Measuring and Enhancing the Contrast and Quality of Digital Images

Gabriele Simone

Faculty of Mathematics and Natural Science
University of Oslo

This dissertation is submitted for the degree of
Philosophiae Doctor (PhD)

December, 2016

Acknowledgements

The present research work has been funded by the Research Council of Norway through the project #182692: “Perceptual image difference metrics–a unifying approach to image representation and reproduction”.
I would like to express my gratitude to my supervisors Jon Yngve Hardeberg, Ivar Farup and Fritz Albregtsen for their outstanding contribution for the success of this research work, the colleagues and coauthors Marius Pedersen for his contribution in the image quality research, Alessandro Rizzi for his contribution in the image enhancement research, Claudio Oleari for his contribution in the colorimetry research.
Thanks for technical suggestions, feedback and for participating in my research work to the all members of the Norwegian Colour and Visual Computing Lab, Norwegian University of Science and Technology, University of Oslo, Gruppo del Colore, University of Milano, University of Parma, CIMET and international students.
Heartfelt thanks for technical and moral support sharing joy and sorrow goes to my parents Mario and Odilia, my fiance’ Estefani Perez De La Cruz and her family, my colleagues, friends, brothers and sisters Daniel Hartung, Steven Le Moan, Christoffer Mogensen, Dibakar Ray Pant, Wilhelm Alexander Stange, Jean Baptiste Thomas, Luca Alfieri, Michele Alfieri, Morgan Amerio, Jessica Bella, Chiara Bornino, Michele Bracco, Daniela Di Bitonto, Silvia Giudici, Giuseppe Gorgoglione, Massimiliano Guerini, Francesco La Barretta, Alessandro Mazza, Alessandra Perrotti, Giulia Porta, Sara Turchetta.
Gratitude goes to the STMicroelectronics and in particular to the Castelletto team, Giancarlo Asnaghi, Riccardo Brugnoli, Paolo Farinelli, Luca Furiato, Gian Antonio Sampietro, Alberto Saviotti, Lorena Simoni, Isabella Zampollo.
Thanks to the all groups of friends, Gjøvik Friends, Drammen Friends, Triple Threat, Gruppo Salute, Gruppo Ciurma, La Base, Gruppo Balordi Lago Maggiore and Felice Cavallotti.
Finally, I would like to acknowledge YOU that you are reading this research work.

Gabriele Simone

P.S.: if you have not read your name, you are mistaken “YOUR NAME” is just here to be acknowledged. WHAT THE EYES SEE AND THE EARS HEAR, THE MIND BELIEVES (“SWORDFISH 2001”).

Abstract

The rapid spread of digital imaging devices (e.g. cameras, TVs and smart phones) has introduced many scientific and technical terms in common language. Among these terms, we can find image contrast, image quality and image enhancement, which from a scientific point of view cover a wide field in different areas of research, e.g. physics, biology and computer science.

Recent research studies have highlighted contrast as one of the key attributes and quality indicators of a digital image. Also manipulating the contrast of an image may lead to image enhancement. For this reason, the purpose of this research is to contribute new knowledge in the field of image contrast, image quality and image enhancement models.

First we briefly revisit the history of contrast and the first studies conducted in the field of optics, which date back to the beginning of the 20^{th} century, before turning focus to the development of global and local measures of contrast for complex images. Building on the theories of local measures developed in these last decades, we present a novel measure of contrast for digital color images named Weighted–Level Framework (WLF), which is based on three mathematical aspects: multilevel analysis, Difference of Gaussians model and variance weighting.

Second we give a brief overview of the four centuries old history of image quality, followed by an introduction to color difference formulae and image quality metrics. Combining the Weighted–Level Framework and a recent color difference formula, named Euclidean color–difference formula for small–medium color differences in log–compressed OSA–UCS space, we propose a novel metric, named WLF–DEE, for estimating the perceived difference in contrast between a reference image and a reproduction.

Finally, following an introduction to the field of image enhancement and spatial color algorithms, we present the well–known, almost 50 year old, Retinex model and its extended family of algorithms. The main characteristics of Retinex algorithms is performing spatial comparisons through a mechanism called reset, and afterwards calculating appearance. Exploiting these characteristics, we present two new spatial color algorithms named Spatio–temporal Retinex–like Envelope with Total Variation (STRETV) and Termite Retinex (TR), respectively. STRETV replaces the stochastic sampling mechanism for reset performed by

its predecessor STRESS with the minimization of the Total Variation method. TR instead replaces the random paths of the original Retinex with a series of paths generated with the Ant Colony Optimization method.

In order to show advantages and disadvantages of the propose measures, metrics and algorithms, the results are supported by statistical analysis.

In particular, performances of WLF in terms of Pearson and Spearman correlation are presented on a custom built image database against three state–of–the–art local contrast measures, Peli, Tadmor and Tolhurst, and Rizzi et al. A psychophysical experiment was performed in a lab controlled environment followed by a web based experiment. A performance of 0.84 in Pearson correlation and of 0.80 in Spearman correlation is achieved in relation to observers' perceived contrast in the lab controlled environment. Confidence intervals for Pearson and Spearman correlation show WLF to have the same performance of Peli and Rizzi et al. contrast measures and to be significantly better than Tadmor and Tolhurst measure. A decrease in correlation in the web based experiment and a comparison of the correlation coefficients indicate that measuring contrast in uncontrolled environments can be significantly different at 99% than measuring in lab controlled environment.

The performance of WLF–DEE is analyzed against five state–of–the–art image quality metrics on four different categories belonging to the well–known Tampere Image Database (TID) and on two databases providing different distortions directly related to color and contrast. Among the several metrics available in literature, PSNR, PSNR–HVS–M, S–CIELAB, SSIM and VSNR were chosen for comparison. Performances in Pearson correlation of 0.68, 0.77, 0.71 and 0.67 are achieved respectively for masked noise, quantization noise, image denoising and contrast change categories of the TID database. Confidence intervals show WLF–DEE to have the same performance as SSIM and S–CIELAB and to be significantly better than PSNR and VSNR on the four categories of the TID database. Performances of 0.65 and 0.49 in Pearson correlation are found on the two databases providing different distortions directly related to color and contrast; Confidence intervals show WLF to have the same performance as other state–of–the–art metrics on both databases except for S–CIELAB on Pedersen database, where the latter has slightly higher performance. Furthermore, the Root–Mean–Square Error confirm WLF–DEE to be in line with other state–of–the–art metrics when estimating the perceived magnitude of contrast between a reference and a reproduction.

Finally, the performances of STRETV and TR are measured against two different Retinex models, respectively. In a psychophysical experiment, STRETV is compared with STRESS on ad–hoc built database. A sign–test at 95% confidence interval shows that STRETV and STRESS are not significantly different in terms of overall image quality, but at the same time a higher preference for STRETV is achieved due to a lower perception of noise. As well, in a

psychophysical experiment TR is compared with RSR on ad–hoc built database, and a sign–test at 95% confidence interval shows that TR is significantly better than RSR both in terms of overall image quality and perceived noise.

Target applications such as unsupervised detail enhancement, dynamic range stretching and color correction are shown for both STRETV and TR.

Summarizing, four novel tools for color digital images are presented: a measure of contrast, an image quality metric and two spatial color algorithms. The applicability and the efficacy of these tools are shown through a variety of use cases.

Table of contents

Papers

List of included papers

PAPER A

G. Simone, M. Pedersen, and J. Y. Hardeberg. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation*, 23(3):491–506, Apr 2012

PAPER B

G. Simone, M. Pedersen, and J. Y. Hardeberg. Measuring perceptual contrast in uncontrolled environments. In *EUVIP 2010 2nd European Workshop on Visual Information Processing*, pages 102–107, Paris, France, Jul 2010

PAPER C

G. Simone, M. Pedersen, I. Farup, and C. Oleari. Multi–level contrast filtering in image difference metrics. *EURASIP Journal on Image and Video Processing*, 2013(1):39, Jul 2013

PAPER D

G. Simone and I. Farup. Spatio–temporal Retinex–like envelope with total variation. In *CGIV 2012 6th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 176–181, Amsterdam, The Netherlands, May 2012

PAPER E

G. Simone, G. Audino, I. Farup, F. Albregtsen, and A. Rizzi. Termite Retinex: A new implementation based on a colony of intelligent agents. *Journal of Electronic Imaging*, 23(1):013006–1–13, Jan 2014

List of related papers

- M. Pedersen, A. Rizzi, J. Y. Hardeberg, and G. Simone. Evaluation of contrast measures in relation to observers perceived contrast. In *CGIV 2008 4th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 253–256, Terrassa, Spain, Jun 2008

- G. Simone, C. Oleari, and I. Farup. Performance of the Euclidean color-difference formula in log-compressed OSA–UCS space applied to modified image-difference metrics. In *AIC 2009 11th Congress of the International Colour Association*, page 81, Sydney, Australia, Oct 2009

- S. A. Ajagamelle, M. Pedersen, and G. Simone. Analysis of the difference of gaussians model in image difference metrics. In *CGIV 2010 5th European Conference on Colour in Graphics, Imaging, and Vision*, IS&T Proceedings, pages 489–496, Joensuu, Finland, Jun 2010

- G. Simone, V. Caracciolo, M. Pedersen, and F. A. Cheikh. Evaluation of a difference of gaussians based image difference metric in relation to perceived compression artifacts. In *ISVC 2010 6th International Symposium Advances in Visual Computing*, Springer Proceedings, pages 491–500, Las Vegas, NV, USA, Nov 2010

- M. Pedersen, G. Simone, M. Gong, and I. Farup. A total variation based color image quality metric with perceptual contrast filtering. In *PCSPA 2011 2nd International conference on Pervasive Computing, Signal Processing and Applications*, Gjøvik, Norway, Sep 2011

Chapter 1

Introduction

Since the '70s, society is witnessing a rapid evolution of digital imaging devices (e.g., cameras, high definition TVs, 3D monitors and laser printers). This growth of technology together with an aggressive electronic media market has propagated many technical and scientific research terms (e.g., resolution, dpi and gamut) into everyday language.

Furthermore, multimedia applications are also developing fast and allowing to manage rather large collections of images (e.g., image archiving, network image transmission, document imaging, digital photography, medical imaging and remote sensing). Thus, the demand for efficient, automatic, unsupervised and versatile tools for image handling is more pressing than it has ever been before.

During acquisition, communication and display, digital images are subjected to a wide variety of distortions (e.g., noise and compression artifacts). These distortions can deteriorate several image attributes and lead to a degradation, in some cases unrecoverable, of the quality of the image. To ensure a certain quality level of multimedia applications and services, automatic means to measure image attributes and evaluate the quality of digital images are required.

From literature to industry, hundreds of attributes can be found to describe and measure the quality of an image (e.g., color, lightness and sharpness). Recent studies, with the purpose to reduce the number of attributes to a manageable set, have identified contrast as one of the fundamental attributes to describe the quality of an image [222]. Furthermore, contrast is a common quality index in technical specifications used for up–selling in the digital media market.

From a scientific point of view, the studies on contrast and quality cover a wide area of research and involve several disciplines (e.g., physics, biology and computer science). Unfortunately, since the initial research in the field of imaging, it has become clear how difficult it is to find a proper definition and an objective measure of contrast and quality in digital images.

The only certainty would remain subjective evaluation, but automatic objective measures are inevitably required when dealing with a large number of images.

The first studies on measuring contrast date back to the beginning of the 20^{th} century in the field of optics [187]. Analysis of patches in uniform background, sinusoids and other periodic patterns led to extend the Weber and Michelson formulae, in particular during the '80s with the development of a series of physical contrast measures [131, 35, 304].

At the same time of the studies in optics and in signal processing theory, contrast is an important factor in analogue photography. Contrast describes how photographic materials reproduce the log exposure range presented to an emulsion in the camera or at the printing stage [8].

With the growth of digital photography and digital display, the physical contrast measures were inherited to measure the contrast of complex images. These measures assign a single contrast value to the whole image and thus are intended as global measures of contrast. The limitations of such measures found at the end of the 20^{th} century [101, 230, 151], where perceived contrast varies across the image and does not correlate with physical measures of contrast, led to the development of local measures of contrast.

As a consequence, contrast emerges as a significant field of research and becomes more and more a relevant topic for color images with the spread of mass color digital media. Nowadays, at least 50 measures of contrast for greyscale/color digital images can be found in the literature. Although there is no a universal consensus on how to define and measure the perceptual contrast objectively [18], different measures of contrast have found success in target applications [155, 318].

The first studies on image quality are even older than contrast and date back to the beginning of the 16^{th} century with the invention of optical instruments [75]. However, image quality needed to await almost three centuries of substantial color milestones before getting serious attention among researchers.

Among the first and most important milestones, the introduction of the CIELAB color space in 1976 and all the related color difference formulae finding their success as tools for measuring perceptual difference between uniform patches of colors in the colorant industry [309]. On the other hand, the extension of these formulae to complex scenes has failed when trying to estimate the difference in terms of quality between a reference image and a reproduction of it using a pixelwise difference in terms of color [316, 132].

This has lead researchers to refine the concept of image quality with respect to image difference and build more complex and efficient tools. The tools for comparing and measuring the difference between two images are known as image metrics and exist of various types based on their main goal (e.g., measuring general image quality, detecting specific distortions,

benchmarking and optimizing a process) [295]. Today, there are hundreds of metrics available in literature with an ongoing debate on how to define and measure the quality of an image universally and objectively [226].

The 70's mark the migration from analog to digital technology and the spread of many commercial imaging devices. Although many old analog devices are off duty, a lot of materials (e.g., negatives, roll films and glass plates) from these devices are still available.

As a consequence, in these last three decades, the phenomenon of digitization and archiving has emerged and grown rapidly in institutions, companies and privately for several reasons. Considering e.g., the process of digitally archiving a set of roll films, one reason can be the amount of physical storage required. Nowadays, a down market usb drive of size of a few cm^3 is able to contain a number of digital images equal to the number of negatives that would fit in a storage room of several m^3. This advantage becomes then remarkable when digitizing color motion pictures [236].

Another motivation is the opportunity to unearth, recollect and redistribute more easily old and historical media. As well, copying, backing–up and sharing digital media are much simpler tasks to perform [236, 54].

Today, transforming an old analog image to a digital one (e.g., scanning a printed photograph and storing it on a computer) is a relatively straightforward process, but many problems may occur regarding the resulting quality. Film based photographs are subject to degradation due to the the instability of chemical substances over the time. Thus, the scanned image may end up with faded colors, bad illumination, wrong tones, etc. If not properly stored, other defects (e.g., scratches and dust) may also be present [236, 54].

Digital imaging devices are not infinitely powerful, and thus defects may occur also in images acquired with these hardwares. These defects are commonly referred to as artifacts and generally caused by hardware failures and/or software errors [236, 8].

In order to estimate and rebuild the original image, a post–process called image restoration is typically performed. This process can be done either manually or automatically by the use of dedicated tools such as professional photo softwares [236, 54].

Often, we come across digital images with a pleasant look, but subtle artifacts may be revealed after close inspection, and image quality might not be optimal any longer. In many cases, subjectivity plays a fundamental role in image quality assessment.

In these situations, a full image restoration might not be not necessary, but a process of image enhancement. These two processes may overlap, as image restoration techniques may bring an enhancement of the processed image, and image enhancement techniques may remove artifacts. In plain terms, image restoration attempts to correct defects while image enhancement modifies the attributes of the image attempting a specific task usually defined by

the user. Both processes cover a lot of territory, and goals are situational, but in general both have the purpose of improving the perceived quality of the processed image objectively and subjectively [236, 8, 180].

When dealing with large collections of digital images, manual restoration or enhancement becomes unsustainable, and for this reason automatic unsupervised tools become necessary.

1.1 Aims

In order to contribute to new knowledge and tools in the fields of image contrast, image quality and image enhancement, a set of research questions, which form a sort of red lines and sub–goals, have evolved during the theoretical and practical development of this research work. In particular, we can outline the following specific research questions:

1. Can we improve the Rizzi et al. measure of contrast using the Difference of Gaussians model instead of using a weighted 8–neighborhood?
2. Are there significant differences between assessing contrast in uncontrolled environments and a lab controlled environment?
3. Can we estimate the quality of a reproduction image with respect to an original image in terms of contrast using the Difference of Gaussians model?
4. Can we improve the STRESS algorithm, estimating the two envelopes through the Total Variation method instead of using stochastic sampling?
5. Can we improve the original Retinex algorithm using the Ant Colony Optimization model for generating intelligent paths across the image?

1.2 Research methodology

As this research work covers a wide area from a theoretical to a practical point of view, in order to answer to the previous research questions, the general research methodology can be divided in two main activities:

1. **Development**: in this first activity, the process of awareness of the problem in the given field is embraced through scientific literature and/or new technical developments. Once the problem is analyzed, possible solutions, either new methods or methods employed from existing literature are proposed in a new way. This activity requires the development of new algorithms;

2. **Validation**: in this second activity, a verification through psychophysical experiments and statistical tools is carried out to investigate whether the developed algorithms fulfill the specific intended use. This activity requires gathering relevant data from visual observations of the users and a set of statistical methods (e.g., correlation index and histogram plots) to evaluate the acquired data.

1.3 Thesis outline

This thesis is intended to give the reader all the necessary information and tools to understand, measure and enhance the contrast and the quality of digital images. The content of this work is divided into 10 chapters, where Chapters 2–5 give the scientific background, while Chapters 6–10 contain the contribution of this research work.

Chapter 2 introduces an overview of the research area related to contrast. Section 2.1 presents the definition of contrast in visual perception, colorimetry and photography. Section 2.2 gives an overview on how to measure contrast, starting from the first attempts with global measures and ending with the most important milestones of local measures. Section 2.3 introduces how to present the information of contrast to the observers.

Chapter 3 introduces an overview of the research area related to quality in imaging. Section 3.1 presents the definitions of image quality in the context of application or marketplace. Section 3.2 recalls the historical background of color difference formulae, which have led to the development of a wide variety of image quality metrics described in Section 3.3. Section 3.4 introduces how to present the information of quality to the observers and how to evaluate the performance of an image quality metric.

Chapter 4 introduces an overview of the research area related to contrast and quality enhancement. Section 4.1 introduces Spatial Color Algorithms, while Section 4.2 presents the well known Retinex theory and the derivated computational models. Section 4.3 gives an overview of Color Constancy.

Chapter 5 gives a general overview of computational methods in image processing. Section 5.1 introduces the Total Variation method followed by an overview of Swarm intelligence in Section 5.2. Section 5.3 presents how to evaluate the performance of computational methods in image processing.

Afterwards, **Chapter 6** sums up the main contribution of this work. The five included papers (three journal articles and two conference proceedings) are summarized in Sections 6.1–6.5. Following in **Chapter 7** the five included papers are discussed in context. **Chapter 8** concludes and sums up the scientific contribution of this research work followed by perspectives to new ideas for future research in **Chapter 9**. Finally, in **Chapter 10** the five included

papers are reproduced in full text.

Chapter 2

Contrast: from the origin to current issues

The word “contrast” comes from the old french word “contrester”, derived of the Italian word “contrastare”. “Contrastare” means *to withstand*, from the Latin “contra” *against* + “stare” *to stand*. “Contrast” was used as a verb to introduce opposition, in order to emphasize differences [60, 181].

As a modern word “contrast” has been re–introduced as an art term and as a noun from the 17^{th} century, and it has various meanings according to the field of use (e.g., Painting, Photography, Radiology and Telecommunication) [60, 181].

In the next section, we will provide the definitions of contrast in the field of visual perception, photography and colorimetry, as related to human vision and as of particular interest of our work. Later we will give an overview of global and local measures of contrast in Section 2.2, and lastly we will introduce how to present the contrast information generated by a measure to a user in Section 2.3.

2.1 Definitions of contrast

Most of the definitions of contrast, in particular in photography, seem to be related to dynamic range.

A generic definition of dynamic range for signals (e.g., sound and light) is [6, p. 127]:

Definition 2.1. Dynamic range is the ratio between the largest and the smallest possible values of a changeable quantity.

Before delving into a series of definitions of contrast, we introduce here a selection of definitions of dynamic range closer to digital imaging. According to:

1. Lee [149, p. 298]:

> **Definition 2.2.** Dynamic range is the ratio of the maximum to the minimum change of intensity over which the photoreceptor can operate with a detectable change in output response.

2. Peres [236, p. 430]:

> **Definition 2.3.** Dynamic range is the difference between the region of highest brightness as compared to the region of least brightness.

3. McCann and Rizzi [180, p. 379]:

> **Definition 2.4.** Dynamic range is the range of useful light described as a ratio of maximum/minimum responses to light.

According to these definitions, dynamic range is the ratio, the difference and the range. This sounds confusing. We underline to the reader that verbal definitions of an image attribute do not mandatory require to express in words a mathematical formula, but they need to express the underlying concept of an attribute (or phenomenon) in the current context. For example, while the second definition seems to be put in context of observing the grey scale (or range of greys) in a photograph, the third definition seems to be put in context of the capacity of a camera sensor.

When a number is required for dynamic range, a ratio–type equation is used [6, p. 127]. According to Allen and Triantaphillidou [8, p. 230] the dynamic range of a camera is often expressed in logarithmic units as follows:

$$DR = \log\left(\frac{I_{max}}{I_{min}}\right) \tag{2.1}$$

where I_{max} and I_{min} are the maximum and minimum light intensities that the sensor can record, respectively.

As we will see later in this chapter, the dynamic range is the basis of most contrast measures. Furthermore, the dynamic range is a feature exploited by Spatial Color Algorithms to achieve contrast and quality enhancement, as we will see in Section 4.1.

2.1.1 Visual perception

Visual perception, also known as sight (or vision), is the physiological capacity to detect light and interpret it. The organ responsible for vision is the eye (Figure 2.1), which can be considered as a spherical optical room. The incident light passes through the cornea, the outer transparent surface of the eye, and then crosses the aqueous humor, a diluted slightly saline fluid, reaching the crystalline lens via the pupil aperture. The pupil opening is controlled by the smooth muscles of the iris. The pigmented iris gives to each individual a specific eye color, determined by the concentration and distribution of melanin within it. The eye focuses on a nearby object being observed, both by increasing the strength of the lens by a relaxation in the tension of the muscles surrounding it and by maintaining convergence of the eyes [71, pp. 9–16].

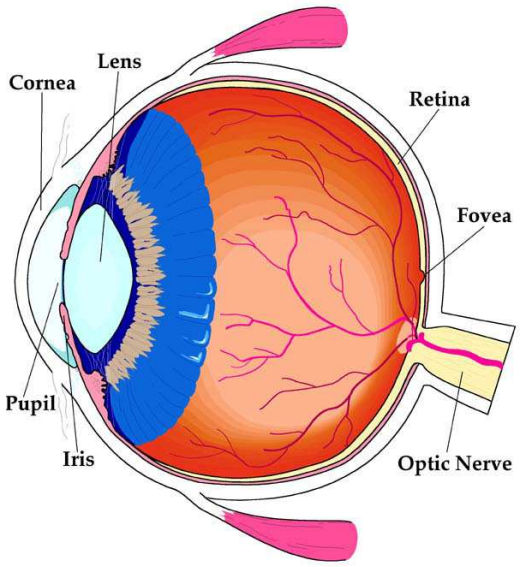


Fig. 2.1 Cross section of the human eye. Figure reproduced from Fairchild [79, p. 2].

After the crystalline lens, the light (or luminous flux) enters in the retina, a layer of tissues outside part of the nervous system. The retina covers about two–thirds of the internal surface of the eyeball and consist of different layers. From the outer to the inner [212, pp. 71–87] [88, pp. 60–68], the layers of the retina are:

1. Retinal pigment epithelium;
2. Photoreceptors;
3. Outer nuclear;
4. Outer plexiform;
5. Inner nuclear;
6. Inner plexiform;
7. Ganglion cells;
8. Nerve fibre.

The retinal pigment epithelium (RPE) is a very dark tissue constituted by a monolayer of flattened cells located just outside the retina. The RPE is attached to what is called the choroid, a layer filled with blood vessels that brings nutrients such as vitamin A, glucose and fatty acids to the photoreceptor cells [212, p. 76] [88, pp. 60–68].

The photoreceptor and the outer nuclear layers (ONL) of the retina contains rods and three sets of cones, which are the eye's color–sensors. The inner and outer segments of the rods/cones are part of the photoreceptor layer, while the cell bodies are part of the ONL divided by a membrane. The rods are responsible for night (or scotopic) vision, that is colorless vision at low levels of illumination, while the cones are responsible for daily (or photopic) vision, that is the color vision at high illumination levels [212, pp. 76–77] [88, pp. 60–68].

The cones are commonly referred to as L, M and S indicating maximum sensitivity at long, medium and short wavelengths, respectively. Sometimes the L, M and S cones are improperly called red, green and blue cones, respectively. The peaks in sensitivity are empirically measured using different approaches such as direct measures of pigment density and electrophysiological studies. The wavelength peaks occur near 565, 530 and 420 nm for L, M and S cones respectively, while for rods near 499 nm [88, pp. 66].

The photoreceptors terminals are linked to the dendrites of the bipolar and horizontal cells in the outer plexiform layer (OPL). The bodies of bipolar and horizontal cells are placed in the inner nuclear layer (INL) together with the amacrine cells. Horizontal cells are connected to both photoreceptor cells and bipolar cells and have the main role of manipulating the incoming signal in order to adjust the eye to see well under different light conditions. Bipolar cells are connected to either rods or cones and have the main role of transmitting the signal from the photoreceptors cells or from the horizontal cell to the amacrine cells [212, pp. 83–84] [88, pp. 60–68].

Bipolar cells are constituted by a receptive field consisting of two concentric regions. These two regions are known as center and surround of the receptive field, respectively, and have antagonistic effects (Figure 2.3). Thus, the role of the surround is to inhibit the signal caused by the center stimulus (center/surround mechanism). Bipolar cells can be anatomically distinguished in ON and OFF types. Simplifying, the first ones are activated with the increment of the light signal, while the second ones are activated with the decrement of the light signal [288, pp. 114–115].

The INL contains also the Müller cells, which are most active in maintaining the chemical stability of the retina environment [88, pp. 62]. The terminals of bipolar and amacrine cells are connected to the ganglion cells in the inner plexiform layer (IPL).

The bodies of ganglion cells are placed in the ganglion cells layer (GCL) and represent the final stage of neural processing in the retina. They have the same constitution of bipolar cells and thus respond to the center/surround mechanism [212, pp. 84–87]. Ganglion cells terminations constitute the nerve fibers and end in the brain visual centers, principally in the Lateral Geniculate Nucleus (LGN) and the superior colliculus [288, p. 99]. As well as ganglion cells, LGNs follow the center/surround response [212, p. 85].

The center surround mechanism is subject of this research work because it has been analytically described and involved in measures of contrast. We defer the reader to Section 2.2.2.

The anatomy of the retina and its processes are much more complex than what it has been described so far, and for this reason we address the reader to Goldstein[88], Valberg [288] and Webvision [303].

For visual perception, a definition of contrast can be found on Wikipedia [305]:

Definition 2.5. Contrast is the difference in luminance and/or color that makes an object (or its representation in an image or display) distinguishable.

In considering real world perception, the color and brightness of the light emitted or reflected by objects plays a fundamental role, and several examples can be found in practical life. For example, supposing to be driving during the night without any illumination and not be able to read traffic signs, it is possible to claim to be in a bad contrast environment.

When dealing with imaging, for example by displaying a scene on a computer monitor, additional parameters determine contrast such as the ambient light, the medium (e.g., screen) and the source of the image. Once again, it is common to claim that the image has bad (or low) contrast (Figure 2.4), if in the representation of the scene, an object is not distinguishable.

With the two aforementioned example, we have shown that contrast may depend on the

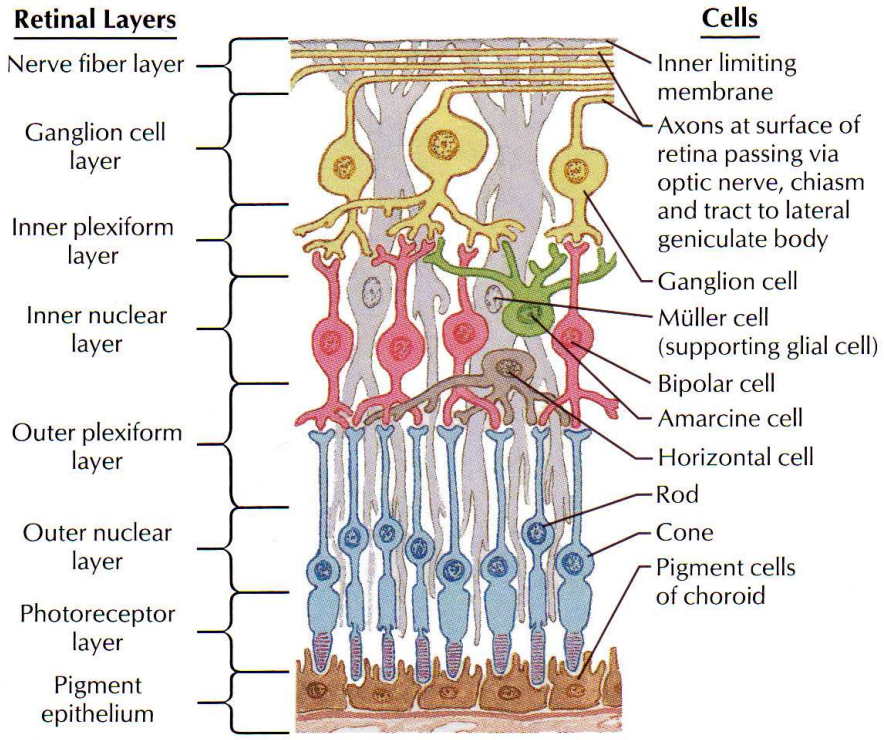


Fig. 2.2 Anatomy of the retina. Figure reproduced from http://www.corpshumain.ca/.

