo-electrical transduction and in the process just discussed leads quite naturally to the curiosity of investigating in more detail why this particular class of fractional transformations works so efficiently for the early stages of human vision. In other terms, it seems logical to ask: behind the simple analytical properties, is this class of functions the optimal solution to a certain, more profound yet still hidden, ecological principle of the early stages of human vision?

This kind of question is not unrelated with image processing, in fact the works of Olshausen and Field [74, 75] about the shape of receptive fields obtained through sparsity maximization have originated a whole new research field in image processing, and not only.

Another subject that I consider worth studying is the comprehension of the effects of chromatic adaptation, either expressed by the von Kries diagonal model (2.6) or by more elaborated ones, on the second and third principal chromatic components described in chapter 6.

Buchsbaum and Gottschalk have obtained some preliminary results about this topic in [18], showing that the order between the energies associated to the opponent channels can switch, see Figure 8.2. Such a study may have direct consequences on the structure of color opponent spaces widely used in the applications as e.g. the $l\alpha\beta$ [106].

Finally, I'm interested in the comparison and possible interaction between our work and that of Alleysson et al. [3, 4], who developed an interesting theory about the influence of cone mosaic in human retina on the spatial and chromatic elaboration of visual signal.

8.4 Analysis of problems related to astrophysical images

Passioned by astrophysics since my youth, I also plan to start studying the problems related to astrophysical imaging. This research field undergone to a revolution in the past 20 years thanks to the application of *adaptive optics* to telescopes. This technique allows correcting in real time the random fluctuations of a wave front thanks to a deformable mirror.

Thanks to adaptive optics, a whole new generation of astrophysical data is now available, e.g. the cosmic background radiation or the black mass maps obtained with gravitational lensing, a general relativistic effect due to the distortion of geometry of space-time provoked by an important accumulation of mass.

The thorough comprehension of problems related to astrophysical images needs, from one side, the knowledge of cosmology and general relativity and, on the other side, the ability to handle the most advanced image processing techniques. Thus, even this one is a highly multidisciplinary setting.

I believe that my background in theoretical physics, in particular the years passed studying quantum gravity, and the competences acquired in image processing during more than ten years can be unified with success in the future for the resolution of some problems in astrophysical imaging.

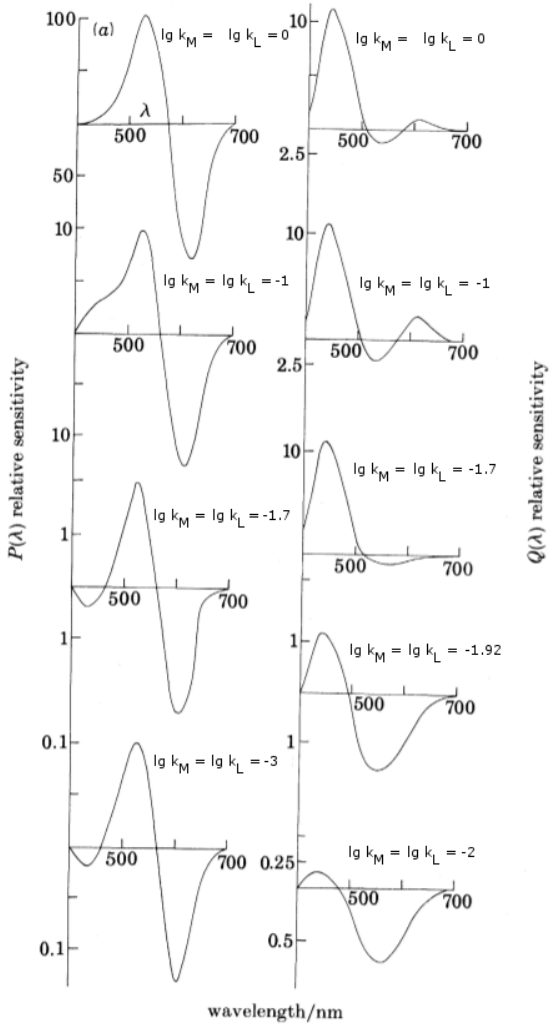


Figure 8.2: If $K = \mathrm{diag}(k_L, k_M, k_S)$ is the von Kries matrix, then the covariance matrix in the adapted state is KCK. The adapted $P(\lambda)$ and $Q(\lambda)$ plots change accordingly to the graphs shown above. Figure adapted from [18], page 104.