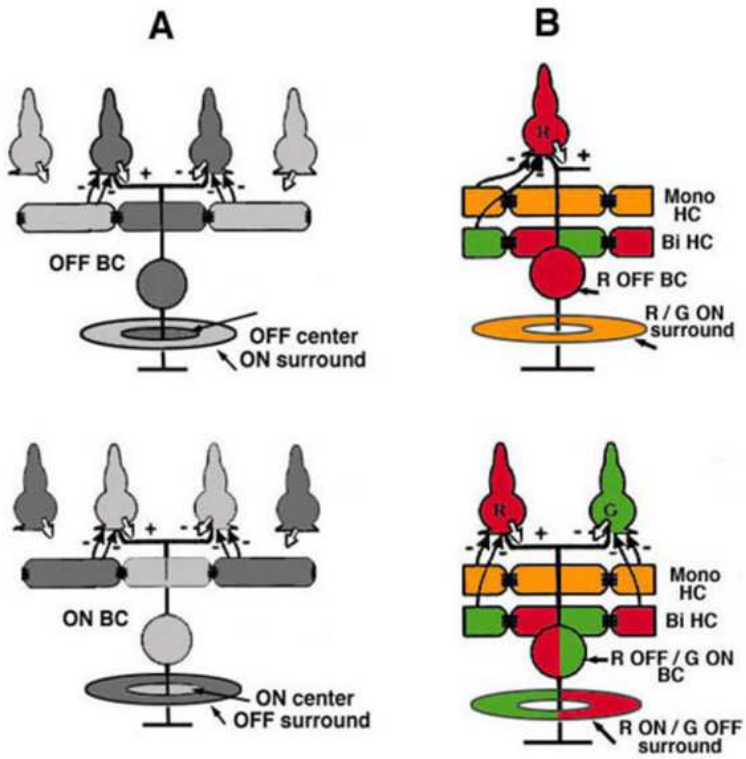


Fig. 2.3 Schematic drawing of the horizontal–cell lateral inhibition onto cones that forms the concentrically organized receptive fields of bipolar cells. In (A), a simple feedback from monophasic (L–type) horizontal cells onto cones can affect the hyperpolarizing OFF–center or depolarizing ON–center bipolar cells to provide the opponent ON– or OFF– surround respectively. In (B), the same mechanism can be used to model the generation of color opponent responses in bipolar cells. The drawing illustrates the manner in which biphasic (Bi) chromaticity horizontal cells and monophasic (Mono) luminosity horizontal cells can feed back through a single cone type to give the opponent color surround to a red OFF bipolar cell. In the lower cartoon, red cones provide excitatory input and green cone inhibitory input to produce a red OFF, green ON center double opponent bipolar cell. The red ON and green OFF surround is added by the chromatic (Bi) and luminosity (mono) horizontal cells. The model is adapted from Kamermans and Spekreijse [126]. Figure reproduced from http://webvision.med.utah.edu/book/.

task of the observer e.g., discrimination of boundaries, legibility of text and recognition of objects in the scene.

(a) (b)

Fig. 2.4 On the left an example of a low contrast image. It may be possible to understand that the depicted objects are insects, but it is impossible to distinguish that they are termites such as in the high contrast image on the right.

The definition provided by Wikipedia can be found reported either in scientific papers and in technical websites, but an official source of this definition is never cited. Fundamentally, it seems derived from the definition of contrast used in the photography field (Section 2.1.3). According to Valberg [288, p. 181], in vision:

Definition 2.6. Contrast is the physical differences in luminance and color, as well as the perception of these differences.

The world of visual perception is wide and associated to various scientific disciplines (e.g., optics, psychophysics and neurophysiology). In these disciplines, almost in all cases, contrast is defined through the Weber ratio [309, pp. 490, 567–569], which gave birth to different measures of contrast presented in Section 2.2.

However, in a recent work by Haun and Peli [99] in 2013, during a psychophysical experiment contrast was explained as:

> "the range of greyscale values you see in the image; brighter bright areas and darker dark areas indicate higher contrast."

This explanation was given to the participants in order to determine which image had higher contrast in a pairwise comparison of a set of over 500 digital images. The experiment was performed to understand how different spatial frequencies contribute to the overall perception of contrast. For this reason, the images were divided and analyzed in subbands according to the well–known (mathematical) definition of local band–limited contrast [230] presented in Section 2.2.2.

From Wikipedia definition to Peli's explanation to the participants, contrast is the difference and contrast is the range. Thus, we face the same problem we have seen for dynamic range. Verbal definitions of contrast follow the same reasoning. They are set in a context, and furthermore they are often combined with a target application and a task for the observer, e.g., object recognition.

According to the state–of–the–art, there is a large number of attributes to describe the characteristics, the content and the quality of an image [288, 222, 180, 8] Here an incomplete list for example:

> lightness, contrast, glare, color, sharpness, texture, contour, compression, size, depth, foreground, background, ...

According to McCann and Rizzi [180, p. 22], glare is a fundamental attribute and defined as:

Definition 2.7. Glare is unwanted stray light that adds to the wanted image of scene radiance at the pixel of interest.

From acquisition to display of an image, glare acts as counterforce to contrast as decreases the edge appearance. Interaction between glare and contrast is out of the perimeter of this research work, and thus only glare free imaging is taken into consideration.

2.1.2 Colorimetry

Colorimetry is the branch of color science concerned with the numerical specification of color and color appearance. The color is a sensation, which cannot be easily measured in a direct way, thus colorimetry, the science for the color measurement, is based on the two sciences, optics and psychophysics.

The term psychophysics was coined by Gustav Theodor Fechner in the 19^{th} century during his studies in attempting to establish relationships between physical and mental quantities. Psychophysics covers a wide territory in different area of research. In color science, psychophysics studies the correspondence between luminous radiations and induced color sensations [212, p. 91–92]. Description and definitions of all psychophysics is out of the perimeter of this research work, and thus we introduce here only those of interest. A detailed overview can be found in Wyszecki and Styles [309], Ohta and Robertson [208], and Oleari [212].

As aforementioned, the luminous radiations enter in the human eye, and then the signals are processed by the photoreceptors in the eye and then sent to the brain for further processing and interpretation of the signal.

This let us to introduce the following definitions:

Definition 2.8. Color stimulus is a visible radiation entering the eye and producing a sensation of color, either chromatic or achromatic.

Definition 2.9. Color sensation: the sense organs register the color stimulus, "decode" it and transform it into a neural signal, the color sensation, that is then transmitted to the brain.

Definition 2.10. Color perception: in the brain, the neural signal is organized and interpreted. Color perception involves "making sense" of color sensations.

Thus a perceived color is a psychological activity after a retinal and neuronal stimulation produced by a visible light. At current writing, color sensation and color perception are subject of debate in the state–of–the–art, and according to some (unofficial) rumors, the two definitions could be merged by future standards organizations. For the moment, in this research work, color sensation and color perception will be used as synonyms.

A perceived color is described by different attributes. We introduce here some attributes, that we will be used throughout this research work:

- Hue [212, p. 47]:

Definition 2.11. Hue is an attribute of a visual perception according to which an area appears to be similar to one of the colors: red, yellow, green and blue, or to a combination of adjacent pairs of these colors considered in a closed ring.

The most common terms representing hues are usually presented with the well–known color wheel (Figure 2.5).

- Brightness [212, p. 50]:

Definition 2.12. Brightness is an attribute of a visual perception according to which an area appears to emit, or reflect, more or less light.

- Lightness [212, p. 57]:

Definition 2.13. Lightness is the brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting.

- Colorfulness [212, p. 49]:

Definition 2.14. Colorfulness is an attribute of a visual perception according to which the perceived color of an area appears to be more or less chromatic.

- Chroma [212, p. 57]:

Definition 2.15. Chroma is the colorfulness of an area judged as a proportion of the brightness of a similarly illuminated area that appears white or highly transmitting.

- Saturation [212, p. 57]:

Definition 2.16. Saturation is the colorfulness of an area judged in proportion to its brightness.

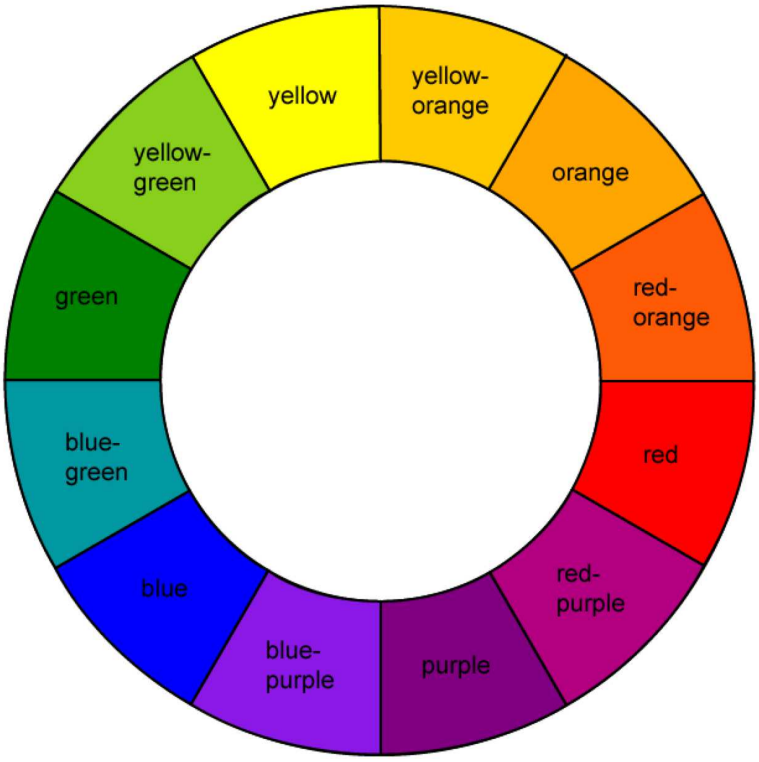


Fig. 2.5 Figure reproduced from http://adult-coloring-101.com.

The interaction among these attributes brings to important visual phenomena. Among them, the simultaneous contrast [212, pp. 58–59]:

Definition 2.17. Simultaneous contrast happens when the appearance of a color stimulus moves away from the color of the inducing stimulus produced by contiguity with other color stimuli. Generally only two colors are present and in mutual contact. Simultaneous contrast is a change in color appearance active within the lightness domain as well as within the chromatic domain (colored shadows, color induction).

Interaction of simultaneous contrast can be distinguished in [212, pp. 58–59]:

1. Simultaneous achromatic contrast (or brightness induction or brightness contrast):

Definition 2.18. Simultaneous achromatic contrast happens when a color stimulus surrounded by another color stimulus, as the luminance of the surround approaches and surpasses the luminance of the color stimulus, its appearance falls rapidly toward dark color. An achromatic unrelated color, when its visual field is modified by introducing another color stimulus, can appear white, brighter white, grey or black.

An example is shown in Figure 2.6.

2. Simultaneous chromatic contrast:

Definition 2.19. Simultaneous chromatic contrast happens when the appearance of a color stimulus moves away from the color of the inducing stimulus.

An example is shown in Figure 2.7.

3. Simultaneous saturation contrast (induced saturation):

Definition 2.20. Simultaneous saturation contrast happens when a color stimulus is surrounded by a field that is of the same color but of stronger saturation, the color stimulus appears to be tinted less vivid (less saturated). On the other hand, when a color stimulus is surrounded by a field that is of the same color but of weaker saturation, the color stimulus appears to be tinted more vivid (more saturated).

An example is shown in Figure 2.8.

4. Crispening:

Definition 2.21. Crispening happens when a very little difference between two color stimuli, which cannot be perceived if the two stimuli are seen on a very different surround stimulus, can appear enhanced if the two color stimuli have a surround that appears between them. Brightness crispening happens if the difference between the stimuli regards only the brightness. Chroma crispening happens if the difference between the stimuli regards only the chroma.

An example for brightness crispening is shown in Figure 2.9, while for chroma crispening in Figure 2.10.

Fig. 2.6 Example of simultaneous achromatic contrast. Two equal grey disks are shown on two different grey backgrounds, respectively. The disk on the left appears darker. Figure reproduced fromhttp://psychologyinrussia.com.

Another important phenomenon is the Helmholtz–Kohlrausch phenomenon [212, p. 150]:

Definition 2.22. Helmholtz–Kohlrausch phenomenon (or effect) is the change in brightness of perceived color produced by increasing the purity of a color stimulus while keeping its luminance constant.

Thus, brightness is not not only dependent on luminance, but also on the chromaticity of the stimulus.

Crispening and Helmholtz–Kohlrausch phenomena are taken into account in the OSA–UCS color system, subject of this research work (see Section 3.2).

2.1.3 Photography

Photography is the art and science of drawing with light [236, p. 27]:

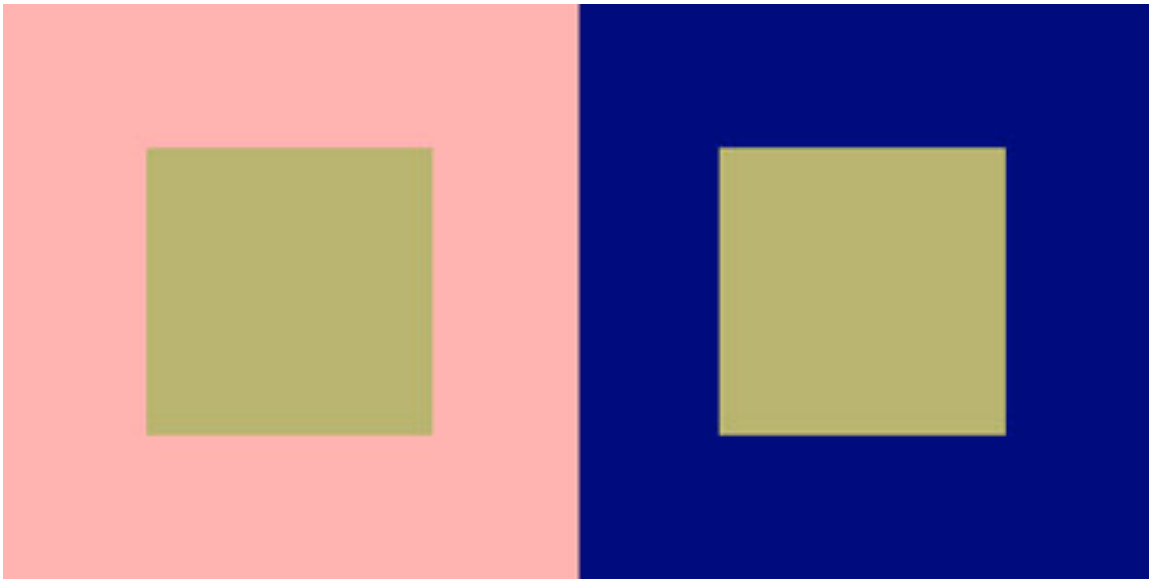
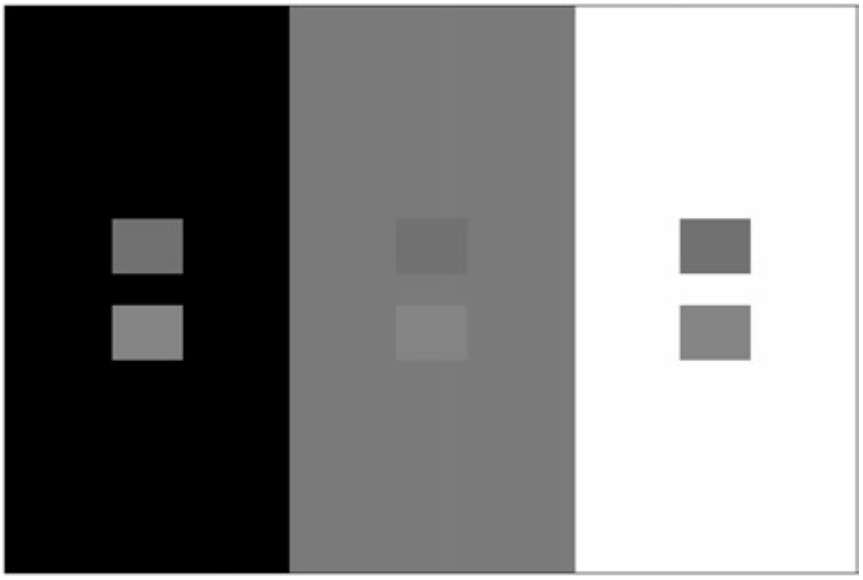


Fig. 2.7 Example of simultaneous chromatic contrast. Two equal brown disks are shown on two different backgrounds, respectively. The disk on the left (pink background) appears browner. Figure reproduced http://www.johnpaulcaponigro.com

Fig. 2.8 Two equal dark orange disks are shown on a bright orange and a grey background, respectively. The disk on the left appears less saturated due to a stronger saturation of the background. Figure reproduced http://www.johnpaulcaponigro.com.



for

Fig. 2.9 An example of brightness crispening. Two different grey squares presented on three different backgrounds. The two grey squares appear more different in brightness difference from each other against a background similar to them (grey in the middle) with respect to the black and white backgrounds. Figure reproduced from Fairchild [79, p. 115].

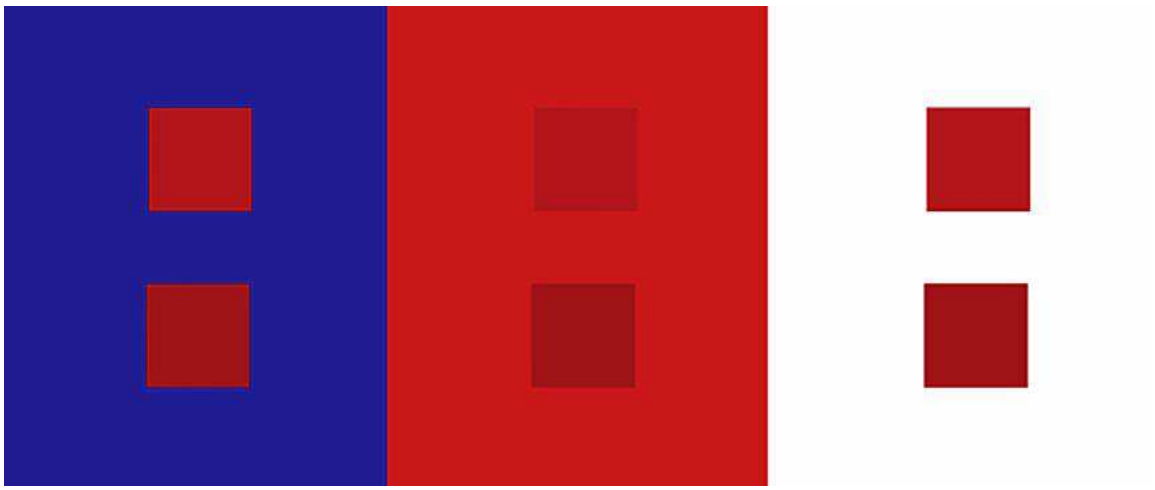


Fig. 2.10 An example of chroma crispening. Two different red squares presented on three different backgrounds. They appear more different on the background with most similar chroma (red in the middle). Figure reproduced from Allen and Triantaphillidou [8, p. 96]

> "Photography seems to record, rather than interpret, a piece of world in front of the camera and lens are often regarded simply as pieces of machinery which allow an image, a duplicate, of the world to be transferred onto film (Kuhn, 1985:26)" [307, p. 13].

In photography, contrast has several definition mostly based on the light distribution available in the picture. According to:

- Dictionary.com [60]:

Definition 2.23. Contrast is the relative difference between light and dark areas of a print or negative.

- The Merriam–Webster Unabridged Dictionary [181]:

Definition 2.24. Contrast is the degree of difference between the lightest and darkest parts of a picture.

- Osterman [236, p. 87]:

Definition 2.25. Contrast is the degree of information visible in the shadow areas.

- Allen and Triantaphillidou [8, p. 5]

Definition 2.26. Contrast of the scene is the ratio between the brightest and darkest tones in the scene and the range of possible intensity levels in between, and may be controlled or manipulated at a number of stages in the imaging chain.

As well for digital photography, contrast follows the same definitions, but according to Burien [34, p. 102] to emphasize the migration from analog to digital imaging:

Definition 2.27. Contrast is the difference in tonal value between the brightest and the darkest pixels that are recorded.

Digital cameras work much like traditional film cameras except that the image is captured on digital media rather than on analog film [236]. In digital photography, contrast is conditioned by the dynamic range that can be recorded by the sensing device (e.g., CCD and CMOS) [27, p. 208]. In other simple words, the range of values that span from black to white [54, pp. 87, 123]. The higher the dynamic range available, the higher the contrast that can be recorded.

Both in traditional photography and digital photography, contrast can be adjusted according to the needs e.g., using filters in the analog case, or manipulating the contrast operating under the direct control of the device or as a post–process in the digital case.

2.2 Measuring contrast

After giving different verbal definitions of contrast, we introduce here a selection of contrast measures defined mathematically. Contrast measures are commonly divided in two groups:

1. Global measures are based on global characteristics of the image (e.g., maximum and minimum luminance).

2. Local measures take into account local characteristics of the image or in other simple words they involve neighboring pixels in the computation.

During the last decades both global and local contrast measures have involved different area of research e.g., optics, color science and neurophysiology, and thus other way of classification are possible.

In this short survey, without a strict focus, we will also distinguish contrast measures in two categories:

1. Biological measures are focused on available physiological information of the Human Visual System and empirical data from experiments. They are generally validated involving observers in psychophysical experiments with a original reference image and a set of reproduction images. Usually, their main intent is targeting and understanding a specific aspect of visual perception.

2. Engineering measures are focused on building objective models and are usually validated through statistical analysis in comparison with other biological and/or engineering measures. Their main intent is providing practical tools for contrast measurement.

A similar distinction has been proposed by Ramirez [247].

During the '90s technologies for heavy computations were available most only at scientific research centers. Furthermore, numerical and image processing libraries were showing some limitations (e.g., only images with size power of two could be processed). A trade-off between fidelity and calculation efficiency used to be taken into account by engineering contrast measures for rather small images. Currently, computing technologies (e.g., General Purpose computing on Graphic Processing Unit and grid computing) have further processed, but with the increasing of image resolution (e.g., 4K and 8K), efficiency is still a concern for engineering contrast measures. We will see some examples in the next section.

2.2.1 Global measures

Global measures of contrast have been mainly developed during the second half of the 20^{th} century, and most of them involve only the luminance concept. In particular for commercial purposes (Figure 2.11), contrast is simply represented by a ratio of the type:

$$C = \frac{L_{max}}{L_{min}}, \tag{2.2}$$

where L_{max} is the maximum luminance and L_{min} is the minimum luminance.

This simple way of giving a quantification to contrast seems to mathematically translate the verbal definition of dynamic range. Actually, this measure has origins from the Weber definition of stimulus contrast [53], which is the most commonly used one in the context of lighting:

$$C^W = \frac{\Delta L}{L_{min}} = \frac{L_{max} - L_{min}}{L_{min}}. \tag{2.3}$$

This definition is directly related to the Weber–Fechner law:

$$\Delta S = k\frac{\Delta L}{L}, \tag{2.4}$$

where ΔS is the sensation magnitude and k is a suitable constant.

Assuming that the minimum variation of the sensation magnitude corresponds to the minimum difference of luminance perceptible, the ratio $\Delta L/L$ becomes constant. A ratio of 0.002 was found by Blackwell [21]. This procedure allows to introduce the more general concept of just–noticeable difference (JND), which is the discrimination between two stimuli that evoke

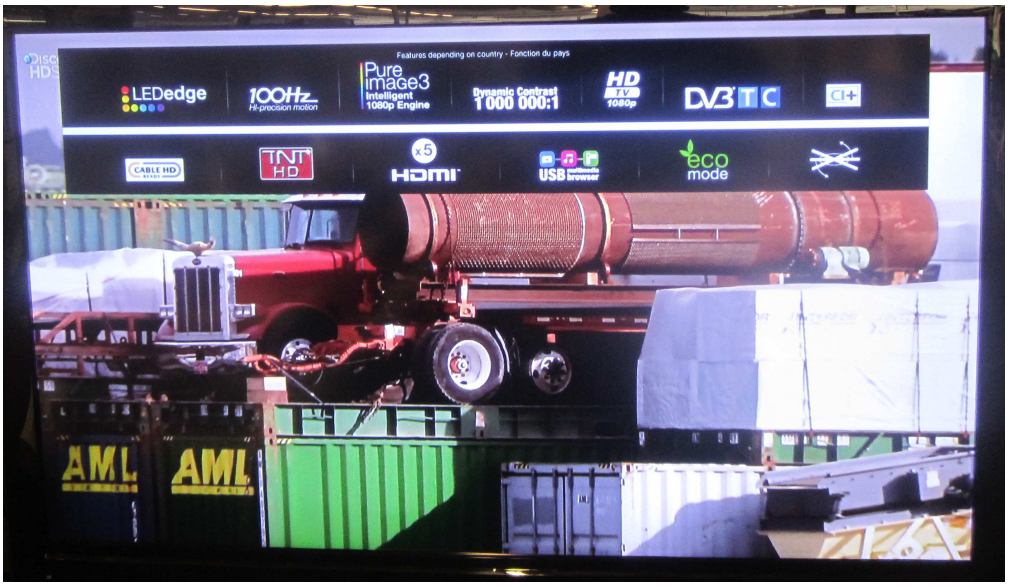


Fig. 2.11 This figure shows the technical specifications of a new generation TV (2011). The contrast is stated to be 1000000 : 1 indicating that there is factor of 10^6 between the maximum white (L_{max}) and the maximum black (L_{min}) reproducible.

only slight differences in visual perception. JNDs can be calculated on luminance, like in this example, and as well on other units of measurement (e.g., lightness and chroma).

When measuring JNDs, an in important mechanism to take into account is adaptation [212, pp. 103–106]:

Definition 2.28. Adaptation is the process by which the state of the visual system is modified by previous and present exposure to stimuli that may have various luminance values, spectral distributions and angular subtenses.

In fact, the Human Visual System (HVS) is capable of decreasing or increasing its visual sensitivity according to the level and the color of illumination [8, p. 94]. Thus, three kinds of adaptation can be distinguished:

1. Light adaptation: decreasing in visual sensitivity with increasing in luminance.
2. Dark adaptation: increasing in visual sensitivity with decreasing in luminance
3. Chromatic adaptation: approximate compensation for changes in the color of stimuli in the case of change in the illuminant (Color Constancy, see Section 4.3).

While adaptation is known to work for a uniform patch on a uniform background [149, pp. 334–347], experiments have shown that adaptation fails in complex scenes because it does not explain color constancy and appearances [178, 1, 176]. Contrast is an instantaneous response to spatial distributions of the scene content that does not require any adaptation state [180, p. 245].

Back to global measures, the very first measure, typically used in signal processing theory, in the case of sinusoids or other patterns of symmetrical deviations ranging from the maximum luminance to minimum luminance, is the Michelson [187] formula:

$$C^M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}. \tag{2.5}$$

King–Smith and Kulikowski [131], Burkhardt et al. [35] and Whittle [304] in 1975, 1984 and 1986 respectively, follow a similar approach replacing either the minimum or the maximum luminance with the average luminance of the pattern/image. These measures so far presented are also known as "physical" measures of contrast as they do not take account to observer's perception and change in illumination. [288, p. 184] [230]. Thus, these global measures assign a single physical contrast value to the whole pattern/image, but perceived contrast can vary across the pattern/image due to different spatial frequencies [101, 230]. Theoretically these measures involve a contrast range $[0, \infty]$, but current technology does not allow for such measurement to be taken.

Another relatively simple measure to calculate is the Root–Mean–Square (RMS) contrast, which is the standard deviation of luminance values in the image [219, 257]:

$$C^{RMS} = \sigma = \sqrt{\frac{\sum_x^M \sum_y^N \left(L(x,y) - \bar{L}\right)^2}{MN}}, \tag{2.6}$$

with:

$$\bar{L} = \mu = \frac{\sum_x^M \sum_y^N L(x,y)}{MN}, \tag{2.7}$$

where $L(x,y)$ is the luminance value at the pixel coordinates x, y, and M and N are the width and the height of the image, respectively. The C^{RMS} has shown to be a good predictor of random noise patterns [199] and of the relative subjective/apparent contrasts of compound grating images [283], when divided by the average luminance of the image, simply the coefficient of variation σ/μ.

The limitations of global measures discovered during the end of the 20^{th} century, where the Weber–Fechner law does not hold true for complex scenes [151], led to the development of local measures of contrast presented in the next section.

However, an alternative empirically based global measure for color images has been developed recently by Calabria and Fairchild [37]. This measure, named Single Image Perceived Contrast, is defined in the CIELAB 1976 color space presented in Section 3.2 as follows:

$$SIP_k = -1.505 + 0.131 \cdot k_C + 0.151 \cdot k_L + 666.216 \cdot k_S, \tag{2.8}$$

where k_C, k_L, and k_S are respectively the standard deviation of image chroma, lightness and high–passed lightness.

2.2.2 Local measures

Local measures of contrast have been developed to overcome some limitations and disadvantages of the global measures (Figure 2.12). In image processing and in particular for the grey level case, the luminance is associated with the intensity corresponding to each pixel, so that important changes in luminance are detected around the edges of the image. The processing of images in the visual system is believed to be local, and therefore the representation of contrast in images should be quasi–local as well.

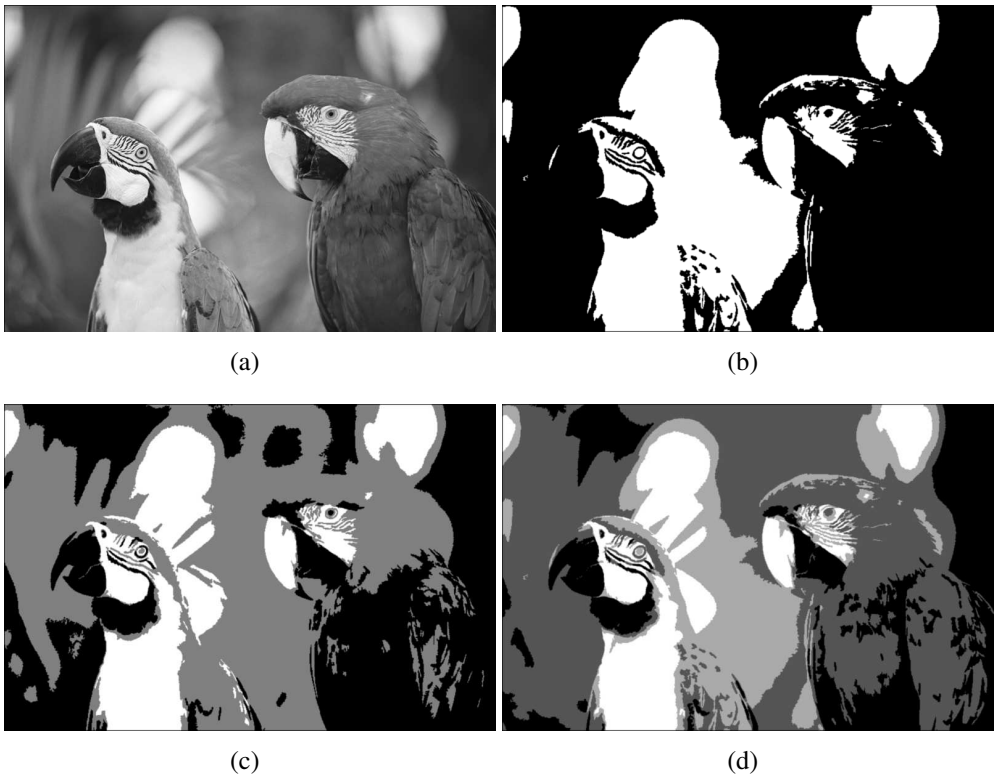
The issue of contrast of complex scenes, at different spatial frequencies, in the context of image processing and perception, has been raised and addressed explicitly by Hess et al. [101]. An important milestone was reached in 1983 when Frankle and McCann [83] followed by Adelson et al. in 1984 [2] proposed the use of the multilevel representation as an important implementation feature to mimic the HVS. This consists of a set of lowpass or bandpass copies of an image, each representing pattern information at a different scale. This data structure used to represent image information is referred as "pyramid" and shown in Figure 2.13.

Based on this approach, comes the work of Peli in 1990 [230], who proposed a local measure which performs a decomposition of the image to be analyzed into subbands using a bank of cosine–log bandpass filters, which are inspired by Gabor patches. Gabor patches are composed by of horizontal (or vertical) sinusoidal gratings in cosine phase and a two–dimensional Gaussian function [27, pp. 369–371]. The luminance distribution of a Gabor patch is defined follows:

$$L(x,y) = L_0 \left[1 + C^M \cos\left(2\pi f_0 x\right) \exp\left(-\frac{x^2 + y^2}{\sigma^2} \right) \right] \tag{2.9}$$

where L_0 is the mean luminance, C^M is the Michelson contrast (referred also as to nominal contrast), and σ is the standard deviation of the Gaussian function.

In Peli's work, contrast is then measured for each pixel, for each subband, as a function of the pixel value and the values of the pixels in the same position in the lower bands. If then the



(a) (b)

(c) (d)

Fig. 2.12 Weakness of global measures. In all four pictures, global contrast measures would typically estimate contrast by the highest and lowest luminance pixel values ($L_{max} = 100, L_{min} = 0$). The visual contrast of these four images are clearly different showing that global measures cannot adequately predict perceived contrast, and the different spatial frequencies must be taken into account.

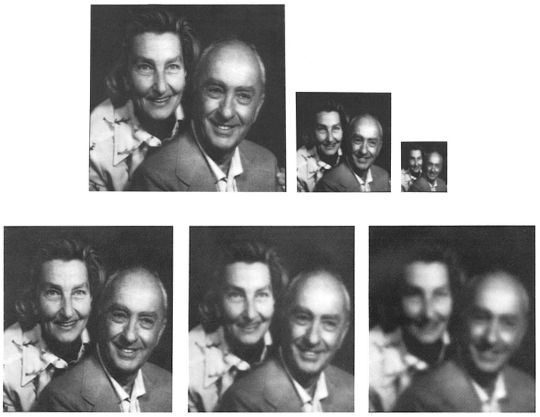


Fig. 2.13 Adelson's pyramid structure, known also as Gaussian pyramid. On the top, the original image and the sequence of two subsampled images. On the bottom, the levels are expanded to the size of the original image and the effects of the filtering are clearly apparent. Figure extracted and reproduced from the original paper by Adelson et al. [2].

contrast of each pixel is above a defined threshold (suprathreshold), the information is kept, otherwise is discarded, and the pixel is set to a luminance value equal to zero (Figure 2.14).

The threshold known as Contrast Sensitivity Threshold (CST) is the inverse of the Contrast Sensitivity Function (CSF), which is simply a simulator of the observer's perception of an image at different cycles per degree of visual angle (Figure 2.15). The CSFs are usually calculated through observations and sometimes corrected through simulations [38, 230, 234, 232]. There are several CSFs available in the literature (Figure 2.16), and in their work Peli et al. [230, 234] propose their own CSF from where they retrieve the contrast sensitivity threshold.

CSFs have been measured not only for luminance, but also for chrominance [93, 201, 200]. Following Movshon and Kiorpes [200] studies, CSFs can be generally defined as follows [79, p. 345]:

$$CSF_{lum}(f) = a \cdot f^{c} \cdot e^{-b \cdot f}, \tag{2.10a}$$

$$CSF_{chrom}(f) = a_1 \cdot e^{-b_1 \cdot f^{c_1}} + a_2 \cdot e^{-b_2 \cdot f^{c_2}}. \tag{2.10b}$$

Johnson and Fairchild propose the following values [125]:

$$a = 75, b = 0.2, c = 0.8, \tag{2.11a}$$

$$a_1 = 109.14130, b_1 = -0.00038, c_1 = 3.42436, a_2 = 93.59711, b_2 = -0.00367, c_2 = 2.16771, \tag{2.11b}$$

$$a_1 = 7.032845, b_1 = -0.000004, c_1 = 4.258205, a_2 = 40.690950, b_2 = -0.103909, c_2 = 1.648658. \tag{2.11c}$$

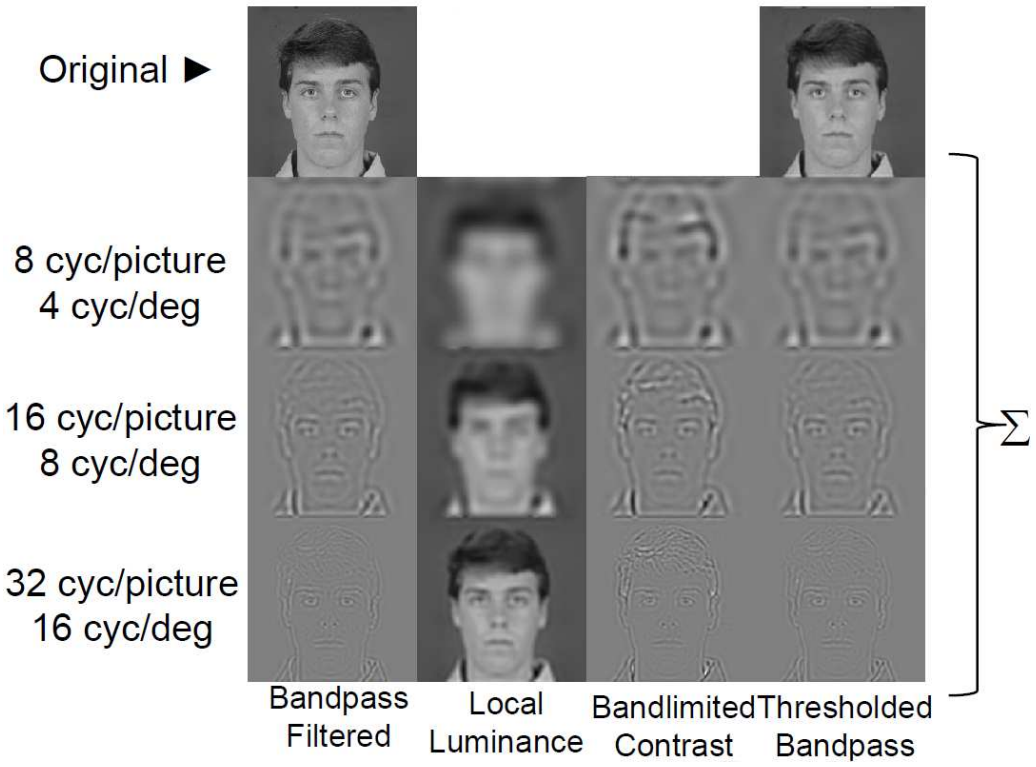


Fig. 2.14 In the first column the original image and the relative bandpass filtered images (frequency bands) obtained with a cosine–log bandpass filter centered at frequency of 2^i cycles/picture. In the second column the corresponding local luminance mean image which are lowpass filtered version of the image containing all the energy below the band. In the third column the band–limited contrast images, which are the ratio between the relative bandpass filtered and the local luminance mean (1st column/2nd column). In the fourth column the thresholded bandpass images obtained keeping the information above a defined threshold and discarding the information below or equal the same threshold. The reconstructed image is shown on top of the fourth column and is the sum of all the thresholded bandpass images. The original and reconstructed image should appear indistinguishable if presented at distance of two meters with size of 512×512. Figure provided and reproduced by courtesy of Eli Peli [233].

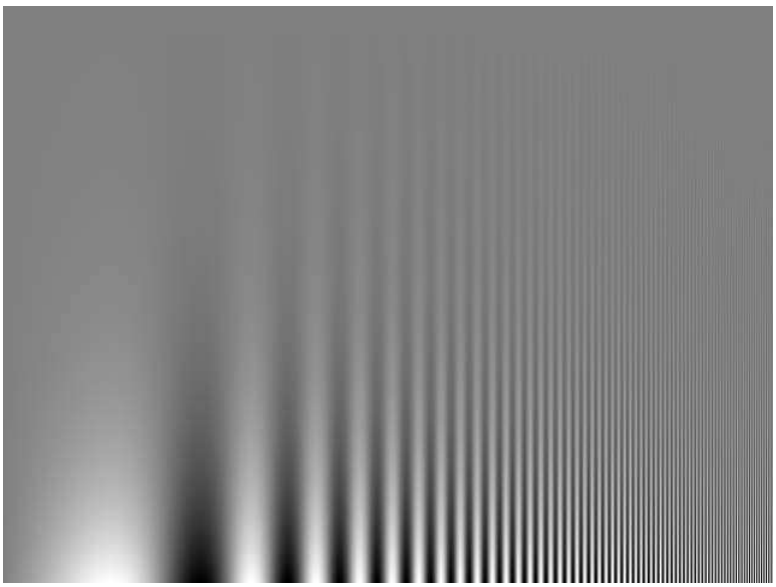


Fig. 2.15 Campbell–Robson CSF Chart (1968). The luminance is modulated sinusoidally along the horizontal dimension. The frequency of modulation (spatial frequency in Cycles Per Degree of visual angle or CPD) increases logarithmically and as well the contrast. The bars appear taller in the middle of the image than at the sides, depending on the viewing distance, and thus showing that the visual system is more sensitive to a specific range of frequencies [38]. Figure reproduced from neurovision.berkeley.edu.

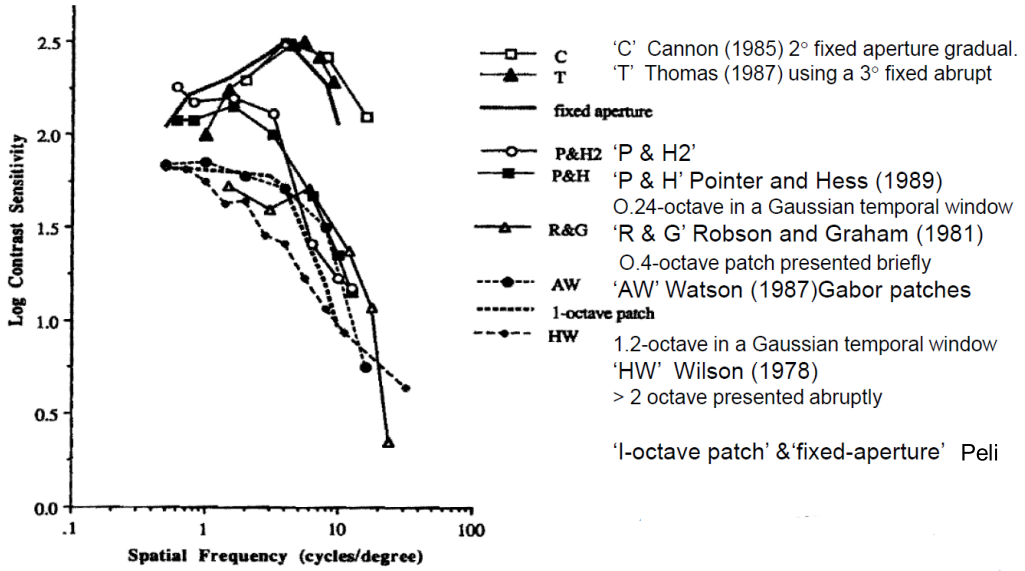


Fig. 2.16 Examples of CFS available in literature for greyscale images. Figure provided and reproduced by courtesy of Eli Peli [233].

The resulting luminance and chrominance CSFs, found success in color appearance models and in image quality metrics presented in Section 3.3.

Back to contrast measures, today the pioneer work of Peli on greyscale images is one of the most important and recognized by the scientific community (over 800 citations in correlated articles [214, 50]) and unofficial resources (e.g., personal websites).

Before introducing a selection of the state–of–the–art biologically and/or engineering inspired measures, we recall the mathematical definition of Peli's local band–limited contrast [230]:

$$C_i^{Peli}(x,y) = \frac{a_i(x,y)}{l_i(x,y)}, \tag{2.12}$$

where $a_i(x,y)$ is the bandpass filtered image at level i of a filter bank, and $l_i(x,y)$ is the corresponding lowpass filtered image at level i containing all energy below the band (or local luminance mean), and x,y indicate the spatial coordinates of the pixel.

The bandpass filtered image $a_i(x,y)$ at level i is calculated as:

$$a_i(x,y) = \psi_i * I(x,y), \tag{2.13}$$

where ψ_i is a chosen filter bank at level i, and $I(x,y)$ is the input image.

In Peli's work ψ_i is a *cosine–log* filter inspired from Gabor patches, centered at frequency of 2^i cycles/picture, and expressed as:

$$\psi_i = G_i(r) = \frac{1}{2}\left(1 + cos\left(\pi log_2 r - \pi i\right)\right). \tag{2.14}$$

As the bandpass filter of a pyramid can be calculated as the difference of two neighboring lowpass filters, Equation 2.13 can be rewritten as:

$$a_i(x,y) = l_i(x,y) - l_{i+1}(x,y), \tag{2.15}$$

and thus:

$$C_i^{Peli}(x,y) = \frac{l_i(x,y) - l_{i+1}(x,y)}{l_{i+1}(x,y)} = \frac{l_i(x,y)}{l_{i+1}(x,y)} - 1. \tag{2.16}$$

The first improvement of Peli's measure date back to 1993 by Lubin [154] with the following proposal:

$$C_i^{L93}(x,y) = \frac{a_i(x,y)}{l_{i+2}(x,y)} = \frac{l_i(x,y) - l_{i+1}(x,y)}{l_{i+2}(x,y)}. \tag{2.17}$$

The difference is clear in the denominator, where the lowpass filter is moved down of one level, using two octave bands instead of one octave.

In a follow–up, Lubin [155] proposed a further improvement:

$$C_{i,k}^{L95}(x,y) = \frac{[l_i(x,y) - l_{i+1}(x,y)] * \vartheta_{i,k}}{l_{i+2}(x,y)}, \tag{2.18}$$

where $\vartheta_{i,k}$ is analytic orientation–selective filter. Lubin formula has found good agreement with observers in psychophysical experiments for predicting the apparent contrast of Gabor patches [231, 155].

In 1998, Ahumada and Beard [3] proposed a method for measuring contrast which can be described as follows:

$$C^{AB}(x,y) = \frac{b(x,y)}{m(x,y)} - 1. \tag{2.19}$$

with:

$$b(x,y) = I(x,y) * g_b(x,y), \tag{2.20a}$$
$$m(x,y) = b(x,y) * g_l(x,y), \tag{2.20b}$$

where $b(x,y)$ is the blurred image, $m(x,y)$ is the local luminance image, and $g_b(x,y)$ and $g_l(x,y)$ are two distinct Gaussian lowpass filters. It is possible to notice a similarity between Equations 2.19 and 2.16. However, with respect to Peli, the measure presented by Ahumada and Beard does not implement the multilevel feature, and the local luminance image is the result of different type of filters.

In 2001, Iordache et al. [111] proposed a measure based on the previous studies of Moon and Spencer [192]:

$$C^{IBL}(x,y) = \frac{I(x,y)}{b_s(x,y)} \tag{2.21}$$

where $b_s(x,y)$ is the average of the 8–neighboring luminances calculated as:

$$b_s(x,y) = \sum_{n8} |I(x,y) - \alpha_n I(x_n,y_n)| \tag{2.22}$$

with:

$$\alpha_n = \begin{bmatrix} \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{8} & 1 & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \end{bmatrix}. \tag{2.23}$$

The measures introduced so far can be clearly categorized as biological. According to the authors, these measures have a good fidelity with observers perception of contrast.

Concurrently to the studies of contrast in digital images, also neurophysiological research

made further progress in understanding and modeling retinal functions in the field of visual adaptation and contrast sensitivity. An important milestone is given by the center/surround mechanism, which is based on the idea that the center stimulus is influenced by the surround. It has been demonstrated that this mechanism is present in neurophysiological processes (see Section 2.1.1) and contributes to the phenomena of lightness and color constancy in the HVS perception [127]. The center/surround mechanism can be described analytically by the Difference of Gaussians (DOG) or Laplacian of Gaussians (LoG) [287]. Both have been used successfully in many studies to describe the receptive fields and responses of mammalian retinal ganglion cells and Lateral Geniculate Nucleus (LGN) neurones.

In 1981, Ullman demonstrated that the LoG can be used to model the output of the retinal operation corresponding to sharp intensity changes in the image [287]. The DOG model, which approximates the LoG at different scales has anyway found more success in studies on biophysics of retinal receptors [265].

In 2000, Tadmor and Tolhurst [280] based their analysis of contrast on modifying and adapting the DOG model. The conventional DOG model [255, 77] assumes that the response of a neuron depends uniquely on the local luminance difference (ΔL) between the the receptive–field center and surround. Thus, the response of a retinal ganglion cell or an LGN neuron can be calculated as a subtraction between the center output and the surround output:

$$DOG = R_c - R_s, \tag{2.24}$$

where R_c and R_s are the response of the center and surround component of the receptive–field respectively and defined as follows:

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c} \sum_{j=y-3r_c}^{j=y+3r_c} Center(i-x, j-y) I(i,j), \tag{2.25a}$$

$$R_s(x,y) = \sum_{i=x-3r_s}^{i=x+3r_s} \sum_{j=y-3r_s}^{j=y+3r_s} Surround(i-x, j-y) I(i,j), \tag{2.25b}$$

with:

$$Center(x,y) = exp\left[-\left(\frac{x}{r_c}\right)^2 - \left(\frac{y}{r_c}\right)^2\right], \tag{2.26a}$$

$$Surround(x,y) = 0.85\left(\frac{r_c}{r_s}\right)^2 exp\left[-\left(\frac{x}{r_s}\right)^2 - \left(\frac{y}{r_s}\right)^2\right], \tag{2.26b}$$

where r_c and r_s are the radiuses of the two Gaussians respectively.

Further studies have shown that, after the light adaptation process, the gain of the ganglion cells of the retina and the neurons depends on the average local luminance $\overline{L}$. As a consequence, the DOG model should be modified by a division by the local mean luminance in order to model the response based on the contrast stimulus $\Delta L/\overline{L}$ [265]. Unfortunately, the physiological information necessary to modify the DOG model accurately are still not available. For this reason, they tested three hypothetical different outputs for contrast as follows:

$$C^{TTC}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y)} \text{ (center–only scheme)}, \tag{2.27a}$$

$$C^{TTS}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_s(x,y)} \text{ (surround–only scheme)}, \tag{2.27b}$$

$$C^{TTCS}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)} \text{ (center–plus–surround scheme)}, \tag{2.27c}$$

Notice the similarity between Equations 2.27a–2.27b and the Weber definition of stimulus contrast (Equation 2.3), and the similarity between Equation 2.27c and the Michelson definition of grating contrast (Equation 2.5). After analysis they propose the scheme in Equation 2.27c for measuring contrast. Although the measure of Tadmor and Tolhurst tries to keep a biological approach, a tradeoff for efficiency is performed calculating the overall contrast of the image as the average local contrast of 1000 random pixel locations. The pixel locations must be taken assuring that the center and surround masks do not exceed the edges of the image.

Another approach for measuring local contrast is presented by Ferraro and Boccignone [24, 82], considering any image as an isolated thermodynamical system. The presentation of this measure requires the introduction of variational methods, and for this reason the reader is deferred to Section 5.1.

In 2004, Rizzi et al. [251] developed a simple and efficient measure able to estimate global and local components of contrast based on two principles:

1. Analyzing the image at various frequency levels.

2. Weighted difference among the 8–neighboring pixels.

Specifically, this measure creates a pyramidal structure in the CIELAB color space [46] (Section 3.2) by halving the image at each iteration. For each level, the local contrast is calculated for each pixel by taking the average difference between the lightness channel value

of the pixel and the surrounding eight pixels:

$$C_i^{RAMM}(x,y) = \sum_{n \in 8} \alpha_n \left| I(x,y) - I(x_n,y_n) \right|, \tag{2.28}$$

with:

$$\alpha_n = \frac{1}{4+2\sqrt{2}} \begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & 1 & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix}. \tag{2.29}$$

The final overall measure is then retrieved with a recombination of the average contrast of each level.

Although some similarity with Iordache et al.'s [111] measure, the Rizzi et al.'s measure clearly shifts the tradeoff towards efficiency, completely sacrificing the fidelity of the measurement with respect to the biological approaches. Four main reasons have invited this contrast measure to be part of this research work [251]:

1. The use of an alternative multilevel approach.
2. Being tested on color images.
3. Objective and subjective tests show an almost linear behavior across the artificial variation of contrast.
4. Computational complexity lower than alternative local methods, keeping a comparable level of correctness in the contrast estimation.

Although some of these measures have been tested on color images, from Peli to Rizzi et al., they do not take into account chromatic information and rely their calculation only on luminance or lightness information. Recent studies [37, 228] have shown that chromatic information plays a fundamental role in contrast perception.

2.3 Quantifying and evaluating contrast

After a short selection of contrast measures, we introduce in this section a brief overview on how to present the contrast information generated by a measure to a user. This is an important issue since contrast is one of the most important attributes of perception for evaluating the quality of an image [222]. Methodologies of evaluation of contrast are mainly inspired from psychophysical experiments in image quality, and for this reason we defer the reader to Section 3.4 for further details.

Deciding how to represent contrast in one number or a series of numbers and using a global or local measure, is generally dependent on the target application. Although we have seen that global measures are inadequate for measuring the perceived contrast of complex scenes (Figure 2.12), they are still commonly used for multimedia device specifications (Figure 2.11). Global measures generally represent contrast as single number and dimensionless unit of measurement.

Also local measures keep the intent of having contrast as a dimensionless unit of measurement, but different ways of representing contrast are available. The representation obtained is usually the result of a process called pooling. Local measures generally produce a contrast map, which in this research work is defined as follows:

Definition 2.29. Contrast map is any array or matrix, where the values can be attributed to visual contrast information.

According to [61, 182] the noun "pool" and the verb to "to pool" assume different meanings with respect to the field of use. Often, the word pool is combined with other terms to indicate aggregation and accumulation of properties, features and resources e.g. swimming pool, car pool and pool of applicants.

In the field of digital imaging, different pooling strategies are available, in particular in the branch of image quality. Pooling studies for image quality will be thoroughly presented in Section 3.4.

In order to help the reader in the comprehension of this research work, we introduce here a definition of contrast pool as follows:

Definition 2.30. Contrast pool is any type of pixel contrast accumulation to reduce a contrast map to a manageable set of values.

Thus, the purpose of pooling is to reduce a large number of values to a lower number of values. This is generally achieved performing statistical methods (e.g., average and RMSE) on the contrast map.

Both contrast map and pooling can be performed, not only using specific mathematical methods, but also in combination with dedicated devices, e.g., in 1999 Reinagel and Zador [249] proposed a measure using the Root–Mean–Square contrast (See Equation 2.6) after extracting gaze information during the viewing process of the observers (Figure 2.17).

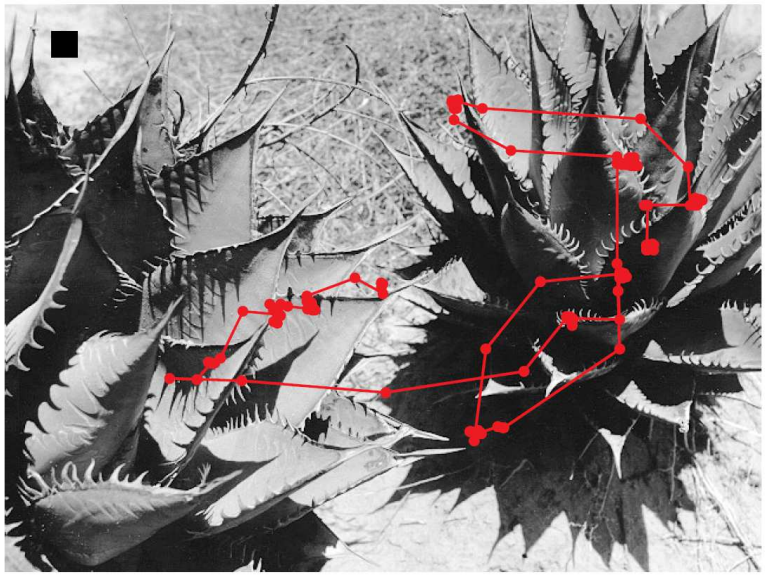


Fig. 2.17 A representative image with eye positions of one subject in the work of Reinagel and Zador. The red circles indicate the position of the center of gaze recorded at 20 *ms* intervals. The calibration square in the upper left corner is 1×1 degrees of visual angle. Figure extracted and reproduced from the original paper Reinagel and Zador [249].

Gaze information are of interest of this research work, with the main and simple idea to use the gaze map as weighting map. Given a contrast map, each pixel can be weighted according to the value retrieved with the gaze map (pixelwise multiplication).

Performing one or more pooling methods leads to several degrees of reduction of a contrast map with different advantages and disadvantages:

1. One number per pixel (or no pooling) has the advantage of keeping the full locality of the information, which can be further accumulated into various measures such as regions or subpixels [149, pp. 412]. Furthermore, pixelwise operations and modifications can be easily performed. The disadvantage is to be image size dependent and require a heavy memory consumption.

2. One number per level (or scale) has the advantage of linking the measure of contrast with the different frequency scales, which can be recombined a posteriori with different weights. The drawback is to be image size dependent and require a vector to store the contribution of each level.

3. One number per color channel is valid only for color images and has the advantage of keeping chromatic information separated and to determine whether a channel has dom-

inance with respect to others. It has the clear disadvantage to be color space dependent (e.g., lightness/chromaticity in CIELAB).

4. One number per image has the advantage of being a compact measure and easy to handle for several applications e.g., can be easily used for comparison in perceptual experiments, where observers usually provide a single assessment for each image. Unfortunately, it has the disadvantage to lose the ability of discriminating various type of images that usually give rise to different contrast perceptions (e.g. greyscale vs color and computer graphics vs natural images). Furthermore, the importance of each level and/or each color channel is lost as weights must be chosen a priori to aggregate all the components into a single number.

Once the degree of reduction is decided, an evaluation task performed by observers is often required. Also on the observer side, the degree of reduction presents advantages and disadvantages for the observer task.