8.5 Development of new pedagogical strategies for scientific teaching

In the last century, the results in scientific disciplines grew at wit a rate faster than ever. The case of mathematics is particularly emblematic: until the end on the 19th century, universalism was still present, personified e.g. by David Hilbert or Henri Poincaré, but nowadays it is practically impossible to be able to produce important results in every important mathematical discipline.

Without having the ambition of universalism, it is already very complicated to become a specialist in just a few scientific disciplines.

Even remaining in the realm of a single discipline, the number of papers published each year is so large that writing an up-to-date the state of the art has become almost as difficult as writing an innovative paper.

In spite of this dramatic change in the research, the pedagogical strategies that are used in schools

and universities are almost identical to those of the beginning of the 20th century, with the only remarkable difference of digital resources.

Studies in cognitive psychology have shown that the most efficient learning is that represented by video: images, movement and sound contribute to facilitate the comprehension and memorization of concepts. This shows the growing success, and consequent development, of MOOC (Massive Open Online Course).

I am interested in collaborating with graphic designers and computer scientists to make collections of didactic videos. These videos must have a professional graphical rendering and interactive contents adapted to all type of student.

Of course, these didactic videos are not intended to replace direct teaching and the professor-to-student relationship, on the contrary, they are thought as a pedagogical help to strengthen and improve this relationship.

Bibliography

[1] A. Abadpour and S. Kasaei. A fast and efficient fuzzy color transfer method. In *Proceedings of the IEEE Symposium on Signal Processing and Information Technology*, pages 491–494, 2004.

[2] C. Aguerrebere, J. Delon, Y Gousseau, and P. Musé. Simultaneous hdr image reconstruction and denoising for dynamic scenes. In *2013 IEEE International Conference on Computational Photography (ICCP)*, 2002.

[3] D. Alleysson. Spatially coherent colour image reconstruction from a trichromatic mosaic with random arrangement of chromatic samples. *Ophthalmic and Physiological Optics*, 30(5):492–502, 2010.

[4] D. Alleysson and D. Méary. Neurogeometry of color vision. *Journal of Physiology*, 106(5):284–296, 2012.

[5] L.A. Ambrosio, N. Gigli, and G. Savaré. *Gradient flows in metric spaces and in the space of probability measures.* Lectures in Mathematics, Birkhauser, 2005.

[6] M. Ashikhmin. A tone mapping algorithm for high contrast images. In *Eurographics Workshop on Rendering*, pages 1–11. P. Debevec and S. Gibson Eds., 2002.

[7] J.J. Atick, Z. Li, and A.N. Redlich. Understanding retinal coding from first principles. *Neural computation*, 4:559–572, 1992.

[8] F. Attneave. Some informational aspects of visual perception. *Physchol. Rev.*, 61:183–193, 1954.

[9] T.O. Aydin, R. Mantiuk, K. Myszkowski, and H.P. Seidel. Dynamic range independent image quality assessment. In *ACM SIGGRAPH 2008 papers*, pages 1–10, 2008.

[10] C. Ballester, V. Caselles, L. Igual, J. Verdera, and B. Rougé. A variational model for p+ xs image fusion. *International Journal of Computer Vision*, 69(1):43–58, 2006.

[11] H.B. Barlow. Possible principles underlying the transformations of sensory messages. *Sensory Communications*, pages 217–234, 1961.

[12] A. Berlin and P. Kay. *Basic Color Terms: Their Universality and Evolution.* University of California Press, 1969.

[13] A. Berman and R.J. Plemmons. *Nonnegative Matrices in the Mathematical Sciences.* SIAM, 1987.

[14] M. Bertalmío, V. Caselles, and E. Provenzi. Issues about the retinex theory and contrast enhancement. *International Journal of Computer Vision*, 83:101–119, 2009.

[15] M. Bertalmío, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. *IEEE Trans. on Image Processing*, 16:1058–1072, 2007.

[16] G. Boynton. Color vision: How the cortex represents color. *Current biology*, 12:R838–R840, 2002.

[17] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:337–350, 1980.

[18] G. Buchsbaum and A. Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proc. Royal Society of London B*, 220:89–113, 1983.

[19] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[20] K. Chiu, M. Herf, P. Shirley, S. Swamy, C. Wang, and K. Zimmerman. Spatially nonuniform scaling functions for high contrast images. In *Proceedings of Graphics Interface 93*, pages 245–253. Morgan Kaufmann, 1993.

[21] P.G. Ciarlet, B. Miara, and J-M. Thomas. *Introduction to numerical linear algebra and optimisation*. Cambridge Univ. Pr., 1989.

[22] I.J. Cox, S. Roy, and S.L. Hingorani. Dynamic histogram warping of image pairs for constant image brightness. In *IEEE International Conference on Image Processing*, volume 2, pages 366–369, 1995.

[23] P. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *Proc. of the 24th annual conf. on Computer graphics*, pages 369–378, 1997.

[24] J. Delon. Fine comparison of images and other problems. *Ph.D. Thesis, École Normale Supérieure de Cachan*, 2004.

[25] J. Delon. Midway image equalization. *J. Math. Imaging Vis.*, 21(2):119–134, 2004.

[26] J.B. Derrico and G. Buchsbaum. A computational model of spatio-chromatic image coding in early vision. *Journal of Visual Communication and Image Representation*, 2:31–38, 1991.

[27] F. Dunn, M. Lankheet, and F. Rieke. Light adaptation in cone vision involves switching between receptor and post-receptor sites. *Nature*, 449:603–606, 2007.

[28] F. Durand and J. Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. pages 257–266. SIGGRAPH 2002, Proceedings of the 29th annual conference on Computer graphics and interactive techniques, 2002.

[29] A. Eden, M. Uyttendaele, and R. Szeliski. Seamless image stitching of scenes with large motions and exposure differences. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2498–2505, 2006.

[30] M.D. Fairchild. *Color appearance models*. Wiley, 2005.

[31] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *ACM Trans. Graphics*, volume 21 (3), pages 249–256, 2002.

[32] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. An analysis of visual adaptation and contrast perception for tone mapping. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(10):2002–2012, 2011.

[33] S. Ferradans, M. Bertalmío, E. Provenzi, and V. Caselles. Generation of hdr images in non-static conditions based on gradient fusion. In *Proceedings of the VISAPP Conference, February 24-26, Rome, Italy*, pages 31–37. SciTePress, 2012.

[34] S. Ferradans, R. Palma-Amestoy, and E. Provenzi. An algorithmic analysis of variational models for perceptual local contrast enhancement. *Image Processing On Line*, 5(10):219–233, 2015.

[35] J.A. Ferwerda, S.N. Pattanaik, P. Shirley, and D.P. Greenberg. A model of visual adaptation for realistic image synthesis. In *Proceedings of SIGGRAPH 96, Computer Graphics Proceedings*, pages 249–258. Addison Wesley, 1996.

[36] D.J. Field. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am.*, 4(12):2379–2394, December 1987.

[37] O. Gallo, N. Gelfand, W. Chen, M. Tico, and K. Pulli. Artifact-free high dynamic range imaging. *IEEE International Conference on Computational Photography (ICCP)*, April 2009.

[38] T. Gevers, A. Gijsenij, J. van de Weijer, and J-M. Geusebroek. *Color in Computer Vision, Fundamentals and Applications*. Wiley, 2012.

[39] B.E. Goldstein. *Sensation and Perception, 9th Edition*. Cengage Learning, 2013.

[40] R.C. Gonzales and R.E. Woods. *Digital image processing*. Prentice Hall, 2002.

[41] M. Granados, B. Ajdin, M. Wand, C. Theobalt, H.-P. Seidel, and H.P.A. Lensch. Optimal hdr reconstruction with linear digital cameras. In *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 215–222, 2010.

[42] R.M. Gray. *Toeplitz and Circulant Matrices: A review*. Now Publishers Inc., 2006.

[43] G. Gronchi and E. Provenzi. A variational model for context-driven effects in higher-level cognition. *To be submitted to Jour. of Math. Psych.*, 2016.

[44] T. Grosch. Fast and robust high dynamic range image generation with camera and object movement. In *Vision, Modeling and Visualization, RWTH Aachen*, pages 277–284, 2006.