One number per pixel is unsuitable for the observer, because a judgment for each pixel is required by the user. Most of the biological contrast measures rely on this degree of detail. In this case, the process of evaluation is often facilitated comparing the filtered image with a reference image [230, 232, 235, 99]. Thus, the evaluation is carried out accounting for the global impression of the image, while preserving the local information. This approach requires to take into account masking phenomena. An example is shown in Peli's 1990 work in Figure 2.14. The contrast image shown on the top–right is compared with the original reference image on the top–left. This degree of reduction has found application in image compression and assessment of distortions in images and video [156].

One number per level requires observers expressing a judgment per frequency content. In this case, the process of evaluation is facilitated presenting to the user all the filtered levels of the pyramid. This degree of reduction has found application in object recognition [156].

Likewise a number per channel requires observers expressing their judgment per each channel. This degree of reduction is commonly used to determine an unwanted color cast [8, p. 126].

One number per image leads to an easy task for the observers and is the most common in psychophysical experiments. Engineering contrast measures almost always rely on one number per image, as they attempt to estimate the perceived contrast of a single image (no reference is available). This type of reduction has found application in the quality evaluation of imaging systems [8, p. 126].

2.4 Summary

The main purpose of the material presented in this chapter was to give the reader an insight on contrast measures. First we introduced the definition of dynamic range, followed by the verbal definitions of contrast in visual perception, colorimetry and photography. Afterwards, we turned focus to global and local contrast measures introducing different concepts such as just–noticeable difference (JND), Contrast Sensitivity Function (CSF), multilevel representation of an image and Difference of Gaussians (DOG).

As currently there is no officially recognized (or standard) local measure of contrast we paid particular attention on Peli, Rizzi et al. and, Tadmor and Tolhurst measures. These measures will be subject of comparison for the local measure proposal of this research work.

As last, we introduced the concepts of contrast map and contrast pooling.

Chapter 3

Quality issues and image metrics

The word "quality" comes from the old French word "qualité" around 12^{th} century, from the Latin "qualitas", coined by Cicero to translate the Greek "poiotes" [62, 183]. In 14^{th} century "quality" assumes the meaning *degree of goodness*. Adjectives such as excellent, high, satisfactory, acceptable, sufficient, adequate, poor, low, bad, etc., are often applied to this word [62, 183]. The word "quality" can be cognated also with the Latin word "quale" meaning *a property as it is experienced as distinct from any source it might have in a physical object* or in other words *things experienced* [184].

In the next section, we will provide definitions of image quality followed by an overview of color difference formulae and of image quality metrics in Sections 3.2 and 3.3, respectively. Finally, in Section 3.4, we will introduce how to present quality information to a user and how it can be evaluated.

Between 2009 and 2012 Pedersen and Hardeberg performed an exhaustive research on definition and classification of color difference formulae, image quality metrics and their evaluation [225, 221, 226]. Thus, the structure and content of this chapter is inspired to their work and our collaboration for contribution [274].

3.1 Definitions of image quality

Like in the case of contrast, several definitions occur for describing the degree of goodness of an image, or in other words to define Image Quality. The term Image Quality, earliest known mentioning at the beginning of the 16^{th} century with the invention of optical instruments [75], has become more and more popular with the introduction of digital media (e.g., cameras, televisions and printers).

In the current literature, we can find several definitions of Image Quality. We report here a small selection of them, quoted as in the original work:

- Yendrikhovskij 1998 [310]:

Definition 3.1. Image Quality is the degree of correspondence between the visual representation and the memory of a natural scene.

- Janssen 1999 [120]:

Definition 3.2. Image Quality is (defined) in terms of two components, usefulness, that is, the precision of the internal representation of the image, and naturalness, that is, the degree of match between the internal representation of the image and representations stored in memory.

- Engeldrum 1999 [75]:

Definition 3.3. Image Quality is the integrated set of perceptions of the overall degree of excellence of the image.

- Fairchild 2002 [78]:

Definition 3.4. Image Quality is the perceptible visual differences from some ideal and the magnitude of such differences.

- Keelan 2002 [128, p. 8], adopted by the International Organization for Standardization (ISO) in 2005 [112]:

Definition 3.5. Image Quality is the impression of the overall merit or excellence of an image, as perceived by an observer neither associated with the act of photography, nor closely involved with the subject matter depicted.

- International Imaging Industry Association (I3A) 2007 [108]:

Definition 3.6. Image Quality is the perceptually weighted combination of all significant attributes of an image when considered in its marketplace or application.

As we can notice, these definitions are quite different and involve different factors such as perceptual attributes, observer memory and customer preference. Yendrikhovskij's definition involves observer memory, and thus a sort "mental recollection" of an earlier perceived scene. Janssen involves two perceptual attributes, naturalness and usefulness. Natural images are very small subset of the set of all possible images, and it is not clear whether very useful but synthetic images (e.g., computer generated graphics) can be subject of image quality assessment. Engeldrum does not specify any perceptual attribute and the quality of the image is judged by the observer in terms of how pleasant it looks like. According to Engeldrum its definition is suitable for the evaluation of commercial imaging systems or devices from non–experts users in a beauty contest. Fairchild provides a definition, where the quality of an image is a comparison between the reproduced scene and a sort of ideal scene in mind. Keelan's definition follows the one proposed by Janssen, but with the requirement that the observer should be far apart from the imaging industry and not be involved with the subject of the image. Although the definition from Keelan was adopted by the International Organization for Standardization in 2005, later in 2007 the International Imaging Industry Association proposed another definition remarking the importance of the target application domain and the expected use of the image. As a consequence of multiple definitions and disagreement among standard organizations, a universal definition of image quality is still not available.

Earlier in 2004, meanwhile waiting a standard or a widely accepted formal definition Engeldrum [76] attempted to make some order distinguishing two major categories of image quality models:

1. Detection/Recognition: the models belonging to this category usually define a specific set of parameters for the detection and recognition of various objects, features or elements in the image content. Most of the medical and security imaging systems fall in this category.

2. Beauty contest: the models belonging to this category are divided into two subcategories:

   (a) Standard or reference: a reference or standard imaging system is defined by characteristics. The standard or reference image is considered the upper bound in terms of image quality that the system can deliver. Thus, subsequent processes can lead

to a decrease in the quality of the image. The models belonging to this class attempt to measure the deviation (or "distance") of the so called reproduction image from the reference image. This distance can be calculated on physical image parameters (e.g., spatial resolution and dynamic range) and perceptual attributes (e.g., brightness and contrast). Observers are involved in psychophysical experiments expressing comparative judgments.

(b) No standard or reference: without any reference, these models set as a goal for the imaging system the achievement of a set of predefined physical image parameters and perceptual attributes. These systems are often evaluated by observers expressing their personal preferences, and thus pleasantness comes into play.

Reconsidering just the term quality, the following definitions can be found according to organization providing standards:

- ISO9001:2008 [113]:

Definition 3.7. Quality is the features of a product (or service) which are required by a customer.

- ANSI/IEEE 2007 [9]:

Definition 3.8. Quality is the degree to which a component, system or process meets specified requirements and/or user/customer needs and expectations.

- Qualinet 2012 [146]:

Definition 3.9. Quality is the outcome of an individual's comparison and judgment process. It includes perception, reflection about the perception, and the description of the outcome.

These three last definitions obviously do not describe the quality of an image, but focusing on the I3A and Qualinet definitions, and as remarked by Engeldrum [75] and Lundström [158], customers and market are in most cases the final recipient of an image. Thus, image quality

assessment through subjective evaluation becomes a precise method and the true benchmark of quality [295].

Unfortunately, subjective evaluation is not always feasible because resource demanding and time consuming when dealing with a large number of images. Furthermore, during the evaluation process, the interference of external factors to observers (e.g., change of the lighting environment, tiredness and stress) may lead to inconsistent results [295, 119, 267]. For these reasons, objective evaluation methods such as image quality metrics (or algorithms) have been developed with the intent to have lower cost, be faster, less resource demanding, and to produce consistent results [295, 267]. From now on, in this research work we will use only the word metric. The interest of this research work are full–reference image quality metrics belonging to the standard or reference category proposed by Engeldrum, and later presented in Section 3.3.

Before giving more details about the state–of–the–art of image quality metrics, we recall in the next section the color difference milestones as many image quality metrics have been developed extending the color difference formulae to the color of a complex image.

3.2 Color difference milestones

The history of color vision theory, based on a set of assumptions or postulates about the Human Visual System, is now almost three centuries long. At the beginning of the 18^{th} century, Newton was the first to explain color theory from his experiments with light [204]. In 1802, Young suggested a three–fold model of color perception [311], followed by an extension quantitatively expressed by Helmholtz later in 1894 [290]. This model, well known as trichromat theory, was developed on the basis of additive and associative laws of color mixture developed by Grassman [92]. The trichromatic theory postulates that the retina comprises three types of photoreceptors (cones) which are sensitive to short (blue), middle (green) and long (red) wavelengths. Thus, all colors are characterized by the degree of response of these photoreceptors when the light is striking the retina, e.g., the sensation of yellow is generated by the simultaneous response of the red and green photoreceptors.

The trichromatic theory is empirically based, and along with the model, is necessary to measure color stimuli from the observers to retrieve the so called Color Matching Functions (CMFs). CMFs are the basic numerical descriptors for the development of a color system. The need of defining a standard observer and colorimetric system led the Commission Internationale de l' Éclairage to recommend the CIE Standard Colorimetric 1931 System (CIE '31 System) based on the experimental color matching data of Guild [95] and Wright [308], using particular real primaries R,G,B to a set of imaginary primary stimuli known as X,Y,Z or

tristimulus values. The representation of color stimuli as three–dimensional vectors is informative, but sometimes practically inconvenient. For this reason a two–dimensional representation, known as chromaticity diagram, is usually preferred to plot colors. The chromaticity coordinates x and y of a color are established at the intersection of the color vector X,Y,Z with the unit plane ($X+Y+Z=1$) as follows [208, pp. 75–76]:

$$x = \frac{X}{X+Y+Z}, \tag{3.1a}$$

$$y = \frac{Y}{X+Y+Z}. \tag{3.1b}$$

Later, the CIE proposed the CIE 1964 Standard Colorimetric System (CIE '64 System) as an extension of the XYZ color system established in 1931, which is applicable for a viewing angle exceeding four degrees. The system is based on color matching experiments performed with a visual field of 10 degrees. The differences between the two degrees and the 10 degrees CMFs are not substantial, but remain meaningful in some applications [208, pp. 72–74, 249–254]. However, the CIE XYZ space is perceptually non–uniform since equal perceptual differences between colors do not correspond to equal distances in the tristimulus space. This lead researchers to attempt finding and developing uniform color spaces.

In 1942, MacAdam [166] conducted a detailed experiment in order to investigate the non–uniformity of the chromaticity diagram. He performed a color matching experiment for 25 different test colors with a 2 degree field and constant luminance. The results of the experiment are known as MacAdam ellipses (Figure 3.1), which represent the standard deviations of color matching.

An important mile stone was reached in 1976 with the CIELAB color system [46, 49]. The CIELAB color space is a three–dimensional color space with approximately uniform scales, spanned by the rectangular coordinates (Figure 3.2). These coordinates known as L^*,a^*,b^* represent the lightness, the red–green opponency and the yellow–blue opponency, respectively.

CIELAB is a metric space in which the color difference between two colors ΔE^*_{ab} is represented by their Euclidean distance:

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}, \tag{3.2}$$

where $\Delta L^* = L^*_2 - L^*_1$, $\Delta a^* = a^*_2 - a^*_1$, $\Delta b^* = b^*_2 - b^*_1$, 1 denotes the reference and 2 the sample.

However, for color difference values in correlation with visual sensation can sometimes be

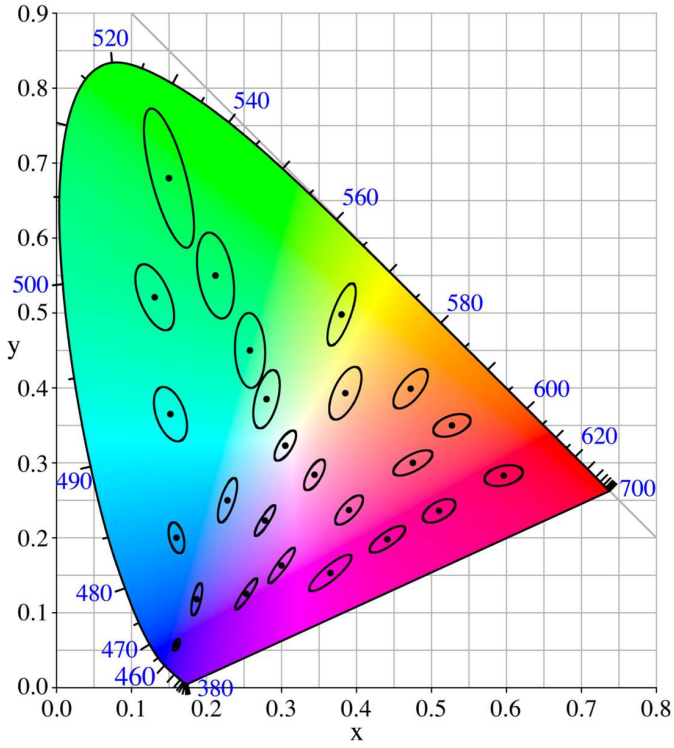


Fig. 3.1 MacAdam ellipses (enlarged 10 times) on CIE' 31 chromaticity diagram. Figure reproduced https://en.wikipedia.org/wiki/MacAdam_ellipse

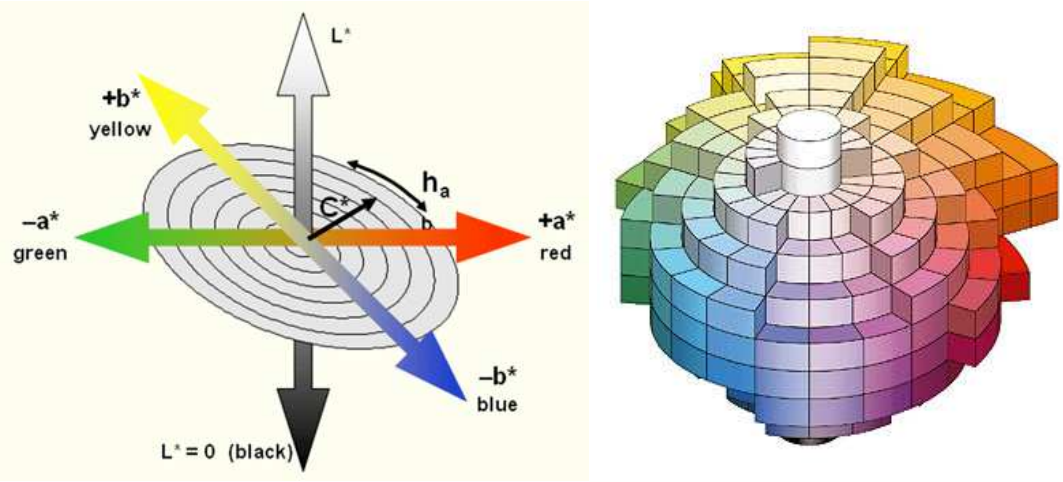


Fig. 3.2 The CIELAB color system: coordinates on the left, color solid on the right. Figures reproduced from http://www.globalspec.com and http://www.dtpstudio.de/english respectively.

preferred to calculate ΔE^*_{ab} using the cylindrical coordinates:

$$\Delta E^*_{ab} = \sqrt{(\Delta L^*)^2 + (\Delta C^*)^2 + (\Delta H^*)^2}, \tag{3.3}$$

where ΔC^* represents the differences in chroma, and ΔH^* the differences in hue.

When considering color differences, the most prevalent three types of visual assessments are [137, 20]:

1. Threshold differences: the magnitude of color difference indicating what is a just noticeable difference (JND).

2. Supra–threshold differences: the magnitude of color difference are not necessarily small or of near threshold, but can be many times JND threshold.

3. Acceptability thresholds: a less concrete concept and depend strongly on application and industry.

Color difference formulae are aimed such that their units correspond to JND's, thus it is commonly stated that any color difference below one unit is predicted as not being perceptible for samples viewed side by side [106, pp. 116–117]. Rules of thumb for practical classification are proposed e.g., in Stokes et al. [279] and in Sharma [266, pp. 139–150].

The dissatisfaction of the CIELAB color space, because not as perceptually uniform as intended, and the failure of the relative color difference formula in industry, has induced the researchers to produce other color difference data and refine the CIE 1976 color difference formulae.

So far, three major improvements of the ΔE^*_{ab} have been proposed, which are mostly based on the empirical color difference data known as the BFD (Bradford) ellipses [163]. All formulae are dependent on some parametric factors, which break the original Euclidean nature. Here we report them briefly:

1. ΔE_{CMC} (CMC $l : c$) 1984 [51]: recommended by British Colour–Measurement Committee (CMC) of the Society of Dyers and Colourist–UK, introduces a local metric tensor in the CIELAB space. As consequences, the color tolerances in the CIELAB space are represented by ellipsoids with semi–axis lengths depending on the point in the space and the axis oriented, one as the lightness, one as the chroma, and one as the hue, respectively. The weighting factors of the formula are thus function of the differences ΔL^*, ΔC^* and ΔH^*, and hue–dependent. A correction of the CMC formula in the blue region has been provided by Luo and Rigg [161, 164]. As result, the CMC formula has been

integrated into some ISO standards [51] and today has acceptance in industrial color control applications.

2. ΔE_{94} (CIE94) [44]: also introduces a local metric tensor in the CIELAB space in analogy with the CMC formula. The main deviations of the CIE94 formula with respect to CMC are the weighting factors, which are mathematically simpler and do not contain hue–dependent correction terms. Today, it is considered an obsolete formula and superseded by the CIEDE2000 formula.

3. ΔE_{00} (CIEDE2000) [160]: based on an enlarged dataset of empirical color differences, known as the COM dataset [160], added to the original BFD color difference data. This formula is only apparently based on the CIELAB space, because a coordinate transformation is made in order to correct hue distortions. Despite the enlarged empirical dataset, the CIEDE2000 formula presents distorted ellipsoids and shows an evident noise on the related data. Furthermore, the high number of parameters to be set on the experimental condition might lead to a difficult use of this formula.

Improved CIE color difference formula should be used when $\Delta E^*_{ab} < 5$. Rules of thumb for practical classification can be found e.g., in Shanda [264, p. 90] and in Hardeberg [97].

In 2002 Moroney et al. [197] proposed the CIECAM02 color appearance model as evolution of the previous work CIECAM97s from Luo and Hunt [159, 45]. The CIECAM02 [47] defines the following color attributes:

1. Brightness Q;
2. Lightness J;
3. Colorfulness M;
4. Chroma C;
5. Saturation S;
6. Hue H and hue angle h.

According to Moroney and Han, the CIECAM02 space is more uniform than the CIELAB space and thus it can be used as a connection space for gamut mapping in color management combined with ICC profiles [281, 96]. ICC profiles are not subject of this researcher work and thus we address the reader to the International Color Consortium [110], ISO 15076–1 [114] and Green [94] for a wider overview.

Following works from Li et al. [162] and Luo et al. [162] provide the CAM02–UCS color space with a related color difference formula indicated with $\Delta E'$ developed for large, small and combined large/small differences, respectively. According to Luo et al. [162], this improvement over the original model is recommended for any type of applications.

Despite the good intent of regularizing the distortions presented by the BFD ellipses in the blue–magenta hues and its relative success for industry applications, the focus of this work is neither color difference formulae based on the CIELAB color system nor on color appearance models, but a recent formula based on the OSA–UCS color system.

The OSA–UCS is a color system developed by Optical Society of America's (OSA) committee on Uniform Color Scales between the years 1947–1974 from an idea of Judd and the consecutive work of MacAdam [167]. The intention of the OSA committee was to show equally perceptible differences between all pairs of adjacent colors [165]. The OSA–UCS is realized by ceramic tiles ordered in a three–dimensional cube–octahedron lattice and constituted by 424 colors, which can be arranged in linear arrays (scales). The three–dimensional lattice is built by a set of parallel planes with square lattices. The cells of contiguous planes are mutually translated in order to have the corners in the centers of the cells of the contiguous planes. As consequence, any color sample is in the center of a three–dimensional cell constituted by 12 samples. These parallel planes are at constant lightness L_{OSA} and defined taking into account:

- The Helmholtz–Kohlrausch effect: the brightness depends on the luminance and on the chromaticity.

- The lightness crispening: represented by the modified Semmelroth formula [263], which asserts that the brightness of a color sample depends on the brightness of the surround (e.g., samples with lower brightness appear darker and with higher brightness lighter).

The planes at constant lightness L_{OSA} are spanned by two Cartesian coordinates (g, j), which recall two hues proper of the color opponencies: green–red opponency is specified by coordinate g, from English green, and blue–yellow by j, from French jaune. The coordinates (L_{OSA}, g, j) correspond to three global axes, which can be approximately related to Hering's opponency mechanisms. The physical color samples of this system have coordinates belonging to these intervals:

$$-7 \leq L_{\mathrm{OSA}} \leq 5, -10 \leq g \leq 6, -6 \leq j \leq 12. \tag{3.4}$$

Given a color sample, the conversion from CIELAB to OSA system requires two steps:

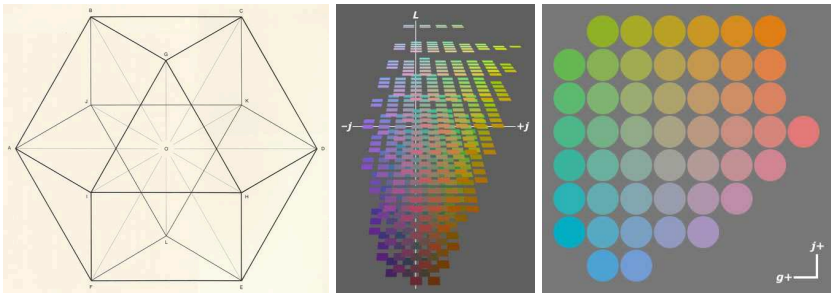


Fig. 3.3 The OSA–UCS color system: Cuboctahedron form on the left, color solid in the middle and constant lightness plane on the right. Figures reproduced from www.colorsystem.com and www.handprint.com

1. From L^*, a^*, b^* to X, Y, Z [264, p. 61]:

$$Y_n = \begin{cases} Y = Y_n \left[\frac{L^*+16}{116}\right] & \text{if } \left(\frac{Y}{Y_n}\right) > \left(\frac{6}{29}\right)^3 \approx 0.008856 \\ Y = Y_n \left[\left(\frac{3}{29}\right)^3 L^*\right] & \text{if } \left(\frac{Y}{Y_n}\right) \leq \left(\frac{6}{29}\right)^3 \end{cases}, \tag{3.5a}$$

$$X_n = \begin{cases} X = X_n \left[\frac{a^*}{500} + \frac{L^*+16}{116}\right]^3 & \text{if } \left(\frac{Y}{Y_n}\right) > \left(\frac{6}{29}\right)^3 \\ X = X_n \left[\left(\frac{3}{29}\right)^3 L^* + \frac{108}{841 \cdot 500} a^*\right] & \text{if } \left(\frac{Y}{Y_n}\right) \leq \left(\frac{6}{29}\right)^3 \end{cases}, \tag{3.5b}$$

$$Z_n = \begin{cases} Z = Z_n \left[-\frac{b^*}{200} + \frac{L^*+16}{116}\right]^3 & \text{if } \left(\frac{Y}{Y_n}\right) > \left(\frac{6}{29}\right)^3 \\ Z = Z_n \left[-\frac{108}{841 \cdot 500} b^* + \left(\frac{3}{29}\right)^3 L^*\right] & \text{if } \left(\frac{Y}{Y_n}\right) \leq \left(\frac{6}{29}\right)^3 \end{cases}. \tag{3.5c}$$

where X_n, Y_n, Z_n is the tristimulus of the reference white. The transformations are valid for both CIE '31 and CIE '64 systems.

2. From X_{10}, Y_{10}, Z_{10} to L_{OSA}, G, J, with 10 indicating the 10,° observer (CIE '64) [209, 210]:

$$L_{\text{OSA}} = \left[5.9 \left(Y_{10}^{1/3} - \frac{2}{3} + C\right) - 14.4\right] \frac{1}{\sqrt{2}}, \tag{3.6}$$

with:

$$C = \begin{cases} +0.042\,|Y_0 - 30| & \text{if } (Y_0 - 30) > 0 \\ -0.042\,|Y_0 - 30| & \text{if } (Y_0 - 30) \leq 0 \end{cases}, \tag{3.7a}$$

$$Y_0 = Y_{10} F, \tag{3.7b}$$

$$Y_{10} = Y_{10}(4.4934 x_{10}^2 + 4.3034 y_{10}^2 - 4.276 x_{10} y_{10} - 1.3744 x_{10} - 2.5643 y_{10} + 1.8103). \tag{3.7c}$$

$$\begin{pmatrix} J \\ G \end{pmatrix} = \begin{bmatrix} 2(0.5735 L_{\mathrm{OSA}} + 7.0892) & 0 \\ 0 & -2(0.7640 L_{\mathrm{OSA}} + 9.2521) \end{bmatrix} = \\ = \begin{bmatrix} 0.1792 & 0.9837 \\ 0.9482 & -0.3175 \end{bmatrix} \begin{pmatrix} \ln\left(\frac{A/B}{0.9366}\right) \\ \ln\left(\frac{B/C}{0.9807}\right) \end{pmatrix} \tag{3.8}$$

with:

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{bmatrix} 0.6597 & 0.4492 & -0.1089 \\ -0.3053 & 1.2126 & 0.0927 \\ -0.0374 & 0.4795 & 0.5579 \end{bmatrix} \begin{pmatrix} X_{10} \\ Y_{10} \\ Z_{10} \end{pmatrix} \tag{3.9}$$

The color difference formula developed for this space, termed ΔE_E, and proposed by Oleari et al. in 2009 [213], is Euclidean and defined as follows:

$$\Delta E_E = \sqrt{(\Delta L_E)^2 + (\Delta G_E)^2 + (\Delta J_E)^2}. \tag{3.10}$$

where L_E, G_E and J_E are appropriate transformations of the original coordinates L_{OSA}, G and J in order to be close to the HVS [213].

Three main reasons have invited this formula in this research work:

1. No hue distortion is present as in CIELAB.

2. Lower complexity than CIEDE2000 and equally good in the prediction of many available empirical datasets [105].

3. Image quality is evaluated in the visual situation of the small–medium color differences ($4 < \Delta E^*_{ab} < 8$ [266, p. 101].

3.3 Image quality metrics milestones

According to Wang and Bovik [294], following works in the state–of–the–art [282, 194] and to an exhaustive review from Pedersen and Hardeberg [226], image quality metrics can be classified in different ways. In a macroscopic view, the easiest and most common way is to divide image quality metrics in the following three classes [294, pp. 12–13]:

1. No–reference image quality metric:

> **Definition 3.10.** No–reference (or objective blind) image quality metric refers to automatic quality assessment of an image using an algorithm such that the only information that the algorithm receives before it makes a prediction on quality is the distorted image whose quality is being assessed.

2. Full–reference image quality metric:

> **Definition 3.11.** Full–reference image quality metrics refers to automatic quality assessment of an image using an algorithm that require as input not only the distorted image, but also a "clean", pristine reference image with respect to which the quality of the distorted image is assessed.

   According to Erola [72], the reference image represents the "ideal quality".

3. Reduced–reference image quality metric:

> **Definition 3.12.** Reduced–reference refer to automatic quality assessment of an image using an algorithm that possess some information regarding the reference image (e.g., a watermark), but not the actual reference image itself, apart from the distorted image.

The aforementioned definitions are extracted from Mittal [189], but almost identical definitions can be found in different state–of–the art works.

The interest of this research work are full–reference image quality metrics. According to Wang and Bovik [294, pp. 12], the term full–reference image quality metrics is often used for historical reason, but the precise term for these class of metrics should be image similarity or

fidelity measurement. According to Pedersen and Hardeberg, based on their target application image quality metrics can be referred also as image difference metrics, image similarity metrics and image fidelity metrics [226].

In this research work, we will focus on those image quality metrics, that are most known as image difference metrics as they aim at predicting the perceived magnitude of differences of one or more perceptual factors. From now on we will refer all of them just as image metrics or metrics, without making any distinction whether they could be more categorized as image difference metrics more than image quality or similarity or fidelity metrics and vice versa.

Before presenting a selection of these metrics, we recall the verbal definition of measure [185] and the mathematical definition of a metric [87, p. 1]:

Definition 3.13. Measure: the dimensions, capacity, or amount of something ascertained by measuring.

Thus, the result of measuring is a number or quantity that records a directly observable value or performance. All measures have a unit attached defined by standard organizations e.g., cd/m^2 for luminance defined by International System of Units.

Definition 3.14. Given a non–empty set X, a distance function d on X, called a *metric* for X, is a function which assigns to each pair of points a real number, (or formally, $d : X \times X \to \Re$), satisfying the following properties: For all $x, y \in X$

1. $d(x,y) \geq 0$ (non–negativity),
2. $d(x,y) = 0$ if and only if $x = y$
3. $d(x,y) = d(y,x)$,

and for all $x, y, z \in X$
$d(x,y) \leq d(x,z) + d(y,z)$ (triangle inequality).

In this research work, we will use metric and measure interchangeably, but in order to show the difference between them, we consider an example in the CIE '31 XYZ and in the CIELAB color spaces (see Section 3.2). If we consider a pair of tristimulus values, they quantify two color stimuli (measure). The difference between the two stimuli is still a tristimulus vector that

represents the difference of stimulation in a particular reference frame of tristimulus space, but does not quantify the perceptual differences between the two colors. On the opposite, the CIELAB space was conceived with the aim to represent the perceived color difference between a pair of color samples by Euclidean distance. In this case, the ΔE^*_{ab} formula (Equation 3.3) provides a metric. Unfortunately, this metric represents the perceived color differences only approximately, and thus over the time other formulae have been proposed to improve the agreement with empirical color difference data and image metrics have been developed with aim of estimating the perceived magnitude of differences of one or more perceptual attributes between two images.

Thus, in a easy manner, we can think a full–reference image metric as a distance function that given two input images (the original reference and the reproduction) produces a real number. These distance functions, or better image metrics, can be classified in several ways [226] e.g., Wang and Bovik [293] propose the two following groups:

1. Mathematically based: the computation rely only on statical measures.

2. HVS based: consider HVS characteristics, although the final computation can still rely on statistical measures. Thus, they introduce typical features for modeling the HVS (e.g., color difference formulae and CSFs) and and take into account aspects that influence perception (e.g., the viewing conditions).

In this short survey, we will not strictly focus on this classification, but for some of the metrics this distinction will be underlined. For a wider overview of image metrics classification and approaches to target specific applications, we address the reader to the work of Pedersen and Hardeberg [226].

We present now a small selection of important metrics focusing on those based on contrast filtering, or those which have been shown to have high correlation with observer perceived difference [238].

Full–reference image metrics are based on many different ideas, a substantial number of them usually follow a general framework shown in Figure 3.4 [221].

One of the first image metrics was proposed by Zhang and Wandell [317] in 1997 with the aim of measuring color reproduction errors of JPEG–DCT compressed images, halftoned images and simple test patterns. The metric is described as a spatial extension of the CIELAB color difference formula, hence the name S–CIELAB. S–CIELAB has a simple framework shown in Figure 3.5.

First, both original and reproduction images are first separated into an opponent–color space (luminance, red–green, and blue–yellow space). Afterwards, each opponent–color image is convolved with a two–dimensional spatial kernel determined by the visual spatial sensi-

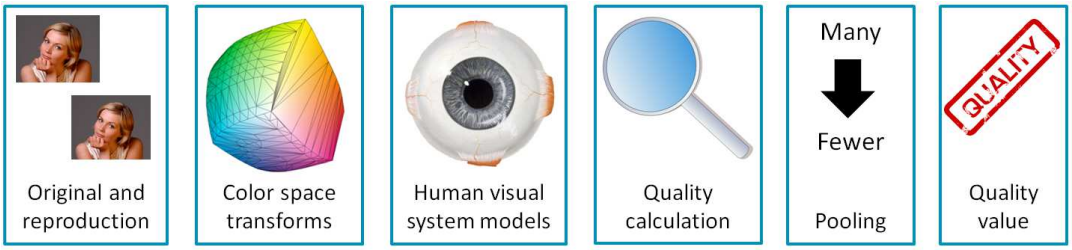


Fig. 3.4 Image quality metric general workflow, consisting of a series of cascading steps [221]. In a first step, the original reference image and its reproduction are first transformed into a suitable color space, preferably a perceptually uniform one. Then, in a second step, a simulation of the HVS is carried out, from simplistic methods as smoothing of the image by a local neighborhood to more complex methods, as using Contrast Sensitivity Functions (see Figures 2.15–2.16 in Section 2.2.2 for more information). After, in a third step, these metrics usually perform a calculation of quality, usually using a color difference formula. Finally, in a fourth step, a process of pooling is executed, where the quality values from the previous steps are reduced to one single number representing the overall quality of the reproduction with respect to the reference. Figures provided and reproduced by courtesy of Marius Pedersen [221].

tivity of that color dimension. Then, both filtered original and filtered reproduction images are transformed back to CIE XYZ, and finally color differences are computed using the CIELAB formula.

According to Zhang and Wandell, the spatial filtering is necessary to simulate the blurring of the HVS and to be consistent with the CIELAB large uniform areas. Both the color transformations and the three spatial filters are estimated from human psychophysical measurements of color appearance.

The results obtained by S–CIELAB have led researchers to extend and/or refine the metric for a wider class of distortions and for a better agreement with observers' judgment in psychophysical experiments. Evolutions and extensions of S–CIELAB follow two approaches:

1. Improvement of the spatial filtering.

2. Replacement of the traditional ΔE^*_{ab} with more performing and recent formulae.

A first improvement was proposed by Johnson and Fairchild [125] in 2001. According to their analysis, the original spatial filtering of the S–CIELAB can be seen as a CSF removing edges information in the scene. This leads to a general decrease of perceived differences of high frequency image information (e.g., halftone dots). In order to avoid this problem, they suggest to use an edge enhancing filter (e.g., Sobel kernel). Practically, they reach the improvement in performance applying the CSFs earlier presented in Equations 2.10–2.11.

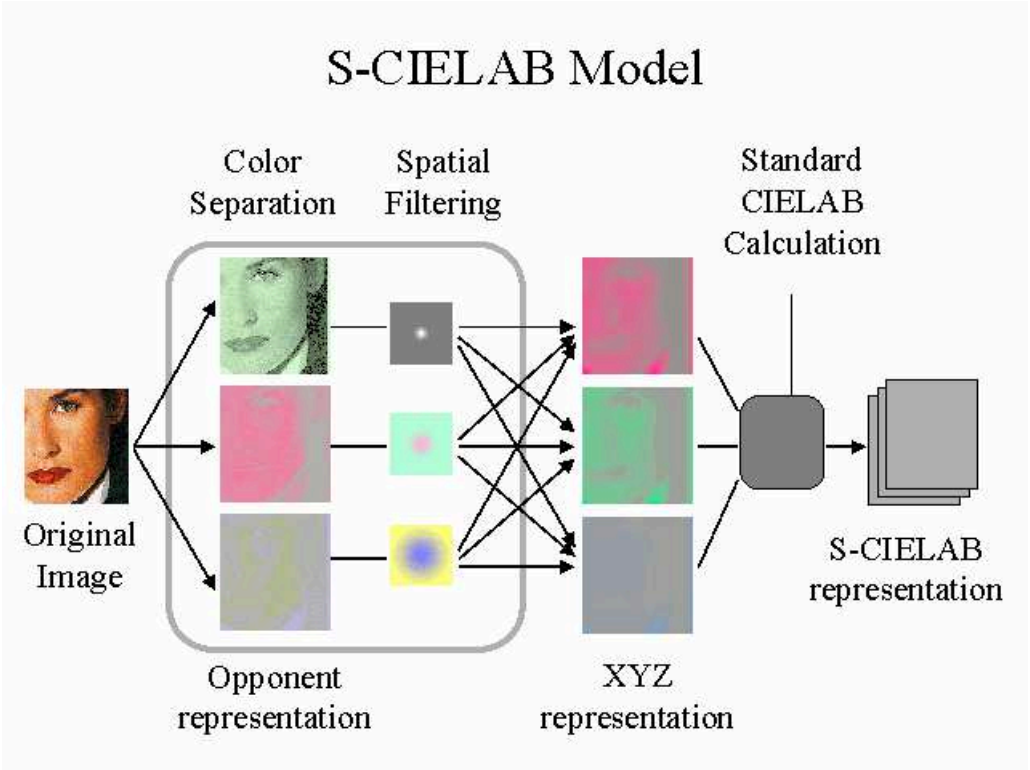


Fig. 3.5 S–CIELAB workflow. Figure reproduced from http://white.stanford.edu.

Later in 2004, Fairchild and Johnson [80] proposed the iCAM framework. The iCAM framework follows the S–CIELAB workflow, but exploits the IPT color space and adds two additional steps:

1. Spatial Localization: an edge enhancement kernel is applied to boost high frequencies in both the reference and the the reproduction images.

2. Local Contrast Detection: a non–linear mask–based local contrast enhancement proposed by Moroney [196] is applied to detect global and local contrast changes between the reference image and the reproduction one.

The IPT opponent color space is derived from a transformation proposed by Ebner and Fairchild [70] of the LMS color space. According to the authors, the transformation from RGB to IPT is much simpler than the transformations proposed in CIECAM02 and used in early stages of iCAM [80]. As last step in iCAM, like in S–CIELAB, color differences are computed using the Euclidean distance, known as ΔI_m in IPT.

In 2008, Roselli et al. [256] proposed the X–CIELAB metric as a novel extension of the S–CIELAB, replacing the original spatial filtering with a CSF adapted from Mannos and Sakrison model [170] and the traditional ΔE^*_{ab} with the most recent CIEDE2000.

In 2009, Wang and Hardeberg [297] proposed the Adaptive Bilateral Filter (ABF) metric. The ABF metric, intuitively by the name, replaces the CSF with an adaptive bilateral filter with the aim of blurring the image, but preserving edges. The filtering is performed in the CIELAB color space, and the color difference formula is still the traditional ΔE^*_{ab}.

In the same year, a metric named Spatial Hue Angle Metric (SHAME) was proposed by Pedersen and Hardeberg [225]. The metric performs a spatial filtering similar to the original S–CIELAB [317] on both original and reproduction images, to be given afterwards as input to the Hue Angle metric proposed by Hong and Luo [103] in 2001.

The Hue Angle metric relies on the known fact that systematic errors over the entire image are quite noticeable and unacceptable. A histogram based on the hue angle is computed, and sorted in ascending order so that weights can be applied to four different quartiles of the histogram. The overall color difference is calculated by multiplying the weighted hue angle of each pixel with the pixelwise ΔE^*_{ab}.

The S–DEE metric, also developed in 2009, is the first metric using the Euclidean color difference formula in log–compressed OSA–UCS space (see Section 3.2), and has been proposed by Simone et al. [273]. This metric is based on the modified S–CIELAB workflow developed by Johnson and Fairchild [125] and simply substitutes the traditional ΔE^*_{ab} with the ΔE_E in the last step of the workflow.

Concurrently to the development of S–CIELAB framework based metrics, other image metrics have been proposed with different mechanisms of evaluation of quality. Among the most known, The Universal Image Quality Index (UIQ) proposed by Wang and Bovik [293] in 2002 and the Structural SIMilarity (SSIM) index proposed by Wang et al. [296] in 2004.

According to Wang and Bovik [293] UIQ is not an image quality metric but rather a "universal" image quality index, where the authors as universal intend a measure that is not HVS model based, and thus independent of the image being tested, viewing conditions and individual observers. Furthermore, it should be able to estimate the quality of an image among various type of distortions. UIQ is defined as follows:

$$UIQ(x,y) = \frac{4\sigma_{xy}\mu_x\mu_y}{\left(\sigma_x^2+\sigma_y^2\right)\left(\mu_x^2+\mu_y^2\right)}. \tag{3.11}$$

SSIM follows the same approach and is defined as follows:

$$SSIM(x,y) = [l(x,y)]^{\alpha}\cdot[c(x,y)]^{\beta}\cdot[s(x,y)]^{\gamma}, \tag{3.12}$$

with:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1}, \tag{3.13a}$$

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2\sigma_y^2 + C_2}, \tag{3.13b}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \tag{3.13c}$$

where $l(x,y)$, $c(x,y)$ and $s(x,y)$ are measures of the luminance, contrast and structure of the image, and C_1, C_2 and C_3 are very small constants to prevent instabilities when the denominator tends to zero. The parameters α, β and γ are used to define the relative importance of the three components, and they are usually set to unitary value.

Both UIQ and SSIM involve three statistical indexes in the calculation:

- μ_x, μ_y: the mean for the signal x and y respectively.
- σ_x, σ_y: the standard deviation for the signal x and y respectively.
- σ_{xy}: the correlation between x and y.

Both metrics have been developed for greyscale images and tested on a wide variety of distortions: impulsive salt–pepper noise, additive Gaussian noise, multiplicative speckle noise, mean shift, contrast stretching, blurring and JPEG compression. Producing results in the range

of $[-1,1]$, make them particular suitable for calculation of correlation coefficients, which will be introduced in the next section.

The first evolution of SSIM was proposed in a follow–up by Wang et al. [300], extending the metric to multilevel analysis. Considering viewing conditions (e.g., display resolution and viewing distance) a drawback of SSIM, the Multi–Scale SSIM (MS–SSIM) was developed with the intent to be more flexible and determine the quality values across different scales, following the same principle of CSFs (see Section 2.2.2). MS–SSIM is defined as follows:

$$MS\text{–}SSIM(x,y) = [l_M(x,y)]^{\alpha M} \cdot \prod [c_i(x,y)]^{\beta_i} \cdot [s(x,y)]^{\gamma_i}, \tag{3.14}$$

with:

$$\sum_{i=1}^{M} \gamma_i = 1,\ \alpha_i = \beta_i = \gamma_i,\ \forall i, \tag{3.15}$$

where M is the number of scales (or levels), and α_i, β_i and γ_i are empirically retrieved parameters to adjust the relative importance of the three different components. It can be noticed that MS–SSIM reintroduces some feature of the HVS based metrics.

At the same time, the results of UIQ and SSIM induced researchers to extend these metrics also to color images. The extension of UIQ to color images, Q_{color}, was proposed by Toet and Lucassen [284] in 2003 and is defined as follows:

$$Q_{\text{color}}(x,y) = \sqrt{\omega_l \left(Q_l(x,y)\right)^2 + \omega_\alpha \left(Q_\alpha(x,y)\right)^2 + \omega_\beta \left(Q_\beta(x,y)\right)^2}, \tag{3.16}$$

where Q indicate the traditional UIQ calculation, and ω_l, ω_α and ω_β are the weights of the l, α and β channel respectively. The l, α and β channels are derived by a transformation of the LMS color space proposed by Ruderman et al [258].

The extension of SSIM to color, SSIM–IPT, was proposed by Bonnier et al. [26] in 2006 and is defined as follows:

$$SSIM\text{–}IPT(x,y) = \sqrt[3]{SSIM_I(x,y) + SSIM_P(x,y) + SSIM_T(x,y)}, \tag{3.17}$$

where I, P and T indicate the color channels respectively. The IPT color space is derived from a transformation proposed by Ebner and Fairchild [70] of the LMS color space.

Since the origins of signal theory, a well–known tool to analyze an original signal with respect to a reconstructed signal is the Signal–to–Noise Ratio (SNR), which has been extended to different fields. In the case of image metrics for image compression and video quality [289], a common metric is the Peak Signal–to–Noise Ratio (PSNR). PSNR is a measure of the peak

error between the compressed and the original image defined as follows:

$$PSNR = 10 \cdot \log_{10} \left[\left(\frac{\max(I(x,y))}{RMSE} \right) \right]^2 = 20 \cdot \log_{10} \left(\frac{\max(I(x,y))}{RMSE} \right). \tag{3.18}$$

The higher the PSNR, the better is the quality of the compressed (or reproduction) image. In order to improve the match with subjective evaluation, in 2006, Egiazarian et al. [73] proposed an evolution named PSNR–HVS. Additionally, PSNR–HVS performs a mean shift removal and contrast stretching similar to UIQ [293]. Furthermore, the MSE is calculated as described by Nill [205]. An extension of this metric incorporating contrast masking, PSNR–HVS–M, was given a short time later in 2007 by Ponomarenko et al. [239].

In the same year, Chandler and Hemami [43] developed the Visual Signal–to–Noise Ratio (VSNR), with the aim of quantifying the visual fidelity of a distorted image with respect to an original. VSNR is based on two main stages. In the first stage, the difference between the distorted and the original image, denoted as E, is processed via an octave–bandwidth wavelet decomposition. Afterwards, for each octave band i, the contrast detection threshold CT_i is computed based on the Contrast Signal to Noise Ratio (CSNR) as follows:

$$CT_i = \frac{C_i^{RMS}(I)}{CSNR_i^{thr}}, \tag{3.19}$$

where:

$$CSNR_i^{thr} = a_0 i^{a_2 ln(i) + a_1}, \tag{3.20}$$

where I indicate the original image, and:

$$a_0 = 59.8, \tag{3.21a}$$
$$a_1 = -0.1258, \tag{3.21b}$$
$$a_2 = -0.1087. \tag{3.21c}$$

According to the authors, the values of the parameters were empirically retrieved averaging the results of 14 images and then adjusted to take into account visual masking and visual summation phenomena.

In order to determine if distortions are visible (suprathreshold), $C_i^{RMS}(E)$ is compared with CT_i. If below the threshold, the distorted image is judged to be perfect in visual fidelity and thus no second stage is required, otherwise, the second stage is performed based on low–level of visual property of perceived contrast and mid–level visual property of global precedence,

which lead to the final measure VSNR defined as follows:

$$VSNR = 10\log_{10}\left(\frac{C_i^{RMS}(I)}{VD^2}\right), \tag{3.22}$$

with:

$$VD = \alpha d_{pc} + (1-\alpha)\frac{d_{gp}}{\sqrt{2}}, \tag{3.23}$$

where $d_{pc} = C^{RMS}(E)$ is the perceived contrast of the distortions, and $d_{gp} \in [0, \sqrt{2}d_{pc}]$ is a measure of the extent to which global precedence has been disrupted. Global precedence states that the HVS visually integrates image edges in a coarse–to–fine–scale (global–to–local) mechanism [202]. According to the authors, a general principle for computational models of global precedence in wavelet domain is not available, and for this reason they provide a method derived from one of their previous studies [42].

In 1986 Watson [302] proposed the Cortex Transform with the purpose to be used for different image processing problems. According to the author, the Cortex Transform has the following seven properties:

1. The output of the transform is similar to a Gabor filter.
2. It generates a pyramid of about one octave bandwidth.
3. It can be calculated at different orientation bandwidths.
4. If the input image is reduced in size, the output is shifted into a new level of the pyramid.
5. The transform can be inverted, thus recomposing (or summing up) the levels of the pyramid yields to the original input image.
6. It generates different pyramids, one per selected orientation.
7. It is a fast algorithm to model the HVS.

The core of the Cortex Transform is the mesa filter defined as follows:

$$\tilde{m_0}(u,v) = \left(\frac{\gamma}{f}\right)^2 \exp\left[-\pi\left(r\frac{\gamma}{f}\right)^2\right] * \Pi\left(\frac{r}{2}f\right), \tag{3.24}$$

with:

$$f = \beta\frac{N}{2}, r = \sqrt{(u^2+v^2)}, \tag{3.25a}$$

where Π is a rectangular pulse filter of unit height and width, and γ is a sharpness parameter. From $\tilde{m_0}$, the following layers of the pyramid can be calculated as follows:

$$\tilde{m_i}(u,v) = \tilde{m_0}(u,v)\left(s^i u, s^i v\right), \tag{3.26}$$

where s is scaling factor and i indicate the level of the pyramid. According to the author, $\beta = 0.9$, $\gamma = 4$ and $s = 2$ are suitable values. When $s = 2$, the bandwidth is exactly one octave.

In 1992, Daly [56] proposed the Visible Differences Predictor modeling three main sensitivity variations of the HVS.

1. Local amplitude non–linearity;
2. Contrast Sensitivity Function;
3. Detection mechanism.

The detection mechanism is based on the Cortex Transform and in particular on the Difference of Mesa (DOM) filters. According to the author, the VDP is not a image quality metric, but rather a image fidelity metric. The output of VDP is not a single number, but a map of detection probabilities approaching image distortions such as blur, noise, banding, blocking, pixellation, algorithm artifacts and tone scale changes. Improvements and extensions to other type of distortions of the VDP (e.g., Lukin [157] and Mantiuk et al.[171]) follow later, but they are not subject of this research work.

In our research work, the S–CIELAB framework is the main interest for the two following reasons:

1. The prefiltering can be easily extended to multilevel.
2. The traditional ΔE^*_{ab} can be easily replaced with another one (e.g., with the ΔE^*_E like proposed by Simone et al. [273]).

3.4 Quantifying and evaluating image quality

Similarly to what was presented in Section 2.3 for contrast measures, an important issue with image quality metrics is how to represent quality to an observer or user.

Most of the image quality metrics produce a quality map, thus also in this case a process of pooling is required. Like in the case of contrast, we introduce here a definition of quality pool as follows:

Definition 3.15. Quality pool is any type of pixel quality (or distortion) accumulation to reduce a quality (or distortion) map to a manageable set of values.

In current literature quality maps and distortion maps have the same meaning. Whether to use one term or the other is target application dependent. Generally, distortion map is used with sets of compressed images. For a wider overview on how to classify metrics and their own target application, we address the reader to Pedersen and Hardeberg [226].

Back to pooling, as well as for contrast, the process may lead to one number per level, one number per channel or one number per image. In current literature, as evaluation of metrics are almost always associated with psychophysical experiments and statistical analysis, one number per image is the most common pooling method to allow the users to judge an image. Pooling methods for image quality metrics have been further investigated with respect to contrast pooling [299, 193, 89].

According to Engeldrum [74], the simplest and most general method remains the Generalized Weighted Mean Hypothesis (GWMH) defined as follows:

$$Q = \frac{\sum_x^M \sum_y^N w(x,y) \cdot m(x,y)}{\sum_x^M \sum_y^N w(x,y)}, \tag{3.27}$$

where $m(x,y)$ is the quality map value at the pixel coordinates x,y, $w(x,y)$ is the given weight for quality map value at the pixel coordinates x,y, and M and N are the width and the height of the quality map respectively.

GWMH suggests that observers take some form of average when evaluating image quality and thus $w(x,y) = 1$ for all pixels. The result is the mean of the quality map.

Following the principle that low quality values should be weighted differently with respect to high quality values, in order to improve the performance of image quality metrics such as SSIM, Wang and Shang [299] proposed:

$$w(x,y) = f\left(m(x,y)\right), \tag{3.28}$$

with $f\left(m(x,y)\right)$ being a monotonically decreasing function if $m(x,y)$ represents a quality value, while $f\left(m(x,y)\right)$ is monotonically increasing function if $m(x,y)$ represents a distortion value.

Another well–known method is the Minkowski pooling [58] defined as follows:

$$Q = \frac{1}{MN} \sum_{x}^{M} \sum_{y}^{N} (m(x,y))^{p}, \tag{3.29}$$

where p is the Minkowsky power. Notice that $p = 1$ results in the simple mean of the quality map. Recent studies have shown that $p = 2$ and $p = 4$ generally improve performance of image quality metrics [299, 89].

In the recent years, many studies have shown that in pooling methods, weights should be determined by the local image content of either or both of the reference and the reproduction images [299, 206, 145, 193, 298, 55, 315], and thus:

$$w(x,y) = g(x,y), \tag{3.30}$$

where $g(x,y)$ indicate the pixel of visual attention at coordinate x,y. In different context, pixel of visual attention may also be referred as pixel–of–interest or pixel–of–importance. An aggregation of pixels–of–interest spatially contiguous may form a Region–of–Interest (or Region–of–Importance), in short ROI.

Regions–of–Interest can be retrieved through dedicated equipment or through mathematical/statistical models. Among the most common:

- Gaze information (or gaze map) [11, 12]: a region indicating where an observer looks, usually recorded with a dedicated equipment (e.g., eye–tracker device). For one particular pixel and/or all pixels belonging to the ROI, a gaze map can reveal e.g., number of fixations, number of fixations multiplied by time, and the total time spent by an observer.

- Saliency map [115, 291]: a region, which is more prominent than other regions in terms of image properties (e.g., color, contrast and orientation). Saliency maps are usually calculated using biologically inspired models, which use computational oriented approaches. They can be used to simulate the natural human fixation (gaze map) or target a specific application (e.g., object detection).

- Gaze–Attentive Fixation Finding Engine (GAFFE) [246]: a simulated gaze map retrieved from four statistically calculated image features: luminance, contrast, luminance–bandpass and contrast–bandpass.

Both gaze information and saliency maps cover a wide area of research with different goals and target applications. For a deeper overview, we address the reader to Zang and Ling [314].

Given the final quality score of the metric after a chosen pooling method, a comparison with the observer score can be performed. It is common practice to have more observers

evaluating the quality of an image, and thus calculating the so called Mean–Opinion–Score (MOS).

The MOS has been proposed and standardized by the International Telecommunication Union (ITU) for the evaluation sound quality systems [116, 117], followed by TV and HD–TV systems [118, 119]. In these fields the observers usually express their judgment of quality in terms of "Excellent, Good, Fair, Poor and Bad".

As well, the MOS has been extensively used in image quality [147, 268, 238], but it is common to ask the observer the judgment through a number. Thus, the MOS can be defined as follows:

$$MOS(I) = \frac{\sum_{o=1}^{O} s_o(I)}{O}, \tag{3.31}$$

where $s_o(I)$ is the score of observer o for the evaluated image I, and O is the total number of observers.

Although we are presenting the MOS as an average of quality scores, we underline to the reader that it can be calculated for any quantifiable image attribute (e.g., lightness, saturation and contrast). An example of MOS with 95% confidence interval is shown in Figure 3.6.

The confidence interval CI at 95% confidence level is defined as follows [285, p. 358]:

$$CI = MOS(I) \pm 1.96\frac{\sigma}{\sqrt{O}}. \tag{3.32}$$

The need of gathering a large number of mean opinion scores for a set of images with various type of distortions has induced researchers to perform psychophysical experiments in order to create image databases suitable for quantitative and qualitative analysis. For a description of public available databases, we address the reader to Winkler [306].

A candidate of this research work is the well–known Tampere Image Database 2008 (TID2008), which currently is the largest image quality database available in the public domain, both for test images and number of subjects. The TID2008 database contains 25 reference images manipulated with 17 different types of distortions over four levels, resulting in a total of 1700 test images. The distortions available are of various type e.g., JPEG and JPEG2000 compression, luminance and contrast changes, noise, blur and quantization. The MOS is the result of more than 800 observers from three different countries attending the experiments.

As well as the TID2008, three databases from the The Norwegian Colour and Visual Computing Laboratory are candidates for this research work. These databases have been developed under strict controlled lab conditions, and they are not available for public domain due to copy-

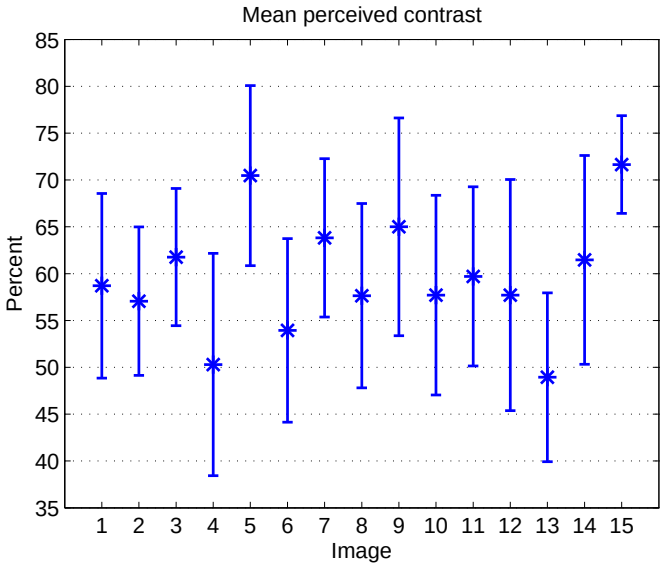


Fig. 3.6 MOS with 95% confidence interval of the 17 observers for the 15 images in Pedersen Contrast 2008 database. Figure reproduced from Pedersen et al. [228].

right restriction of some images. [1] Here follows a brief description:

- Pedersen Lightness 2008 [224]: contains four original images manipulated in lightness. For each image, four brighter and four darker reproductions were created with change in global and local lightness respectively, resulting in a total of 32 test images. The psychophysical experiment was performed in a windowless neutral grey room, with approximately 17 lux illumination, and the observers were seated approximately at 80 cm from a calibrated CRT monitor, LaCIE electron 22 blue II. A total of 25 observers were recruited for a pair–wise comparison experiment with the task to choose the reproduction image most similar to the original.

- Pedersen Contrast 2008 [228]: contains 15 original and no reproductions. The psychophysical experiment was carried out with the same settings of the previous one. A total of 17 observers, 9 considered experts (experience/job in color science, image processing, photography or similar) and 8 considered non–experts (none or little experience in these fields), rated the contrast of each image in a scale from 1 to 100. Additionally to MOS, this database provides gaze information recorded by eye–tracking, which are candidates of this research work.

[1] Public available databases from The Norwegian Colour and Visual Computing Laboratory are after this research work and available at http://www.colourlab.no/research_and_development/software.

This database has been designed to evaluate contrast measures scores with respect to observers perceived contrast. Actually, it would belong to Section 2.3, but has been voluntarily presented in this section in order to introduce first the concept of MOS and gaze information.

- Ajagamelle 2009 [4]: contains 10 original images manipulated with separate and simultaneous variations of contrast, lightness and saturation, resulting in a total number of 80 test images. The psychophysical experiment took place in a windowless neutral grey room, with ambient illumination around 40 lux. The display was a calibrated Eizo ColorEdge CG241W digital LCD display, placed at a viewing distance of 70 cm from the observers. The experiment was carried out as a category–judgment experiment with 14 observers, evaluating in a scale from 1 to 7 the overall difference as well as the differences in terms of contrast, lightness and saturation.