[45] M. Grundland and N.A. Dodgson. Color histogram specification by histogram warping. In *Proceedings of SPIE*, volume 5667, pages 610–624, 2005.

[46] J. Hays and A. Efros. Scene completion using millions of photographs. In *ACM Transactions on Graphics (SIGGRAPH)*, volume 26, 2007.

[47] Y. Heo, K. Lee, S. Lee, Y. Moon, and J. Cha. Ghost-free high dynamic range imaging. In *Computer Vision - ACCV 2010*, volume 6495 of *Lecture Notes in Computer Science*, pages 486–500. Springer Berlin-Heidelberg, 2011.

[48] J. Huang and D. Mumford. Statistics of natural images and models. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, 1999.

[49] D.H. Hubel. *Eye, Brain, and Vision*. Scientific American Library, 1995.

[50] R.W.G. Hunt and M.R. Pointer. *Measuring Colour*. Wiley, 2014.

[51] A. Hyvärinen, J. Hurri, and P.O. Hoyer. *Natural Image Statistics: A Probabilistic Approach to Early Computational Vision*. Springer, 2009.

[52] K. Jacobs, C. Loscos, and G. Ward. Automatic high-dynamic range image generation for dynamic scenes. *IEEE Computer Graphics and Applications*, 28:84–93, 2008.

[53] J. Jiang, D. Liu, J. Gu, and S. Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *WACV, IEEE Workshop on the Applications of Computer Vision*, pages 168–179, 2013.

[54] D.J. Jobson, Z. Rahman, and G.A. Woodell. A multiscale Retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on image processing*, 6(7):965–976, july 1997.

[55] S.B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22:319–325, July 2003.

[56] E.A. Khan, A.O. Akyuz, and E. Reinhard. Ghost removal in high dynamic range images. In *IEEE International Conference on Image Processing*, pages 2005–2008, 2006.

[57] H. Kolb, E. Fernandez, and R. Nelson. *Webvision. The Organization of the Retina and Visual System*. University of Utah Health Sciences Center, 1995.

[58] H. Kotera. A scene-referred color transfer for pleasant imaging on display. In *IEEE International Conference on Image Processing*, volume 2, 2005.

[59] G. Krawczyk, K. Myszkowski, and H.P. Seidel. Lightness perception in tone reproduction for high dynamic range images, 2005.

[60] J. Kuang, G.M. Johnson, and M.D. Fairchild. icam06: A refined image appearance model for hdr image rendering. *J. Vis. Commun. Image R.*, 18:406–414, 2007.

[61] E.H. Land. The Retinex theory of color vision. *Scientific American*, 237:108–128, 1977.

[62] E.H. Land and J.J. McCann. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, January 1971.

[63] Funt B. Logvinenko, A.D. and C. Godau. Metamer mismatching. *IEEE Trans. on Im. Proc.*, 23(1):34–43, 2014.

[64] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 3rd edition, 1999.

[65] S. Mann. Comparametric equations with practical applications in quantigraphic image processing. *Image Processing, IEEE Transactions on*, 9(8):1389–1406, 2000.

[66] S. Mann and R. W. Picard. On being undigital with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proceedings of IS&T*, pages 442–448, 1995.

[67] R. Mantiuk, K. Myszkowski, and H.P. Seidel. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP)*, 3 (3):286–308, 2006.

[68] John J. McCann. Capturing a black cat in shade: past and present of Retinex color appearance models. *Journal of Electronic Imaging*, 13(1):36–47, January 2004.

[69] A.A. Michelson. *Studies in Optics*. Chicago University Press, 1927.

[70] T. Mitsunaga and S.K. Nayar. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, pages 374–380 Vol. 1, 1999.

[71] J. Morovic and P.L. Sun. Accurate 3D image colour histogram transformation. *Pattern Recognition Letters*, 24(11):1725–1735, 2003.

[72] L. Neumann and A. Neumann. Color style transfer techniques using hue, lightness and saturation histogram matching. *Computational Aesthetics in Graphics, Visaulization and Imaging 2005, L. Neumann, M. Sbert, B. Gooch, and W. Purgathofer, editors.*, pages 111–122, 2005.