We can notice that all the three databases share a small number of participants. In recent years, this issue has always been considered as a strong drawback of psychophysical laboratory experiments and has lead researchers to move image quality evaluation under uncontrolled conditions in order to gather a large number of observers.

A pioneer work was presented Zuffi et al. [320] in 2008, performing an experiment for the evaluation in terms of preference of 21 printed images under controlled and uncontrolled viewing conditions. The analysis of the results indicated that the preferences expressed by the observers in the controlled and uncontrolled experiment were equivalent for most of the cases. In a related work, based on these results Zuffi et al. [319] designed a Web–based visual experiment management system with the purpose of facilitating online research investigation. During the development of the system, five key features were suggested and discussed for a successful online experiment: Communication, Seriousness of the study, Quality of the collected data, Solicitation to participate and Publicity. Communication provides guidelines on how to inform the users about the main objective of research investigation and on how to increase reliability (e.g., treatment of personal information and contact researchers). Seriousness of the study gives guidelines on how to provide information regarding the importance of the experiment and instructions for test execution (e.g., time of execution). Quality of the collected data provides some guidelines in order to avoid incorrect information and rejection of valid data (e.g., calibration). Solicitation to participate give some suggestions how to attract users (e.g., proposing a kind of game competition and giving rewards). As last, Publicity gives some hints on how to advertise the experiment (e.g., newsletters and URL in search engines).

When the metric scores and the MOS are available for each image, a common way to evaluate the performance of a metric is through statistical indexes, in particular with correlation

coefficients.

Given two discrete variables x and y, where x is representing the metric scores and y the MOS respectively for a set n of images, the two following correlation indexes recommended by ITU and Video Quality Expert Group (VQEG) are generally calculated:

1. Pearson product moment correlation coefficient [285, pp. 519–520]:

$$r_P = \frac{\sum (z_x, z_y)}{n-1} \tag{3.33}$$

where z_x and z_y are the z scores respectively for x and y, defined as follows:

$$z_x = \frac{x-\mu_x}{\sigma_x}, \tag{3.34a}$$

$$z_y = \frac{x-\mu_y}{\sigma_y}. \tag{3.34b}$$

The coefficient r_P is a dimensionless measure quantifying the strength of the linear correlation between x and y in the interval $[-1,1]$, where:

- 0: no linear association (x and y are linearly uncorrelated);
- 1: total linear association, with x and y varying in the same direction. This actually means that the metric scores follow the MOS, or in simple words, both metric and observer have claimed that all the images have good/bad quality;
- -1: total linear association, with x and y varying in the opposite direction. This actually means that the metric scores are in full contradiction with MOS, or in simple words, the metric has estimated good quality while the observers have judged bad quality respectively for all images.

2. Spearman rank correlation test [285, pp. 691]:

$$r_S = 1 - \frac{\sum_{d=1}^{D} d^2}{n\left(n^2-1\right)}, \tag{3.35}$$

where d is the difference between the ranks, and D is the total number of ranks.

The coefficient r_S is a non–parametric measure quantifying an association between two variables, using a set of ranked data instead instead of actual values. It does not make any assumptions on the frequency distribution of the variables and on how tightly are the ranked data clusters. A non–parametric test has the advantage to be able to detect relationships that are not linear, although will not specify which kind of non–linearity.

According to VQEG [289] and Sheikh et al. [267], the relation between image quality metrics scores and MOS are generally non–linear. This might be due by the fact that subjective ratings are most often at the end of the chosen scale of rating. In order to remove this weakness, they propose to use a non–linear mapping to fit in a common space the metric scores and the MOS, before calculating correlation coefficients.

VQEG [289] proposes the use of cubic polynomial:

$$f(y) = ay^3 + by^2 + cy + d, \tag{3.36}$$

where a, b and c and d are empirically obtained constants by fitting the function to the MOS.

Sheikh et al [267] instead propose the use of a logistic function defined as follows:

$$f(x) = \theta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\theta_2 (x - \theta_3)}} \right) + \theta_4 x + \theta_5, \tag{3.37}$$

where θ_i with $i = 1,2,3,4,5$ are parameters obtained by fitting the function to the metric scores. A well–known fitting method for logistic functions is the Nelder–Mead search [203], which in this specific case minimizes the sum–squared error between the transformed metric outputs and the corresponding MOS.

As correlation coefficients have uncertainties, that are necessary for the comparison of metrics results, is common to additionally perform:

1. Significance of the difference between the Pearson correlation coefficients.

2. The visual inspection of the confidence intervals.

The significance of the difference between the Pearson correlation coefficients proposed by VQEG [289] is based on the two following statistical hypotheses:

1. Null hypothesis (symbolized by H_0): it states that there is no significant difference between correlation coefficients.

2. Alternative hypothesis (symbolized by H_1): it states that the difference is significant, but it does not specify if better or worse.

Assuming that the normal distribution is a good fit for observers scores, this test uses the Fisher–z transformation to calculate the normally distributed statistic Z_n as follows:

$$Z_n = \frac{z_1 - z_2 - \mu_{(z_1 - z_2)}}{\sigma_{(z_1 - z_2)}}, \tag{3.38}$$

with:

$$z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right), \tag{3.39a}$$

$$z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right), \tag{3.39b}$$

$$\mu_{(z_1-z_2)} = 0 \text{ for the null hypothesis}, \tag{3.39c}$$

$$\sigma_{(z_1-z_2)} = \sqrt{\frac{1}{n_1-3}+\frac{1}{n_2-3}}, \tag{3.39d}$$

where r_1 and r_2 are the Pearson correlation coefficients for the first and the second metric respectively, and $n_1 = n_2 = n$ is the number of images.

Afterwards Z_n is evaluated at 95% confidence level against the t–Student value for the two–tail test [285, p. 754], and the null hypothesis H_0 can be rejected, thus two metrics can be significantly different, if:

$$-1.96 < Z_n < 1.96. \tag{3.40}$$

If the first metric has a higher positive correlation coefficient than the second metric, it is common to claim that the first metric has a better performance with respect to the second one.

The distribution of the correlation coefficients r is generally assumed not to be normally distributed, and thus the confidence intervals require to be calculated in Fisher–z domain. The confidence interval c_z at 95% confidence level is defined as follows [289]:

$$c_z = z \pm 1.96\sigma_z, \tag{3.41}$$

with:

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right), \tag{3.42a}$$

$$\sigma_z = \sqrt{\frac{1}{n-3}}, \tag{3.42b}$$

where n is the number of images.

In order to be visually inspected, a confidence interval c_z require to be converted back in the r domain to get the confidence interval of correlation r:

$$c_r = \frac{\exp(2c_z)-1}{\exp(2c_z)+1}. \tag{3.43}$$

An example is shown in Figure 3.7. In this scenario, c_r are used informally to compare different metrics, visually inspecting if one or more of them do not overlap with the others.

Given the CIs of two metrics, if the CI of the first metric does not overlap with the CI of the second one, it is possible to claim that two metrics are significantly different.

Afterwards, in the case the first metric has a higher positive correlation coefficient than the second metric, it is common to claim that the first metric has a better performance with respect to the second one.

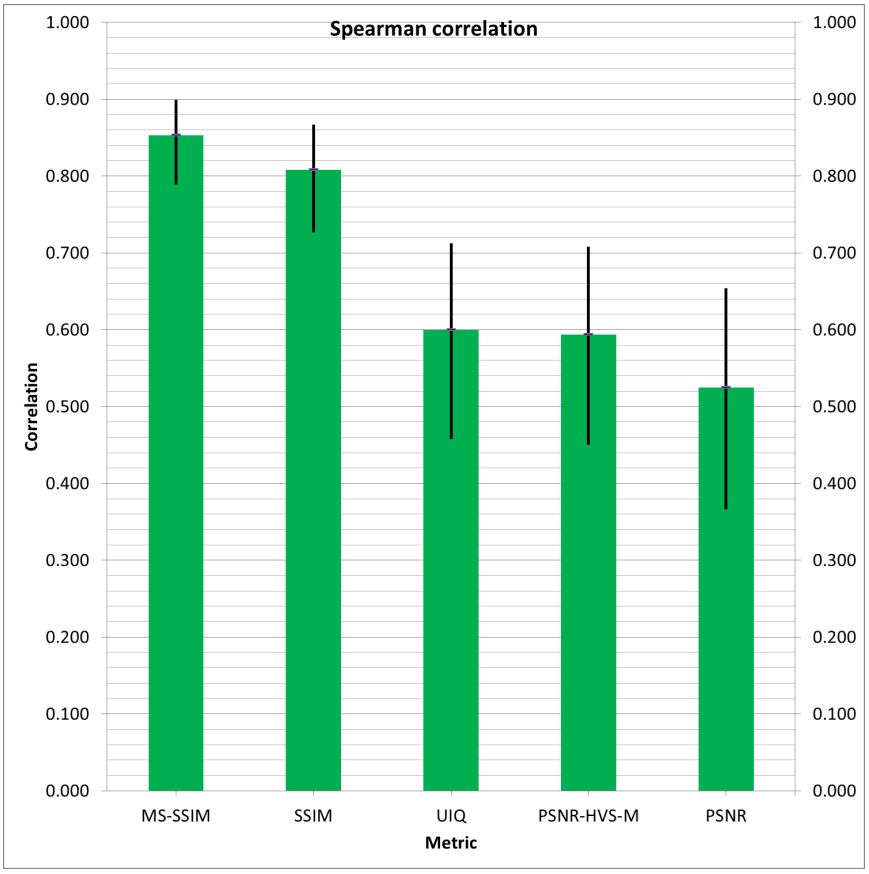


Fig. 3.7 Spearman correlation and confidence intervals of a small selection of image quality metrics on TID–full database. Numerical results from Ponomarenko et al. [238]. As we can see, MS–SSIM shows the highest positive correlation, but as its confidence interval overlaps with SSIM, it cannot be claimed that MS–SSIM and SSIM are significantly different. Both confidence intervals of MS–SSIM and SSIM do not overlap with those of UIQ, PSNR and PSNR–HVS–M, and thus it can be claimed that MS–SSIM and SSIM are significantly different with respect to the other metrics. As MS–SSIM and SSIM have also higher positive correlation with respect to the other metrics, we could conclude that they have a better performance.

3.5 Summary

The main purpose of the material presented in this chapter was to give the reader an insight on image quality metrics. First we introduced different definitions of image quality and then focused on CIELAB and OSA–UCS color spaces with a brief overview of the related color difference formulae.

Afterwards, we presented a selection of the state–of–the–art image quality metrics, which are the main core of this chapter. As currently there is no officially recognized (or standard) image quality metric, we spent particular attention on S–CIELAB, SSIM, VSNR and PSNR, which will be subject of comparison for the image metric proposal of this research work.

Finally, as image quality metrics are generally evaluated through statistics, we presented the concepts of quality pool, Mean–Opinion–Score (MOS), confidence intervals and, Pearson and Spearman correlation. Additionally, we reviewed some available databases for image quality metrics performance evaluation.

Chapter 4

Contrast and quality enhancement

Image quality deterioration is a common problem faced every day. In digital images, the degradation might be due to hardware failures and/or software errors. These undesired and unwanted changes in a digital image are named artifacts, and from the acquisition to the storing process, may occur individually or simultaneously. Artifacts can be prevented by hardware correction, but in most cases require to be treated in a post–process. Artifacts can be of various types and caused by the sensor, optics and internal image processing algorithms of the camera (Figure 4.1).

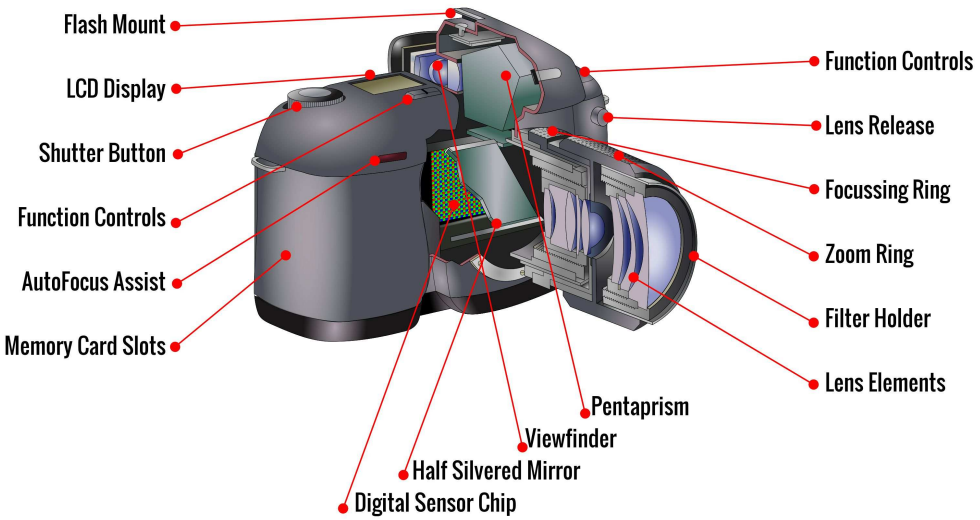


Fig. 4.1 Anatomy of a digital camera. Figure reproduced from http://www.ulearn.photography/

We present here a selection of the most common artifacts [8]:

- Blooming [260, p. 7] [8, p. 9] occurs because a digital camera sensor collects too many photons overcharging its photodiodes, resulting in overexposed pixel values (Figure 4.3(a)).

- Chromatic aberration [260, p. 774] [8, p. 175], known also as color fringing, occurs because the camera lens does not focus different wavelengths of light into the same focal plane (Figure 4.3(b)).

- Compression noise [260, pp. 125–126] [8, pp. 546–547] occurs during the storing process and is caused by the application of a lossy data compression algorithm (e.g. CPC, PGF, TIFF, JPEG and its variants), which may not be intelligent enough to discriminate among distortions of little subjective importance. (Figure 4.3(c)).

- Quantization noise [27, p. 330] [8, p. 168] occurs during image registration in the Analog–to–Digital Conversion (ADC) process, and is either due to rounding or truncation of the converted continuous signal into the discrete digital representation; (Figure 4.3(d)).

- Blocking [27, pp. 34–35] [8, p. 551] occurs because a group of pixels are mapped or rounded to the same grey/color values creating a contiguous region of pixels of the same value in the output image (Figure 4.3(e)). Blocking may be caused by different types of processing (e.g., encoding schemes, compression and interpolation).

- Ringing [27, pp. 77] [8, p. 551] occurs because of a conversion of a signal into or out of frequency domain (Fourier space). Ringing artifacts appear as dark and light ripples around features of the image in particular near edges (Figure 4.3(f)). Ringing may be caused by different image processing methods (e.g., blurring or inversion techniques, spatial sampling methods, iterative methods, inadequate filter design and compression).

- Blurring [27, p. 125] [8, p. 229] occurs because of a relative motion between the camera and the original scene or by an optical system that is out of focus. Blurring appears like smears along the direction of relative motion (Figure 4.3(g)).

  Blurring is not only caused by the involuntary shaking of the camera by the user or other external factors, but also by the combination of the limitations of the lenses and of the built–in anti–aliasing mechanism of the sensor. Today, the Modulation Transfer Function (MTF) is one of the most common measures for the performance of these two components [8, p. 128]. According to Allen and Triantaphillidou [8, p. 138], MTF measurements summarizes image quality attributes that contribute to subjective

impressions of sharpness and resolution. For a wider overview of MTF measurements of an imaging system, we address the reader to Allen and Triantaphillidou [8] .

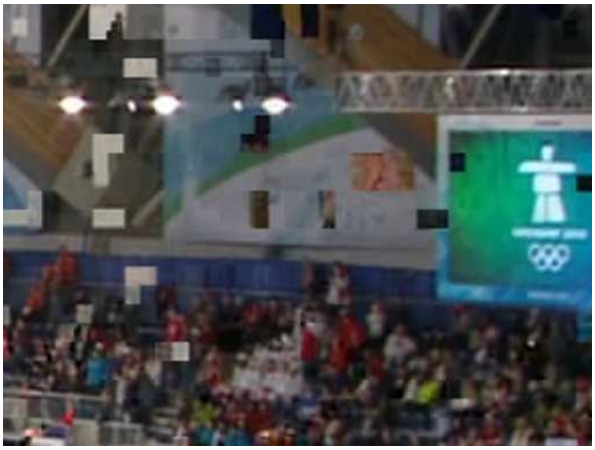
- Barrel [149, pp. 239–250] [8, p. 178] occurs because of lens errors applying an unequal magnification within the field of view. The magnification gradually decreases towards the corners of the image field, bending straight lines outward (Figure 4.3(h)).
- Pincushion [149, pp. 239–250] [8, p. 178] also occurs because of lens errors applying an unequal magnification within the field of view. In this case, the magnification gradually increases from the center axis to the corners of the image field, bending straight lines inward (Figure 4.3(i)).
- Vignetting [149, pp. 239–250] [8, p. 114, p. 192, p.443] occurs because of the reduction in effective lens aperture and the additional thickness of glass through which the light must pass, causing the progressive darkening of the image field towards the corners (Figure 4.3(j)).

(a) Blooming

(b) Chromatic aberration

(c) Compression noise (JPEG)

(d) Quantization noise



(e) Blocking

(f) Ringing

(g) Blurring

(h) Barrel

(i) Pincushion (j) Vignetting

Fig. 4.2 Example of artifacts occurring in digital imaging.
Figure (a) reproduced from https://nikoneurope-en.custhelp.com/app.
Figure (b), (c), (d) and (g) reproduced from Ponomarenko et al. [237].
Figure (e) reproduced from http://www.fame-ring.com/forum/.
Figure (f) reproduced from http://imaging.cs.msu.ru/en/research/ringing.
Figure (h) and (i) reproduced http://www.roborealm.com/help/Radial.php.
Figure (j) reproduced from https://more4u2c.wordpress.com/2010/10/23/.
All images have been manipulated to make them suitable for a printable version.

The need of automation, either to detect and remove artifacts or to restore and/or enhance individually or concurrently image attributes, has led to the development of image processing frameworks. From the software point of view, these frameworks may include drivers to manage hardware, application programming interface (API), compilers and debuggers. From the pure image processing point of view, these frameworks may include e.g., convolution methods, color space conversion and, design and application of complex filters (e.g., Marvin [10] and Cimg [150]. Thus, image processing frameworks consist of a set of versatile tools to develop from basic image processing algorithms to complex software for dealing with any imaging science problem.

A well–known framework has been proposed by the International Color Consortium (ICC) for color management [109]. Target applications using this framework have been developed by e.g., MacDonald for image sharpness [168] and Nussbaum for printing workflows [207]. For a wider overview we address the reader to ISO 15076–1 [114] and Green [94].

The interest of this research work is not building a framework for detecting and removing one or more of the aforementioned artifacts, but focusing on image enhancement frameworks. Although the definition and the specification of an image processing framework consider a wide software environment, in current literature algorithms such as wide variety of Spatial Color algorithms and Retinex model of computation are able to assess different target applications, and for this reason are also referred to as frameworks.

Before delving into the Spatial Color Algorithms and into the Retinex models, of main interest for this research work, we give a brief overview of image enhancement.

According to Gonzales and Woods, image enhancement is among the easiest and most interesting areas of digital image processing. The main idea image enhancement is to reveal obscured details and to highlight certain features of interest of the image. A common example of enhancement is increasing the contrast of the image because “it looks better!” [90, p. 52].

Here a selection of image enhancement definitions in a wider contest. According to:

- Gonzales and Woods [90, p. 102]:

Definition 4.1. Image enhancement is processing an image so that the result is more suitable than the original image for a specific application.

- Bovik [27, p. 16]:

Definition 4.2. Image enhancement is removing noise from an image while retaining the perceptual fidelity of the visual information.

- Maini and Aggarwal [169]:

Definition 4.3. Image enhancement is basically improving the interpretability or perception of information in images for human viewers and providing "better" input for other automated image processing techniques. The principal objective of image enhancement is to modify attributes of an image to make it more suitable for a given task and a specific observer.

- Bedi and Khandelwal [17]:

Definition 4.4. Image enhancement is improving the visual appearance of the image, or to provide a "better" transform representation for future automated image processing, such as analysis, detection, segmentation and recognition.

- Celebi et al. [39]:

Definition 4.5. Enhancement of digital images is the process of increasing the quality of the visual information by improving its visibility and perceptibility. Enhancement is a necessary step in image processing applications when the conditions under which is a scene is captured result in quality degradation, e.g. increased/decreased brightness and/or contrast, distortions of colors, and introduction of noise and other artifacts such as blotches and streaks.

The various definitions presented so far share some keywords e.g., improving and better, together with perceptual attributes such as visibility and contrast, meaning a sort of subjective evaluation in addition to objective results for image enhancement techniques. We defer the reader to Section 5.3 on how evaluate image enhancement techniques and other image processing methods.

Image enhancement techniques can be divided in two main categories [90, p. 75]:

1. Spatial domain methods;

2. Frequency domain methods.

In recent years, variational approaches have taken an important place in the field of image enhancement. However, considering them as a third category would be improper as their effective computation rely on methods in the frequency or in the spatial domain.

In next section, we will give a brief description of Spatial Color Algorithms, followed by an overview of the history of Retinex and its development in Section 4.2. Finally, we will present the color constancy and lightness constancy phenomena in Section 4.3.

4.1 Spatial color algorithms

A Spatial Color Algorithm (SCA) is a spatial domain image enhancement method with the main characteristic of recomputing the color of each pixel through the spatial distribution of values in the image [253]. Examples of SCAs are iCAM and its evolutions [80, 136], ACE [86, 252] and Retinex [144], while examples of spatial domain methods but not SCAs are gain/offset correction, logarithmic scale mapping and Miller–Ratio–Preserving Operator [188].

According to the studies of Rizzi and McCann [253], SCAs have the fundamental property of performing dynamic range remapping (or rescaling), and thus given an input image with a determined range, remap the output into the range available in the output device, eventually according to a perceptual post–Look–Up–Table. Look–Up–Tables (LUTs) are efficient and inexpensive memory tables used to transform color data and control the internal values within the minimum and the maximum of a color space [180, p. 380].

Three distinct remappings may occur, which are a generalization of several target applications:

1. Stretching (or maximization): from a smaller to a wider range.

2. Compression (or shrinking): from a wider to a smaller range.

3. Redistribution (or equalization): in the same range.

In all occurrences, the remapping output is strictly characterized by the behavior of the SCA and determined by the following features [253]:

- Local properties define how to explore and weight a set of neighboring pixels (e.g., paths, sprays and convolution masks);

- Final global scaling defines the final dynamic range output. Two image enhancement approaches are typically used:
  - Gray World (GW): independent channel averaging on the middle dynamic range value.
  - White Patch (WP): independent channel maximization to the white.

According to the task of SCA, a post–Look–Up–Table to match quantitatively the human color sensation can be used as additional final scaling.

In this research work, these two features (local properties and final global scaling) are unified under the definition of glocality. We introduce glocality as follows:

Definition 4.6. glocality = global scaling + local properties defining the amount and spatial distribution of changes between the original input image and the SCA filtered one.

SCAs may work on different types of images, among them:

- Low Dynamic Range (LDR): common digital images working in a 24 bit–depth RGB space, with 8 bits/pixels allocated to each color channel. The advantage of this representation is a reasonable reproduction on most commercial displays. The disadvantage is clearly that many colors cannot be represented, particularly values that are either too dark or too bright. Human observers can easily perceive details in scenes that span 4–5 orders of magnitude in luminance through local adaptation and reach to over 9 orders of magnitude in minutes. As consequence this representation covers only a few perceivable colors.

- High Dynamic Range (HDR): images typically encoded in a floating color space with higher bit–depth representation. These representations works from 16 bits/pixels to 96 bits/pixels for each color channel spanning from $\approx$ 10 to 79 orders of magnitude. These types of images are not directly displayable on common monitors and thus require a specific process explained later in this section. HDR images are a minor part of this research work, and for this reason the reader may refer to Ward [301], Reinhard et al. [250] and, Rizzi and McCann [180] for a wider overview.

The glocality lets the SCA target one or more applications. We recall here a short selection of them:

- Quantization: the process of mapping a broad range of input values to a limited number of output values defined in bits. Commonly quantization occurs during the Analog–to–Digital conversion (ADC) during the image registration. However for SCAs, quantization may refer to a process more known as bit–depth reduction, where a digital image is reconstructed in a lower bit precision. This a typical application of dynamic range compression.

- Dequantization: the opposite process of quantization, where a small range of input values defined in bits are mapped to a broad range of output values. Dequantization commonly refers to a process of Digital–to–Analog conversion (DAC). However for SCAs, dequantization may refer to bit–depth expansion, where a digital image is reconstructed into higher bit precision. This a typical application of dynamic range maximization.

- HDR to LDR: in current literature, the process that maps HDR to LDR images follows different names depending on whether the image is being acquired, displayed, stored and/or transmitted. SCAs generally focus on displaying, and common names for this process are tone reproduction, tone rendering and tone mapping. Most of the time these terms are used interchangeably, although the first term has a specific definition proposed by Tumblin and Rushmeier [286]. Tone reproduction requires a visual matching in controlled conditions, where a human observer compares a real scene to the image obtained by the tone reproduction operator. As in this research work the real world scene content will not be available, and thus a visual matching cannot be performed, we will adopt the term unsupervised tone rendering. This is another typical application of dynamic range compression.

- Gamut mapping: the set of all colors reproducible by a given device is named gamut. The process that maps the gamut of a given input device to the gamut of a given output device is known as gamut mapping. Examples are display to display, display to printer and camera sensor to display.

  Traditional gamut mapping algorithms works mainly using global methods, in particular colorimetric pixelwise principles. On the contrary, SCAs, due to the glocality, take into account the spatial properties of the image. In gamut mapping, all the three occurrences of dynamic range remapping may occur distinctly.

- Color correction (or color balancing): the process of improving the appearance of a image affected by color cast. Color correction is based on altering the balance of the color channels of an image in order to achieve a desired adjustment in terms of lighting and color. Most of the time, color correction is an application of dynamic range equalization.

- Illuminant/Reflectance estimation: the process of separating illuminant from reflectance. A general assumption is that the intensity of each color channel is proportional to the product of the reflectance and the light falling onto the object scaled by a geometry factor. This is another typical application of dynamic range equalization, most of the time resulting in details enhancement.

According to Hunt, when targeting an application, different types of color reproduction objectives can be aimed [106, pp. 164–177]. Hunt defines the following six ones [106, pp. 164–177]:

1. Spectral color reproduction:

Definition 4.7. Spectral color reproduction aims at equality of spectral reflectances or of relative spectral power distributions.

2. Colorimetric color reproduction:

Definition 4.8. Colorimetric color reproduction aims at equality of chromaticities and relative luminances.

3. Exact color reproduction:

Definition 4.9. Exact color reproduction aims at equality of chromaticities, relative luminances and absolute luminances.

4. Equivalent color reproduction:

Definition 4.10. Equivalent color reproduction aims at chromaticities, relative luminances and absolute luminances such as to ensure equality of appearance.

5. Corresponding color reproduction:

Definition 4.11. Corresponding color reproduction aims at chromaticities and relative luminances such as to ensure equality of appearance when the original and reproduction luminance levels are the same.

6. Preferred color reproduction:

Definition 4.12. Preferred color reproduction aims at whether at equal or at different absolute luminance levels.

According to McCann and Rizzi [180, pp. 3–4], when targeting an application, one of the most important factors is the careful definition of the goal of the output image. This goal is called rendering intent and defined as follows [180, p. 382]:

Definition 4.13. Rendering intent is the goal for image modification.

They propose the following rendering intents:

1. Reproduce exactly the light from the scene.
2. Match the appearance of the scene.
3. Calculate the surface reflectance of objects in the scene.
4. Calculate the appearance of the scene and print the appearance.
5. Abstract the important features of the scene.
6. Introduce a personal style to the image of a scene.
7. Use the scene for inspiration.
8. Make a pretty picture.

The main interest of this research work is the Retinex algorithm developed by Land and McCann [139, 144] in 1971, for the following four reasons:

1. It is a Human Vision Color Constancy model.

2. It is a White Patch Spatial Color Algorithm.

3. It can perform one of the three type of dynamic range remapping individually (or stretching or compression or redistribution), and thus target different applications (e.g., color correction and HDR unsupervised tone rendering).

4. According to Hunt, it aims at preferred color reproduction. According to McCann and Rizzi, it has the rendering intent of calculating the appearance of the scene and print the appearance. Often the result is an enhanced and prettier output image with the respect to the original.

The state–of–the–art of SCAs and their applications covers a wide area and more than 50 years of history. Other SCAs (or methods) e.g., ACE [86], Fattal [81], RACE [242] and iCAM06 [136] share the same aforementioned properties. Other SCAs (or methods) e.g., histogram based, variational based and wavelet based may share one or more of these properties. For a wider overview we refer to Bovik [27, pp. 71–206], Reinhard et al. [250, pp. 277–365], Chan and Shen [41] and, McCann and Rizzi [180].

4.2 Retinex: from early theory to actual models

The well–known publication "Retinex theory of color vision" written by Edwin H. Land begins with a well–explaining sentence [140]:

> "A retina–and–cortex system (retinex) may treat a color as a code for a three–part report from the retina, independent of the flux of radiant energy but correlated with reflectance of objects."

The term Retinex is a mix of the words retina and cortex, where in Land's theory both of them play an important role in the vision process.