[73] Y. Ohta, T. Kanade, and T. Sakai. Color information for region segmentation. *Computer graphics and image processing*, 13:222–241, 1980.

[74] B. Olshausen and D.J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Letters to Nature*, 381:607–609, 1996.

[75] B. Olshausen and D.J. Field. Sparse coding with an overcomplete basis set: A strategy employed by v1? *Vision research*, 37:607–609, 1997.

[76] X. Otazu, M. Vanrell, and C.A. Párraga. Multiresolution wavelet framework models brightness induction effects. *Vision Research*, 48:733–751, 2008.

[77] R. Palma-Amestoy, E. Provenzi, M. Bertalmío, and V. Caselles. A perceptually inspired variational framework for color enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(3):458–474, 2009.

[78] N. Papadakis, E. Provenzi, and V. Caselles. A variational model for histogram transfer of color images. *IEEE Transactions on Image Processing*, 20(6):1682–1695, 2011.

[79] C. Párraga, T. Troscianko, and D. Tolhurst. Spatiochromatic properties of natural images and human vision. *Current biology*, 6(12):483–487, 2002.

[80] S.N. Pattanaik, J. Tumblin, H. Yee, and D.P. Greenberg. Time-dependent visual adaptation for fast realistic image display. In *Proceedings of SIGGRAPH*, pages 47–54, 2000.

[81] M. Pedersen and J. Y. Hardeberg. Full-reference image quality metrics: Classification and evaluation. *Found. Trends. Comput. Graph. Vis.*, 7(1):1–80, 2012.

[82] E. Peli. Contrast in complex images. *Journal of Opt. Society of Am. A*, 7(10):2032–2040, 1990.

[83] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. In *ACM SIGGRAPH 2003 Papers*, pages 313–318, 2003.

[84] D. Philipona and J.K. O'Regan. Color naming, unique hues and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23:331–339, 2006.

[85] E. Pichon, M. Niethammer, and G. Sapiro. Color histogram equalization through mesh deformation. In *IEEE International Conference on Image Processing, Barcelone*, volume 2, pages 117–120, 2003.

[86] G. Piella. Image fusion for enhanced visualization: A variational approach. *International Journal of Computer Vision*, 83:1–11, 2009.

[87] F. Pitié. Statistical signal processing techniques for visual post-production. *Ph.D. Thesis, University of Dublin, Trinity College*, 2007.

[88] F. Pitié and A. Kokaram. The Linear Monge-Kantorovitch Colour Mapping for Example-Based Colour Transfer. In *IEEE European Conference on Visual Media Production (CVMP'06)*, London, 2007.

[89] F. Pitie, A.C. Kokaram, and R. Dahyot. N-dimensional probability density function transfer and its application to colour transfer. In *International Conference on Computer Vision, Beijing*, 2005.

[90] F. Pitié, A.C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.*, 107(1-2):123–137, 2007.

[91] E Provenzi. Boosting the stability of wavelet-based contrast enhancement of color images through gamma transformations. *Journal of Modern Optics*, 60:1145–1150, 2013.

[92] E. Provenzi and V. Caselles. A wavelet perspective on variational perceptually-inspired color enhancement. *International Journal of Computer Vision*, 106:153–171, January 2014.

[93] E. Provenzi, L. De Carli, A. Rizzi, and D. Marini. Mathematical definition and analysis of the retinex algorithm. *Journal of the Optical Society of America A*, 22(12):2613–2621, December 2005.

[94] E. Provenzi, J. Delon, Y. Gousseau, and B. Mazin. On the second order spatiochromatic structure of natural images. *Vision Research*, accepted for publication.

[95] E. Provenzi, M. Fierro, A. Rizzi, L. De Carli, D. Gadia, and D. Marini. Random spray retinex: A new retinex implementation to investigate the local properties of the model. *IEEE Transactions on Image Processing*, 16:162–171, January 2007.

[96] E. Provenzi, C. Gatta, M. Fierro, and A. Rizzi. Spatially variant white patch and gray world method for color image enhancement driven by local contrast. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 30:1757–1770, 2008.

[97] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer Graphics and Applications*, 21:34–41, 2001.

[98] E. Reinhard and K. Devlin. Dynamic range reduction inspired by photoreceptor physiology. *IEEE Trans. on Visualization and Computer Graphics*, 11(1):13–24, 2005.

[99] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. on Graph.*, 21:267–276, 2002.

[100] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging, Acquisition, Display, And Image-Based Lighting.* Morgan Kaufmann Ed., 2005.

[101] H.L. Resnikoff. Differential geometry and color perception. *Journal of Mathematical Biology*, 1:97–131, 1974.

[102] A. Rizzi, C. Gatta, and D. Marini. A new algorithm for unsupervised global and local color correction. *Pattern Recognition Letters*, 24:1663–1677, 2003.

[103] M.E. Rudd and I.K. Zemach. Quantitive properties of achromatic color induction: An edge integration analysis. *Vision Research*, 44:971–981, 2004.

[104] D.L. Ruderman. Origin of scaling in natural images. *Vision Research*, 37:3385–3398, 1996.

[105] D.L. Ruderman and W. Bialek. Statistics of natural images: Scaling in the woods. *Phys. Rev. Lett.*, 73:814–817, 1994.

[106] D.L. Ruderman, T.W. Cronin, and C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *J. Opt. Soc. Am. A*, 15(8):2036–2045, August 1998.

[107] G. Sapiro and V. Caselles. Histogram modification via differential equations. *Jour. of Diff. Eq.*, 135:238–266, 1997.

[108] C. Schlick. Quantization techniques for visualization of high dynamic range pictures. In *Proceedings of the 5th Eurographics Workshop on Rendering Workshop*, pages 7–20. Springer Verlag, 1994.

[109] R. Shapley and C. Enroth-Cugell. *Visual adaptation and retinal gain controls*, volume 3, chapter 9, pages 263–346. Progress in Retinal Research, 1984.

[110] S. Shevell. Saturation in human cones. *Vision Research*, 17:427–434, 1977.

[111] G. Strang and T. Nguyen. *Wavelets and Filter Banks.* Wellesley-Cambridge Press, 1996.

[112] D. Tamburrino, D. Alleysson, L. Meylan, and S. Susstrunk. Digital camera workflow for high dynamic range images using a model of retinal processing. In *Proceedings SPIE*, volume 6817, 2008.

[113] M.W. Tao, M.K. Johnson, and S. Paris. Error-tolerant image compositing. In *Proceedings of the 11th European conference on Computer vision: Part I*, ECCV 2010, pages 31–44, 2010.

[114] A. Tomaszewska and R. Mantiuk. Image registration for multi-exposure high dynamic range image acquisition. In *Proc. Intl. Conf. Central Europe on Computer Graphics, Visualization, and Computer Vision (WSCG)*, 2007.

[115] J. Tumblin and H. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, pages 42–48, 1993.

[116] J. Tumblin and G. Turk. Lcis: A boundary hierarchy for detail-preserving contrast reduction. In *SIGGRAPH: Conference Proceedings*, pages 83–90, 1999.

[117] J. Valeton and D. van Norren. Light adaptation of primate cones: An analysis based on extracellular data. *Vision Research*, 23:1539–1547, 1983.

[118] J. Vazquez-Corral, J.K. O'Regan, M. Vanrell, and G.D. Finlayson. A new spectrally sharpened sensor basis to predict color naming, unique hues, and hue cancellation. *Journal of Vision*, 12:1–14, 2012.

[119] C. Villani. *Topics in optimal transportation.* American Mathematical Society, 2003.

[120] J. von Kries. Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, pages 145–158, 1902.

[121] A. Wade and B. Wandell. Chromatic light adaptation measured using functional magnetic resonance imaging. *Journal of neuroscience*, 22:8148–8157, 2002.

[122] H. Wallach. Brightness constancy and the nature of achromatic colors. *Journal of Experimental Psychology*, 38(3):310–324, 1948.

[123] G. Ward, H. Rushmeier, and C. Piatko. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Transactions on Visualization and Computer Graphics*, 3:291–306, 1997.

[124] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from handheld exposures. *Journal of graphic tools*, 8:17–30, 2003.

[125] G. Wyszecky and W. S. Stiles. *Color science: Concepts and methods, quantitative data and formulas.* John Wiley & Sons, 1982.

[126] S. Zeki. *Vision of the brain.* Blackwell Publishing Incorporated, 1993.