4.2.1 Land's experiments and the original model

For at least 30 years, Land dedicated his studies understanding color vision, including a set of experimental setups named "Mondrians". These experiments were using a set of black and white or colored patches, inspired to the art works of Pieter Cornelis "Piet" Mondriaan, a dutch painter who lived between the 19^{th} and the 20^{th} century.

In apparent self–contradiction, black and white objects were the starting point of this color investigation. In one of his experiments Land demonstrated that a quantity, named by him

lightness, is associated to every object of a scene and not correlated to the illumination conditions or to the object location. Thus, the HVS perceives lightness without directly involving the actual light flux received by the eye [180, pp. 70–72, 116–117].

A particular array of white, grey and black papers, similar to a Black and White Mondrian (Figure 4.3), was presented to an observer in a non–uniform illumination from a single lamp. The position of the lamp was adjusted in order to have the luminance from a black paper near the lamp equal to the luminance from a white paper at the top of the Mondrian. Although the white at the top sent the same luminance to the eye as the black at the bottom, observers reported that the white paper was appearing white and the black paper was appearing black. Thus, appearances of white and black were correlating from exactly the same luminances. In other words, two areas with identical luminance (or pixel values in terms of imaging) generated the white and black lightnesses in the same scene at the same time [180, pp. 70–72, 116–117]. This phenomenon is called lightness constancy and briefly explained in Section 4.3.

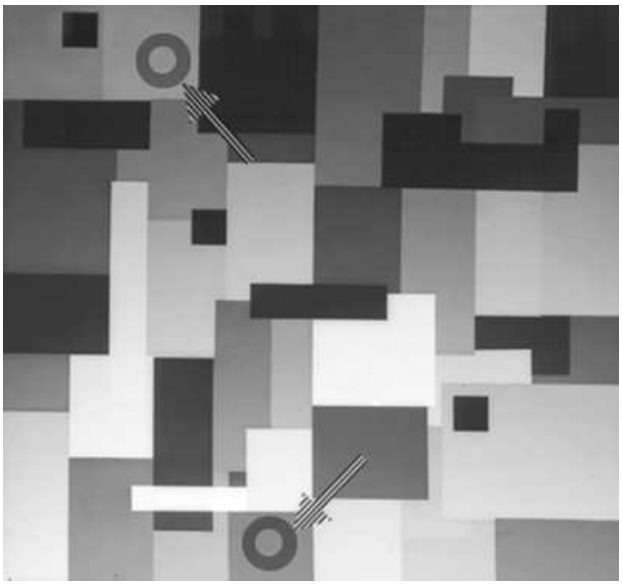


Fig. 4.3 Black and White Mondrian. Different size and shapes of the black, white and grey rectangular papers were used in order to suppress afterimages [57]. Figure provided and reproduced by courtesy of Alessandro Rizzi.

Afterwards, through another experiment Land demonstrated that is possible to obtain a full color sensation, producing a light stimulus to rods and cones sensitive to long wavelength (red). Thus, also in this case the color sensation is dependent on a comparison of lightness and not on spectral fluxes reflected by each object [180, pp. 225–235].

A set of colored patches, similar to a Color Mondrian (Figure 4.4), were presented to an

observer. Three white lights illuminated the Mondrian, one projecting with a red filter, one with a green filter and one with a blue filter, respectively. The observer was asked to adjust the intensities of the lights making a particular patch in the display appearing white. Then, the intensities of red, green and blue lights reflected from the white appearing patch were measured. At this point the observer needed to identify the color of a neighboring patch, telling which color he was perceiving. After that, the lights were adjusted in order to have the intensities of red, blue and green lights reflected from the neighbor patch the same as were originally measured from the white patch (the same of the starting point). The result was showing that all the patches during this light modification were keeping their original colors [180, pp. 225–235]. This phenomenon is called Color Constancy and briefly explained in Section 4.3.

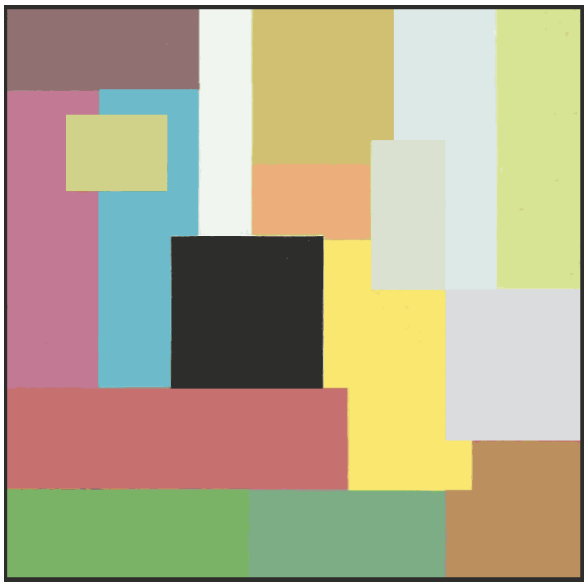


Fig. 4.4 Color Mondrian. Figure extracted and reproduced from a reprint of the original paper from Land in 1977 [140]

This experimental result led Land to the following conclusion [140]:

> "...the eye in determining color, never perceives the extra red because it does not depend on the flux of radiant energy reaching it."

Land continues explaining that there is a scientific tradition of simplifying the condition of an experiment, but paradoxically the modern technology of color photography has reinforced the belief that the colors discerned by Newton in the spectrum are the colors of the world around us.

The original paper was written to demonstrate how the eye perceives colors with the principal assumption of a strict dependency between color perception and neural structure of the HVS. In Land's model, the HVS is based on three retinal cortical systems that are processing separately the low, the middle and the high frequencies of the visible spectrum. Thus, a separate image is formed by each independent process and determined by the relative lightness values of the various regions of a scene [140]. Edges among adjacent areas of an image play a fundamental role in color perception and thus the choice of lightness ratio between two areas, which must be done as a dimensionless property describing their relationship.

This type of spatial comparison results in the following statements:

- If the lightness is very different between the two areas, the ratio is far from the unitary value. On the opposite, it tends toward the value of 1 where lightnesses tend to become equal.

- If the computed ratio tends to the average in many areas of the image, Retinex can discount a possible chromatic cast.

In the original Retinex model developed by Land and McCann [139, 144] in 1971, the previous statements are formalized computing the relative channel lightness (L) at a point i as the mean value of the relative channel lightnesses (l) computed along N random paths from point j to the point i (Figure 4.5(a)):

$$L^i = \frac{\sum_{h=1}^{N} l_h^{i,j}}{N}, \tag{4.1}$$

where:

$$l_h^{i,j} = \prod_{x=j}^{i} \delta_{x \in path} \left[\left(\frac{I_{x+1 \in path}}{I_{x \in path}} \right) \right], \tag{4.2}$$

where I_x is the lightness intensity of the pixel x, and I_{x+1} is the lightness intensity of the pixel $x+1$ and h is indicating the path. An example of random paths is shown in Figure .

The reset mechanism δ forces the chain of ratios to restart from the unitary value, considering the lightness value found at the reset point a new local reference white [180]:

$$\delta_{x \in path} = \begin{cases} 1 & \text{if } \frac{I_{x+1 \in path}}{I_{x \in path}} > T \\ \frac{I_{x+1 \in path}}{I_{x \in path}} & \text{else} \end{cases}, \tag{4.3}$$

where T is a defined threshold.

Subsequent studies have shown that the retinal receptors have a logarithmic response to

light [143, 142], and thus Equations 4.2 and 4.3 can be rewritten as follows:

$$l_h^{i,j} = \sum_{x=j}^{i} \delta_{x \in path} \left[\log \left(\frac{I_{x+1 \in path}}{I_{x \in path}} \right) \right], \tag{4.4}$$

$$\delta_{x \in path} = \begin{cases} 1 & \text{if } \left| \log \left(\frac{I_{x+1 \in path}}{I_{x \in path}} \right) \right| > T \\ 0 & \text{else} \end{cases}. \tag{4.5}$$

In the original Retinex model the three color channel R, G and B are processed independently and thus the lightness is represented by the triplet (L_R, L_G, L_B) of lightness values in the three chromatic channels.

As it can be noticed, the original model is characterized by five operators (or mechanisms):

1. Ratio;
2. Product;
3. Reset;
4. Threshold;
5. Average.

4.2.2 The extended Retinex family

Since the origins of the Retinex model and in the light of current developments, the respective role and importance of both threshold and reset and as well the use of paths for exploring the image content have incited a debate in literature [142, 177], particularly when Land retired as Polaroid Chairman of the Board and presented the first two departures from the original Retinex model in 1983 and in 1986, respectively.

The first departure was the same as Land and McCann [144] with a series of paths beginning in the surround and traveling to a central pixel of interest, but without the reset mechanism [141]. Since each path did not reset, the average of pixel values was implemented as novel way of normalizing the output to the maxima in each channel.

The second departure, instead, was proposed with the name of Designator [143, 142]. The Designator was differing from the original version of 1971 and the consecutive in 1983 in calculating the ratio of a central area to an extended fixed surround (center/surround technique). In details, the new algorithm was computing a distance weighted average and subtracting this average from the intensity value at that point, with the intent of finding lateral inhibition.

Threshold operation was replaced by averaging over a large area (local–average dependence), linking the lightness output to global and local averages.

The results of Retinex theory and the ongoing debate have lead the researcher to an extended family of models/algorithms. The only certainty is that there is not consensus among the different school of thoughts, although in 2005 Provenzi et al. [240] proved that the Retinex algorithm maintains its fundamental properties without the threshold mechanism. The spatial reset mechanism raises another issue, seen as the most important feature of Retinex.

Removing or changing the reset mechanism can modify the glocality and in some cases destroy the Retinex nature e.g., the Designator changed the glocality from WP like the original model to GW, although Land was pointing out that the Designator was still a three channels Retinex model of color.

Before giving a short review of the extended family, we highlight to the reader that in current literature the two terms Retinex model and Retinex algorithm are used as synonyms. However, if we want to underline a slight difference, Retinex model is often used for the mathematical formalization and its properties, while Retinex algorithm is generally used when discussing implementation details and target applications. As well, some algorithms define themselves Retinex–inspired or Retinex–like algorithms because share one or more mechanisms of the original Retinex (e.g., reset) or are derived from a re–formalization of the model.

In this research work, this extended family is conveniently divided in three classes based on the computational model and the way the image is explored:

1. Random–based use paths or extract random pixels around the pixel of interest.

2. Scale–based compute values over the image with convolution masks or weighting distances.

3. Variational–based use differential mathematical techniques based on Poisson–type equations and variational approaches.

All the algorithms belonging to this extended family has three unique attributes:

1. They use all the pixels in the image as input.

2. They are influenced by the content of the image.

3. They attempt to mimic human vision or the best rendition for observers.

We now present a brief description of these three classes, touching the most important milestones of the history of Retinex models.

The models/algorithms belonging to Random–based class are directly derived from the original Retinex model. A particular interest has been given by the "Milano Group" at the beginning of the '90s, which started the reformulation of the Retinex theory developing the so called MI–Retinex models/algorithms, inspired by the original work of Land in 1971 and his further contribution in 1983. A specific characteristic of these algorithms is the point of reset, which is always at the end of the path with respect to the original Retinex.

The first model was developed in 1993 by Marini and Rizzi [172] with the use of straight paths between the pixels while the second in 2000 [173] with the use of Brownian paths (Figure 4.5(b)).

A k–dimensional Brownian motion is a continuous time stochastic process $B(t) : t \geq 0$ taking values in $\Re^k$ and having the following properties [190]:

1. For all times $t_0 \leq t_1 \leq t_2 \leq ... \leq t_n$, the random variables $B(t_n) - B(t_{n-1}), B(t_{n-1}) - B(t_{n-2}), ..., B(t_1) - B(t_0)$ are independent, thus the process has independent increments.

2. The distribution of the increment $B(t+h) - B(t)$ does not depend on t, thus the process has stationary increments. The distribution is a normal distribution with zero mean and variance h.

3. $B(t) : t \geq 0$ is almost certainly (i.e., with probability 1) continuous.

Earlier than Marini and Rizzi, Brownian motion methods for generating Retinex paths were studied from a theoretical point of view by Brainard and Wandell [29] in 1986. In their work they analyzed Brownian paths formally using stochastic methods but only considering the case of infinitely long paths because it could be fit by an analytic expression. This limited behavior has no real interest as it loses the local behavior and keeps only the global scaling.

Marini and Rizzi instead adopted an approximation of Brownian paths using a mid–point displacement technique, inspired by the results of some neuro–physiological researches about human cortical vision areas, where the distribution of receptive fields centroids mimics a Brownian path, as demonstrated in many experiments by Zeki [312] on macaque monkeys. In their work locality influence is kept with the selection of pixels and the numbers of paths and their interaction with the spatial content. However, with increased numbers of paths and with many pixels per path, also the output of their approach may tend to the global normalization asymptote.

In 2011, Montagna and Finlaynson [191] re–investigated Brownian motion proposing an efficient algorithm with several important features:

1. The probability $P(n,d)$ of the so called walker being at distance d from the origin is

given by:

$$P(n,d) = \sum_h \sum_k w^n_{h,k}, \tag{4.6}$$

where h and k represent all the positions at city–block (Manhattan) distance d from the initial position, and $w^n_{h,k}$ is the probability that after n steps, depending on the probability that at step at $n-1$, the walker is in any of the neighboring and thus:

$$w^n = \left\{W^0, W^1, W^2, ..., W^n\right\}, \tag{4.7}$$

with:

$$W^0 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W^1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{1}{4} & 0 \\ 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad W^2 = \begin{bmatrix} 0 & 0 & \frac{1}{16} & 0 & 0 \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} & 0 \\ \frac{1}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16} \\ 0 & \frac{1}{8} & 0 & \frac{1}{8} & 0 \\ 0 & 0 & \frac{1}{16} & 0 & 0 \end{bmatrix} ... \tag{4.8}$$

2. The scene content is represented on a random multigraph, where visiting a vertex produces a random path on the original graph. The randomness is generated modifying the geometry of the initial graph (representing the original image content) with the Marsaglia and Tsang's ziggurat algorithm [174], which allows normal distributed paths of random length.

3. A lower bound to the number of visits to each pixel is guaranteed. Actually, they prove that in average a path visits each pixel $m = 2K$ times, where K is the number of vertexes.

4. The ratio–product–reset–average is defined as Frankle and McCann (see Equation 4.14 in the next section), and also paths are generated in a multilevel framework. This actually makes this algorithm a crossover between this class and the next one.

According to the authors, their model generates paths that are neither a pure random walk nor a Brownian motion as the first property is not satisfied (Figure 4.5(c)). For this reason, it takes the name of Constrained Pseudo–Brownian Motion.

Back in 2007, Provenzi et al. [241] investigated the effects of different spatial samplings, replacing paths with random sprays (Figure 4.6). The lightness calculation is redefined as follows:

$$L^i = \frac{\sum_{h=1}^N \delta spray_h^{i,M}}{N}, \tag{4.9}$$

where in this case N is the number of sprays, and $spray_h^{i,M}(x,y)$ is a two–dimensional point distribution across the image and defined in polar coordinates as follows:

$$spray_h^{i,M}(x,y) = \begin{cases} x & = x_i + f(\rho)\cos(\theta) \\ y & = y_i + f(\rho)\sin(\theta) \end{cases}, \tag{4.10}$$

where $f(\rho)$ is the density of the spray with $\rho \in RAND_n[0,R]$ and $\theta \in RAND_n[0,2\pi]$ the corresponding uniform random distributions of n points and radius R and 2π, respectively. The reset δ in Equation follows the Mi–Retinex mechanism and thus is defined as follows:

$$\delta = \frac{I(x_i,y_i)}{I(x_s,y_s)}, \tag{4.11}$$

with:

$$s = \begin{array}{c} \max\ m_j \\ j \in \{1,2,...,M\} \end{array}. \tag{4.12}$$

This re–formalization of the model takes the name of Random Spray Retinex (RSR).

In a follow–up, Kolås et al. [135] developed in 2011 the Spatio–Temporal Retinex–like Envelope with Stochastic Sampling (STRESS), where the random sprays are used to calculate two envelope functions, $E_{\max}$ and $E_{\min}$, representing the local reference black and white points, respectively. (Figure 4.7).

Given the two envelopes, for each pixel a dynamic range remapping is performed as follows:

$$O(x,y) = \frac{I(x,y) - E_{\min}(x,y)}{E_{\max}(x,y) - E_{\min}(x,y)}. \tag{4.13}$$

The models/algorithms belonging to Scale–based class are mostly derived from the Frankle–McCann algorithm patented in 1983 [83] and the Designator proposed by Land in 1986 [142].

Exploiting the engineering progress of the 70s, the Frankle–McCann algorithm was implemented in hardware, and the ratio–product–reset–average was defined as follows:

$$O^{k+1}(x,y) = \frac{reset\left[\left(I(x,y) - I(x \pm d^k, y \pm d^k)\right) + L^k(x \pm d^k, y \pm d^k)\right] + L^k(x,y)}{2}, \tag{4.14}$$

where d^k is the shift distance from the original image I at iteration k, and O is the output image. The values of d^k are chosen 90 degrees apart progressively reduced towards zero, forming the so known "spiral". As Frankle–McCann works in logarithm domain, $I(x,y) - I(x \pm d^k, y \pm d^k)$ takes the ratio, $L^k(x \pm d^k, y \pm d^k)$ the product, and $L^k(x,y)$ the average. We underline to the reader that in current Retinex state–of–the–art, the formalization $\log(I)$ is most often replaced

by I.

The Frankle–McCann algorithm follows a re–investigation in 1999 and in 2004 with a traditional software implementation and the addition of a post process Look–Up–Table (LUT) [175, 84]. The algorithm is known as McCann99 or Gamut Retinex.

In the aftermath, different implementations have been presented with the intent of improving spirals computation and/or the reset mechanism. In these algorithms, if present, the reset mechanism is referred as ClipToWhite function and actually presented as soft–reset, as weightings are applied usually before to the operator. Among them:

1. Cooper and Baqai [52] 2004: dual–spirals iteration:

$$O^{k+1}(x,y) = \frac{reset\left[w\left(I(x,y) - I(x \pm d^k, y \pm d^k)\right) + L^k(x \pm d^k, y \pm d^k)\right] + L^k(x,y)}{2}. \tag{4.15}$$

2. Pan et al. [215] 2013: jumping–spirals iterations, but no reset:

$$\begin{aligned} O^{k+1}(x,y) = &\left(1 - \frac{g\left(\left|I(x \pm d^k, y \pm d^k) - I(x,y)\right|\right)}{2}\right) I(x,y) + \\ &+ \left(1 - \frac{g\left(\left|I(x \pm d^k, y \pm d^k) - I(x,y)\right|\right)}{2}\right) \max\left(I(x,y), I(x \pm d^k, y \pm d^k)\right), \end{aligned} \tag{4.16}$$

   where $g(\cdot)$ is an edge stepping function.

3. Sobol [277, 278] 2006: which introduced several effective improvements including the Ratio–Modification–Operator (RMO), the spatially varying contrast gain and the partial contrast strength mask. As well, Sobol Retinex found success in hardware implementation on HP digital cameras, with the marketing name Digital Flash. The name was implying that it was able to perform the digital equivalent of adding automatic fill–flash to the scene to equalize the illumination. The ratio–product–reset–average is defined as follows:

$$O^{k+1}(x,y) = \frac{reset\left[RMO\left(I(x,y) - I(x \pm d^k, y \pm d^k)\right) + L^k(x \pm d^k, y \pm d^k)\right] + L^k(x,y)}{2}. \tag{4.17}$$

Land's Designator gave a new shape and light to the Retinex model, but only 10 years later in 1996, Jobson et al. picked up the idea and reformulated the Designator first in a Single–Scale Retinex (SSR) and then in a Multi–Scale Retinex (MSR) [122, 124, 123]. MSR was improving Land's designator and SSR implementing multilevel resolution based on Frankle

and McCann approach [83] and Gaussian surround instead of Land's inverse–square distribution [244, 243]. Thus, MSR is defined as follows:

$$O(x,y) = \sum_{h=1}^{N} w_h F_h(x,y) * I(x,y), \tag{4.18}$$

where in this case N is the number of scales, w_h is the weight of scale h, and F_h is a Gaussian surround at scale h given as follows:

$$F_h(x,y) = r \cdot exp\left(-\frac{x^2+y^2}{h^2}\right), \tag{4.19}$$

where r is the radius of the Gaussian. According to the author the number of scales is application dependent, but at least three are required [244]. The weights must be chosen such that:

$$\sum_{h=1}^{N} w_h = 1. \tag{4.20}$$

The authors state that in most cases equal weighting is adequate [244].

Jobson et al. Retinex, known also as NASA Retinex, found its application in remote sensing and satellite imaging. A first improvement of MSR was given short time later by Barnard and Funt [14] extending MSR with gamma adjustment and color balance.

Following, new implementations were proposed with the main purpose to replace the Gaussian kernel with more efficient filters or improving the multilevel resolution approach. Among them:

1. Meylan and Süsstrunk [186] 2006: adaptive filter with shape following the high contrast edges (or in other words adapting the shape to follow image contours):

$$F(x,y) = \frac{\sum_{\theta=0}^{360}\sum_{r=0}^{r_{max}} I\left(x + r\cdot\cos(\theta), y + r\cdot\sin(\theta)\right) \exp\left(-\frac{r^2}{\sigma_{\theta,r}^2}\right)}{\sum_{\theta=0}^{360}\sum_{r=0}^{r_{max}} \exp\left(-\frac{r^2}{\sigma_{\theta,r}^2}\right)}, \tag{4.21}$$

where θ is the angle of the radial direction and r is the distance to the central pixel (x,y) defined as follows:

$$\sigma_{\theta,r} = \begin{cases} \sigma_0 & \text{no high–contrast edge was crossed along } \theta \\ \sigma_1 & \text{a high–contrast edge was crossed along } \theta \end{cases}. \tag{4.22}$$

According to the authors $r_{max} = 3\sigma_0$.

2. Saponara et al. [262] 2007: low pass edge preserving recursive rational filter:

$$F(x,y) = \frac{S_h f_h + S_v f_v + I(x,y)}{S_h + S_v + I(x,y)}, \tag{4.23}$$

$$f_h = \alpha F(x-1,y) + (1-\alpha) I(x,y), \tag{4.24a}$$

$$f_v = \alpha F(x,y-1) + (1-\alpha) I(x,y), \tag{4.24b}$$

$$S_h = \frac{10^{-1}}{10^{-5} + \left(\log_{10}\left(\frac{I(x-1,y)+1}{I(x+1,y)+1}\right)\right)^2}, \tag{4.24c}$$

$$S_v = \frac{10^{-1}}{10^{-5} + \left(\log_{10}\left(\frac{I(x,y-1)+1}{I(x,y-1)+1}\right)\right)^2}. \tag{4.24d}$$

The algorithms belonging to Variational–based class are the most recent ones, mainly developed in the 21^{st} century even though a mathematical alternative to the Retinex algorithm that was substantially differing in form was proposed by Horn in 1974 [104]. Based on the main assumption of Retinex, where illumination varies smoothly over the image, while sharp discontinuities are present only where the reflectance changes, Horn proposed to decompose the image intensity I as a product of the reflectance R and of the incident illumination intensity L, so that $I = R \cdot L$. Taking the logarithm of I, applying the Laplacian and finally the threshold operator T, the relation can be viewed as a Poisson equation:

$$\Delta(\lg R) = T(\lg I). \tag{4.25}$$

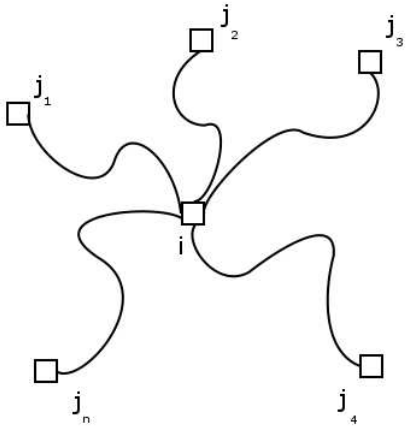
A tight mathematical connection between Land's and Horn's computations has been shown by Hurlbert [107]. A decade later Blake proposed a refinement, thresholding the norm of the gradient intensity with a smooth step function, which still yields to a Poisson equation.

In the aftermath of Horn's and Blake's work, the models in this class propose to translate the Retinex principles into a more physical form, leading to a set of equations or to an optimization problem. In particular, Horn's model was strengthened recently by Morel et al. [195] in 2010 first and then by Limare et al. [153] in 2011, proposing the so called PDE–Retinex or Retinex–Poisson formalizing the calculation of the lightness in the Fourier domain as follows:

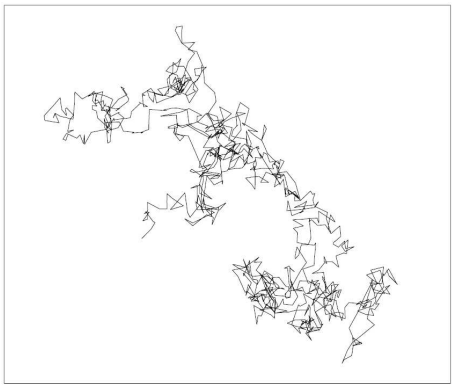
$$\hat{L}(u,v) = \frac{\hat{I}(u,v)}{4 - 2\cos\frac{2\pi u}{M} - 2\cos\frac{2\pi v}{N}}, \tag{4.26}$$

where in this case M and N are the width and the height of the image, respectively.

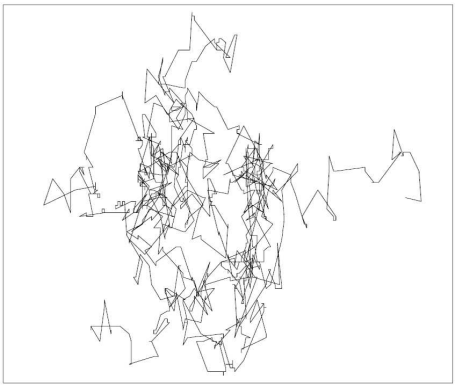
Earlier in 2003, Kimmel et al. [130] presented the first variational framework, or in other words the use of the Total Variation method. The Total Variation method is of particular interest of this research work and will be thoroughly introduced in the next chapter.



(a) Random paths



(b) Brownian paths

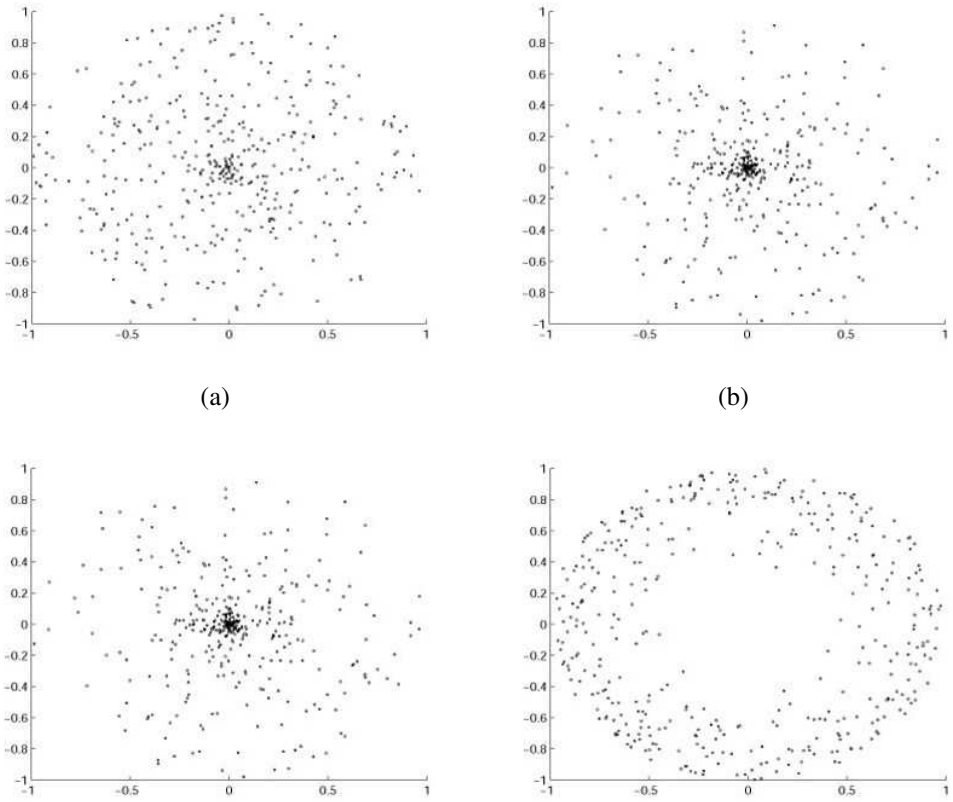


(c) Pseudo–Brownian paths

Fig. 4.5 Representation of N random paths from point j to the point i of the original Retinex model from Land and McCann on the top. Traditional Brownian motion investigation proposed by Marini and Rizzi on the bottom–left, Pseudo–Brownian motion investigation proposed by Montagna and Finlayson. Figures provided and reproduced by courtesy of Roberto Montagna.

4.3 Color constancy

The study of the appearances of objects in different illuminations date back to the late 19^{th} century. A famous quotation from Karl Ewald K. Hering in 1872 [100] states:



(a) (b)

(c) (d)

Fig. 4.6 Four examples of random sprays generated by RSR. The density of a spray depends on three parameters: the radius of the spray, the radial density function, and the number of pixels per spray. Figures provided and reproduced by courtesy of Alessandro Rizzi.

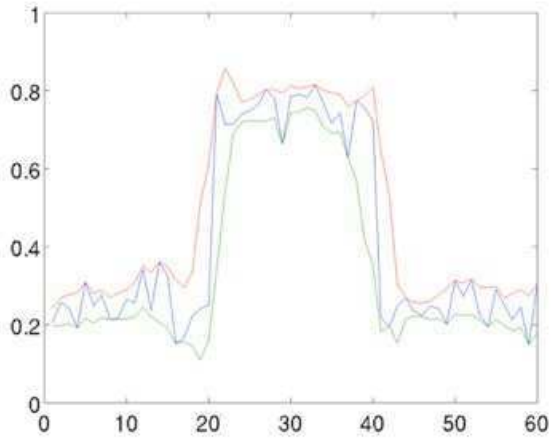


Fig. 4.7 An example of envelopes produced by STRESS. Figure provided and reproduced by courtesy of Ivar Farup.

> "The approximate constancy of the colors of seen objects, in spite of large quantitative or qualitative changes of the general illumination of the visual field, is one of the most noteworthy and most important facts in the field of physiological optics. Without this approximate constancy, a piece of chalk on a cloudy day would manifest the same color as a piece of coal does on a sunny day, and in the course of a single day it would have to assume all possible colors that lie between black and white."

Nowadays, the studies between the physics of light and the visual appearance it produces have defined the following two phenomena, which according to Ebner [71, pp. 1, 16]:

- Color constancy:

Definition 4.14. Color constancy is the phenomenon in which the HVS is able to determine the colors of objects irrespective of the illuminant.

- Lightness constancy:

Definition 4.15. Lightness constancy is the phenomenon in which a surface will appear equally light relative to the other surfaces of the scene, independent of the illumination.

Thus, color constancy enables a stable perception of the scene regardless changes in the color of the illuminant, while lightness constancy enables a stable perception of the scene regardless of changes in the mean luminance intensity [179].

So far, the investigation and the modeling of these phenomena have led to two different approaches (or schools):

- Computer Vision Color Constancy (or Computational Color Constancy) has the goal of separating illuminant from reflectance or alternatively estimating the physical reflectance of objects in different illuminants, or alternatively estimating the illuminant spectral (or colorimetric) component [180, pp. 259, 270].
- Human Vision Color Constancy has the goal to generate appearances that are close to reflectance with perceived stable colors within changes in illumination. Thus, Human Vision Color Constancy neither accurately estimates the light from the scene nor removes the illumination. Furthermore, appearance does not isolate materials from illuminations, according to McCann and Rizzi [180, p. 31]:

Definition 4.16. Appearance is the result of the spatial interactions of gradients and edges.

Although Computer Vision Color Constancy (CV CC) and Human Vision Color Constancy (HV CC) come from the same scientific history, have distinct goals, and thus different kind of outcomes should be expected.

Unfortunately, early experiments proved that appearance was correlating with scaled integrated reflectance [178], and for many years reflectance and appearance, which distinguish the two schools, have been treated as an unique correlated feature. As a consequence, the two different outputs were considered identical, making an incorrect assumption and a weak discrimination between CV CC and HV CC.

Recent studies have shown that human vision color constancy involves the spatial content of the scene, which depends on the arrangement of the scene, the reflectance of the objects, and the spectral content and spatial distribution of the illumination [218, 217]. Thus, correlation between reflectance and appearance holds true only for flat surfaces in uniform illumination [218, 217]. Furthermore, phenomena such as simultaneous contrast (Figure 4.8) show that human appearance does not correlate with reflectance [7].

This is due to the fact that separating the reflectance from the illuminant in the color signal (reflectance · illumination) is a well–known ill–posed problem [13, 15], and as a consequence

the methods belonging to CV CC require constraints or assumptions on the scene content, illumination and/or geometry.

In CV CC, the gloss and the roughness of a surface play an important role, as gloss involves mixed reflection and is responsible for a mirrorlike appearance of the surface [309, p. 52], while the roughness is responsible of different refraction and reflection [212, p. 26]. For a wider overview we address the reader to Oleari [212, pp. 23–26, 341–342] and, Kumar and MacDonald [138].

In Human Vision CC (HV CC) the illuminant component is not totally canceled and responsible of generating appearances that are close to reflectance but with significant departures. These departures are most often not considered in CV CC models and given by many factors (e.g., non–uniform illumination of the scene, glare and gradients).

As a consequence, considering these two approaches as similar has led to simplified assumptions about vision, in some cases useful for CV CC but unrealistic for HV CC. Although they have the common point of introducing a sort of normalization with regard to the illuminant, they clearly differ in the estimated pixel output according to the content of the scene.

As Retinex is a SCA belonging to HV CC, CV CC methods are not subjects of this research work, but the reader may refer to Ebner [71] for a detailed overview.

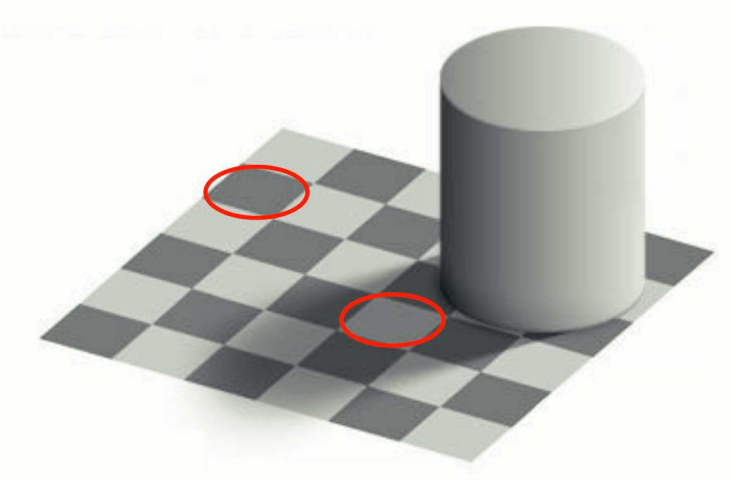


Fig. 4.8 CV CC would report the two grey squares as identical, while HV CC would report different appearance of greys on white and black surrounds respectively. Figure provided and reproduced by courtesy of Alessandro Rizzi (original figure from Ted Adelson).

4.4 Summary

The main purpose of the material presented in this chapter was to give the reader an insight on image enhancement methods and specifically the Retinex–based models. First we gave a definition of image enhancement, and then focused on Spatial Color Algorithms (SCAs). The definition of glocality and the property of dynamic range remapping were as well introduced to show target applications that SCAs can fulfill.

Afterwards, we presented the original Retinex model and the extended family of algorithms, which are the main core of this chapter. From the origins to current days, the various Retinex models/algorithms have been developed with different purposes: some have been implemented for investigating a different glocality (e.g., RSR and STRESS), some for reducing halos (e.g., Meylan and Süsstrunk), some for exploiting specific hardware (e.g., Sobol). Particular attention was given to RSR and STRESS, which will be subject of comparison for the two Retinex proposals of this research work.

As last, we introduced the concepts of color constancy and lightness constancy.

Chapter 5

Computational methods in image processing

In the past decades, with a rapid development of image technologies and a growth of new mathematical theories, a substantial number of novel computing methods and tools have emerged for image processing (e.g., stochastic approaches, Bayesian inference theory, wavelets and variational methods). Variational methods in image processing are of specific interest of this research work and presented in Section 5.1. Following, we introduce the Swarm Intelligence field in Section 5.2. Swarm Intelligence is actually more than a computational method for image processing covering other areas of research, but for convenience is presented in this chapter. Finally, we give a brief overview on how to evaluate computational methods in image processing in Section 5.3.

5.1 Variational image processing

From historical point of view, variational methods have origins in calculus of variations, which according to some manuscripts may date back to studies in ancient Egypt and with the birth of modern mathematics in ancient Greece. Commonly, the calculus of variations is associated with the brachistochrone curve problem raised by Johann Bernoulli in 1696, but the term was first introduced by Euler in 1756 describing a new method for solving geometrical problems proposed by Lagrange in the previous years. In the late modern era calculus of variations was more often called variational calculus and found most of the success in physics [22].

Nowadays, variational methods is a common term since the spread of variational calculus for optimization problems in different fields of research e.g., statistics, economics, control theory, quantum mechanics and image processing, from which comes the term Variational

Image Processing (VIP).

The definition of variational methods may differ slightly from field to field, but share and maintain the key idea that the problem of interest is formulated as an optimization problem.

A unifying concept was proposed by Rockafellar and Wets [254]:

Definition 5.1. The problem solution is the minimizer of an energy functional $E(u)$:

$$\arg\min_{u \in v} E(u), \tag{5.1}$$

with $v = \Re^n$ in the discrete domain and $v = \mathscr{L}^2(\Omega)$ in the continuous domain.

Typical energies consist of a regularization term and a data term:

$$E(u) = R(u) + D(u, f) \tag{5.2}$$

where $R(u)$ is the regularization term, $D(u, f)$ is the data term, f is the input data, and u is the unknown solution. In general, an energy functional is designed such that low–energy states reflect the physical properties of the problem while the minimizer provides the best (in the sense of the model) solution to the problem.

A pioneer work was proposed by Rudin [259] in 1992 for edge preserving image denoising. The variational method, named ROF or TV–L2 is defined as follows:

$$E(u) = \int_\Omega |\nabla l|_2 \, d\Omega + \frac{1}{2\lambda} \int_\Omega (u - f)^2 d\Omega, \tag{5.3}$$

where u is the output desired clean image, f is the input noisy image, Ω is the domain of the image, $\int_\Omega \mid \nabla u \mid d\Omega$ is the Total Variation (TV) term with $\|\cdot\|$ Euclidean norm, and $\frac{1}{2\lambda} \int_\Omega (u - f)^2 d\Omega$ is the non–smooth fidelity term with λ weighting factor for the data attachment. The factor of 2 is incorporated just for computational convenience.

Since its introduction, the TV method has been extended to several areas of image processing (e.g., deblurring, denoising, inpainting, segmentation, resolution enhancement, gamut mapping and image restoration). We refer the reader to Chan and Shen [41] for a detailed overview of TV methods for image processing. Most of the TV methods have been originally developed for greyscale images, but an extension to color images has been proposed by Blomgren and Chan [23] in 1998. This variational method, named Color–TV, is of interest in this

research work and defined as follows:

$$TV_c = E(u_c) = \int_\Omega |\nabla u_c| \, d\Omega + \frac{\lambda}{2} \int_\Omega (u_c - f_c)^2 d\Omega, \tag{5.4}$$

The corresponding Euler–Lagrange Equation is as follows:

$$\nabla \cdot \left(\frac{\nabla u_c}{\| \nabla u_c \|} \right) - \lambda (u_c - f_c) = 0, \tag{5.5}$$

where u_c is the output image solution for channel c, f_c is the input channel c of the original image, $\nabla \cdot \left(\frac{\nabla u_c}{\|\nabla u_c\|} \right)$ is the driving force to the smoothness, and $\lambda(u_c - f_c)$ is the driving force to the data attachment with $0 < \lambda \ll 1$. The terms ∇u_c and $\|\nabla u_c\|$ are the gradient and the gradient magnitude for channel c respectively and defined as follows:

$$\nabla u_c = \begin{bmatrix} \frac{\partial u_c}{\partial x_c} \\ \frac{\partial u_c}{\partial y_c} \end{bmatrix}, \tag{5.6a}$$

$$\|\nabla u_c\| = \left[\left(\frac{\partial u_c}{\partial x_c} \right)^2 + \left(\frac{\partial u_c}{\partial y_c} \right)^2 \right]^{\frac{1}{2}}. \tag{5.6b}$$

There are different ways of calculating the gradient (e.g., forward difference, central difference and Sobel filter) and the choice is usually given by a tradeoff between computational time and accuracy (Figure 5.1).

Before showing how to get to Equation 5.5, we recall the following mathematical principles:

1. Given a vector space $u = (u_0, u_1, ..., u_n)$, the p–norm and its derivative are given as follows:

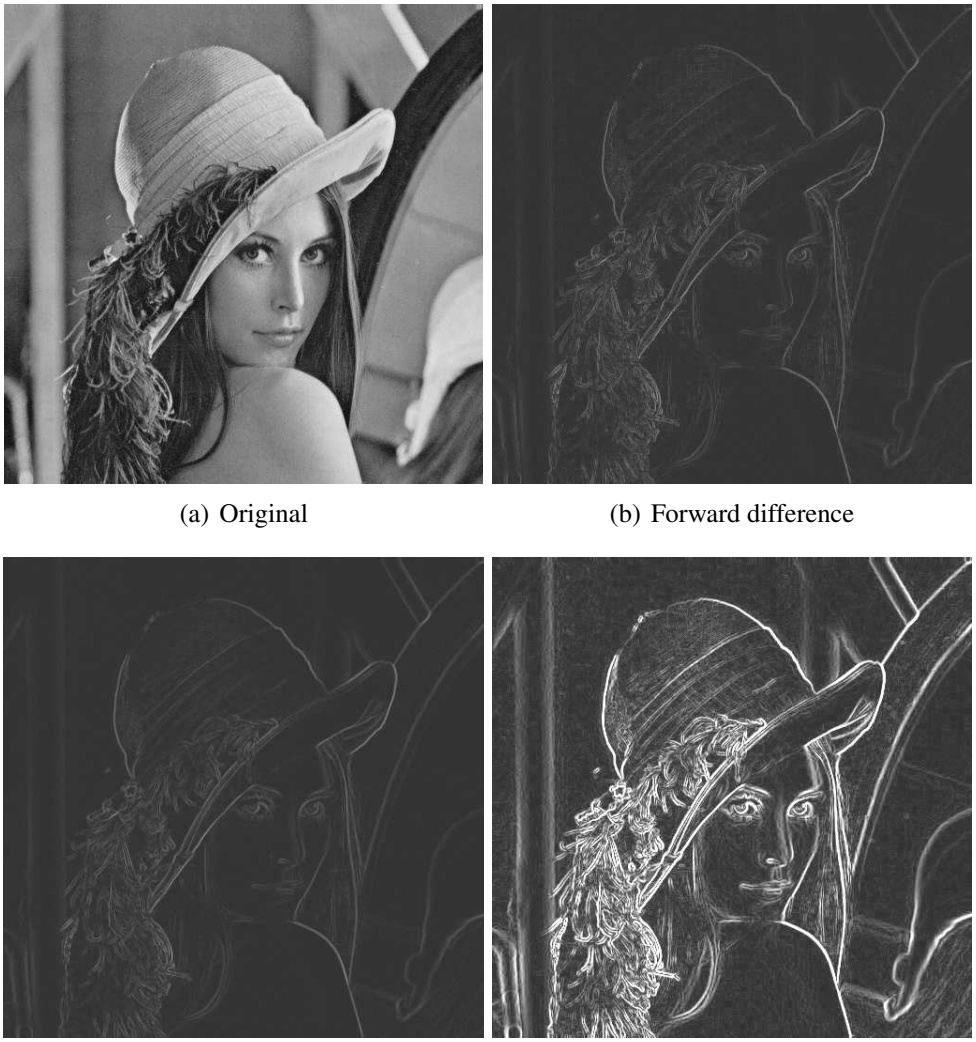
$$\|u\|_p = \left(\sum_{i=1}^{n} |u_i|^p \right)^{\frac{1}{p}}, \tag{5.7a}$$

$$\frac{\partial}{\partial u} \|u\|_p = \frac{u|u|^{p-2}}{\|u\|_p^{p-1}}. \tag{5.7b}$$

   For example for $p = 2$ (Euclidean norm):

$$\frac{\partial}{\partial u} \|u\|_2 = \frac{u}{\|u\|_2}. \tag{5.8}$$

2. A functional is a mapping E which assigns to each element of a vector–space (to each



(a) Original (b) Forward difference

(c) Central difference (d) Sobel filter

Fig. 5.1 Example of norm of the gradient (or gradient magnitude). Original on top left, forward difference on top right, central difference bottom left and Sobel filter on the bottom right, respectively. All images have been boosted in luminosity to make them suitable for a printable version of the manuscript.

function u) an element from the underlying field (a number). A functional in canonical form, called (from physics) as Lagrangian, is defined as follows:

$$E(u) = \int_{\Omega} \mathscr{L}(u, u')dx, \tag{5.9}$$

with:

$$u' = \frac{du}{dx}. \tag{5.10}$$

3. Gateaux derivative: the derivative of the functional $E(u)$ in direction $h(x)$ is defined as:

$$\left.\frac{dE(u)}{du}\right|_h = \lim_{\varepsilon \to 0} \frac{E(u+\varepsilon h) - E(u)}{\varepsilon}. \tag{5.11}$$

The Gateaux differential can be thought as the directional derivative of E at u in direction h, and generally expressed as follows:

$$\left.\frac{dE(u)}{du}\right|_h = \left\langle \frac{dE(u)}{du}, h \right\rangle = \int \frac{dE(u)}{du} h(x)dx. \tag{5.12}$$

A classical term for the Gateaux differential is variation of E, under which is credited the actual term variational methods.

We refer the reader to Chambolle [40] for full proofs of these principles and related theorems.

The central idea for deriving the Euler–Lagrange equation is to compute the Gateaux derivative of E at u in direction h:

$$\left.\frac{dE(u)}{du}\right|_h = \int \underbrace{\left(\frac{\partial \mathscr{L}(u)}{\partial u} - \frac{d}{dx}\frac{\partial \mathscr{L}(u)}{\partial u'}\right)}_{\frac{dE}{du}} h(x)dx. \tag{5.13}$$

The necessary condition for minimality of the functional $E(u)$, the variation of E in any direction $h(x)$ must vanish. Thus at the extremum we have:

$$\frac{dE(u)}{du} = \frac{\partial \mathscr{L}(u)}{\partial u} - \frac{d}{dx}\frac{\partial \mathscr{L}(u)}{\partial u'} = 0. \tag{5.14}$$

In the case of Equation 5.4, for 1–D we get:

$$\mathscr{L}(u, u') = \left|u'(x)\right| + \frac{1}{2}\lambda(u(x) - f(x))^2, \tag{5.15}$$

with:

$$\frac{\partial \mathscr{L}(u)}{\partial u} = \lambda \left(u(x) - f(x)\right), \tag{5.16a}$$

$$\frac{d}{dx}\frac{\partial \mathscr{L}(u)}{\partial u'} = \frac{d}{dx}\left(\frac{\partial}{\partial u'}\left|u'(x)\right|\right) = \frac{d}{dx}\left(\frac{u'(x)}{\|u(x)\|}\right). \tag{5.16b}$$

The final Euler–Lagrange equation back in the discrete domain is then:

$$\lambda \left(u(x) - f(x)\right) - \frac{d}{dx}\left(\frac{u'(x)}{\|u(x)\|}\right) = 0. \tag{5.17}$$

The extension to 2–D is straightforward, and we obtain Equation 5.5.

Back to Blomgren and Chan [23], of particular interest for this research work is the addition of a further regularization term ϕ to be applied at each time iteration (or step) of the Euler explicit time marching scheme, defined as follows:

$$\frac{\partial}{\partial t} = \phi \nabla \cdot \left(\frac{\nabla u}{\| \nabla u \|}\right) - \lambda (u - f) = 0. \tag{5.18}$$

According to the authors, the reason why and whether color channels should be coupled is still unknown, and for this reason they analyze three versions of ϕ, with c identifying the color channel, and C the total number of channels:

1. Channel–by–channel or decoupling:

$$\phi = 1. \tag{5.19}$$

2. Sapiro and Ringach [261] 1996: each color channel at each iteration is multiplied by the square root of the sum of the gradient norm of all channels:

$$\phi = \sqrt{\sum_{c=0}^{C} \| \nabla u_c \|^2}. \tag{5.20}$$

3. Blomgren and Chan [23] 1998: each color channel at each iteration is multiplied by the TV of the channel and divided by the square root of the sum of the TV of all channels:

$$\phi = \frac{TV_c}{\sqrt{\sum_{c=0}^{C} TV_c}}. \tag{5.21}$$

According to their investigation, the third regularization term, which is of particular inter-

est in this research work, provides the following features with respect to the others:

1. It is edge preserving.

2. It takes the form of a global channel–wise scaling. As the TV may wipe out the weaker of the channels, the regularization attempts to maintain a balance of how much each channel is smoothed.

Other techniques of channels coupling (or inter–channel dependencies) can be found in literature, e.g., inspired by Principal Component Analysis (PCA) [260, pp. 325–336]. These methods, widely investigated in multispectral imaging, are not subject of this research work.

In a general overview, variational methods lead to a wide variety of partial differential equations, which can be solved with the following traditional methods:

1. Euler explicit time marching scheme;

2. Euler implicit time marching scheme;

3. Jacobi Method;

4. Gauss–Seidel;

5. Finite Element Method (FEM).

Since all the methods have advantages and disadvantages, a rather large number of variants have been developed in the recent years giving rise to these five families of solvers. However, the performance of a variational method is not only given by the solving method, but also by two important additional factors:

1. Boundary conditions impose the conditions of the independent variable in the equation at the boundaries (or extreme points) of the domain. The two most known are:

   - Dirichlet: specify that the values of the function u are fixed on the boundaries, and thus variations are only inside of the domain:

   $$u(x)\Big|_{\text{boundary}} = \alpha. \tag{5.22}$$

   - Neumann: specify the normal derivative of the function u and allows for variations at the boundaries:

   $$\frac{d\mathscr{L}}{du'}\Big|_{\text{boundary}} = 0. \tag{5.23}$$

2. Choice of a initial solution for u: it can be a set of predefined values (e.g., white), guess values or estimated values (e.g., salt and pepper noise).

The most suitable solving method and as well the boundary conditions and the initial solution are in general identified by the target problem or application. As these solving methods improve iteratively the initial solution of u are called descent methods. According to current literature [41], Neumann boundary conditions seem to be more successful in VIP.

As mentioned in Chapter 2, variational methods have been investigated for measuring contrast. Assuming state that any image can be considered as an isolated thermodynamical system by identifying the image intensity with thermodynamical variables, Ferraro and Boccignone [24, 82] proposed a local contrast measure as follows:

$$\tilde{c}(x,y) = \int_0^\infty I(x,y,t)\sigma_{an}(x,y,t)dt, \tag{5.24}$$

with:

$$\begin{aligned}\sigma_{an}(x,y,t) &= \sigma'_{an}(x,y,t) - \sigma''_{an}(x,y,t) = \\ &= \chi(x,y)\frac{\nabla I(x,y,t)\cdot\nabla I(x,y,t)}{I(x,y,t)^2} - \chi(x,y)\frac{\nabla I(x,y,t)\cdot\nabla I^*(x,y,t)}{I(x,y,t)I^*(x,y,t)},\end{aligned} \tag{5.25}$$

where $\sigma'_{an}(x,y,t)$ represents the density of entropy production during anisotropic diffusion, whereas $\sigma''_{an}(x,y,t)$ tends to prevent entropy production, and χ is a non-negative decreasing function of the magnitude of the local image gradient, which forces convergence of the diffusion process toward some desired image representation, and $*$ denotes the stationary point of the transformation.

As mentioned in the previous chapter, variational methods have been also investigated for the Retinex modeling. Always assuming that an image can be decomposed as a product of the reflectance for the incident illumination intensity $u = L \cdot R$, and in the logarithm domain:

$$\log(u) = \log(L\cdot R) \Rightarrow \log(u) = \log(L) + \log(R), \tag{5.26a}$$

$$s = \log(u), l = \log(L), R = \log(R) \Rightarrow s = l + r, \tag{5.26b}$$

in 2003 Kimmel et al. [130] proposed the following energy functional:

$$E(l) = \int_\Omega |\nabla l|^2 + \alpha(l-s)^2 + \beta\,|\nabla(l-s)|^2\, d\Omega \tag{5.27}$$

where the first term of the energy functional forces spatial smoothness on the illumination im-

age, the second term forces a proximity between l and s weighted by the free parameter α, and the third term forces r to be spatially smooth weighted by the free parameter β, respectively.

5.2 Swarm Intelligence: from natural to artificial ants

Swarm Intelligence (SI) is a branch of artificial intelligence born at the end of the '80s. The term was introduced for the first time by Beni and Wang [19] in 1989 in the context of cellular robotic systems. Today, the term is used to refer to any attempt of problem–solving dealing with the collective behavior of simple agents interacting with the environment and among themselves. From the theoretical point of view, SI models are in general, physically or biologically, inspired by social insect colonies and other natural phenomena. From the computational point of view instead, SI models are largely stochastic search algorithms, which provide metaheuristics for a wide set of combinatorial optimization problems. There are several families of SI models e.g., Bat, Bee Colony, Firefly, Glowworm, Gravitational Search, Particle Swarm, River Formation Dynamics, and the Ant Colony System (ACS) [67], which is of particular interest in this work. For a wider overview of SI models and their applications, the reader can refer to Bonabeaut et al. [25], Eberhart et al. [69] and Lim et al. [152]

The first studies on the development of artificial ants date back to the '90s from the doctoral studies and consecutive works of Dorigo et al. [63, 67], inspired by the observations of Goss et al. [91] and Deneubourg et al. [59], where the interest was to understand how almost blind animals like ants can manage to establish shortest route paths from their colony to feeding sources and back. Biological studies have shown that social insects coordinate their activities via "stigmergy", a form of indirect communication mediated by modifications of the environment. In the case of ants the communication among individuals or between individuals and the environment, is based on the production of a chemical substance called pheromone.

Inspired by these behavioral properties of the ants, Dorigo et al. developed the so called Ant Colony System (ACS) for the Traveling Salesman Problem (TSP) [65, 64]. Today, their pioneer work is known as Ant Colony Optimization (ACO) model and it has been diversified to solve a wider class of numerical problems [68]. Since this SI model is relevant for this work, we recall it briefly in the next subsection.

5.2.1 Ant Colony Optimization for the Traveling Salesman Problem

The TSP is an NP–hard problem in combinatorial optimization and theoretical computer science, where [133, p. 474]:

Definition 5.2. Given a list of cities and their pairwise distances, the task is to find the shortest possible tour that visits each city exactly once.

Three ideas from the natural behavior of the ants are inherited to the artificial ant colony:

1. The preference for paths with a high pheromone level.
2. The higher rate of growth of the amount of pheromone on shorter paths.
3. The trail mediated communication among ants.

An artificial ant k in city r chooses the city s to move to among those which do not belong to its working memory M_k by applying the following probabilistic formula [66]:

$$p_k(r,s) = \begin{cases} \frac{(\tau_{r,s})^{\alpha}(\eta_{r,s})^{\beta}}{\Sigma_{u \notin M_k}(\tau_{r,u})^{\alpha}(\eta_{r,u})^{\beta}} & \text{if } s \notin M_k \\ 0 & \text{otherwise} \end{cases}, \tag{5.28}$$

where $\tau_{r,u}$ is the amount of pheromone trail on edge (r,u), $\eta_{r,u}$ is a heuristic function called visibility, which is the inverse of the distance between cities r and u and, α and β are parameters that allow a user to control the importance of the trail versus the visibility. The memory M_k is the tabu list of the k^{th} ant, which contains the cities that it has already visited. City s is inserted in this list when the ant transits from city r to city s. The choice criteria of the parameters α and β can differ widely according to the problem for which the ACO is used. A guideline on how to choose the values of the different parameters for the TSP problem can be found in Dorigo and Gambardella [64].

As artificial ants build and travel on artificial paths, ACO will be candidate of this research work. ACO will be used in conjunction with the original Retinex model to build a novel engine for generating paths across the content of the image. In particular, Equation 5.28 will be properly modified to constitute a new probabilistic formula for choosing a new local reference white pixel along a path, and then rescaling the appearance of a point of interest through the MI–Retinex spatial reset mechanism.

5.3 Evaluating image processing methods

At current writing of this research project, there is no universal method to evaluate computational methods in image processing (e.g., SCAs or image processing frameworks). For this

reason, the analysis for evaluating computational methods in image processing will be handled in a wider context.

First we introduce the definition of Quality of Experience according to Qualinet [146]:

Definition 5.3. Quality of Experience (QoE) is the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state.

In terms of evaluation of an image, the degree of delight in QoE is analogous to a measure of how pleasing an image appears. Instead, in terms of evaluation of a SCA, if we consider a SCA like an aftermarket product purchased by a user/customer, the degree of delight in QoE is analogous to the satisfaction of his needs and expectations for a target application. Potential users, above all non–experts, prefer a fast and easy to use product.

This issue leads us to identify four common variables of evaluation:

1. Efficiency (or computational complexity): according to a definition proposed by Kleinberg and Tardos [133, p. 33]:

   Definition 5.4. An algorithm is efficient if it has a polynomial running time.

   The efficiency (or computational complexity) is indicated with $O = f(n)$, where n is the input size of the problem. In the field of image processing, the whole processing of one pixel independently by the number of effective operations is considered one fundamental operation with $O = 1$. As a consequence the polynomial $f(n)$ identifies the complexity of processing a whole image of size n (total number of pixels). From now on, for convenience, in this research work we will adopt only the term computational complexity.

   Choosing an algorithm with lower computational complexity may lead to lower computational time. However, non–experts do not always rely on the evaluation of this variable because hard to analyze.

2. Computational time (or time of execution): the time required to process the whole image.

Computational time depends on many factors e.g., language of implementation, memory consumption, parallelism involved and noteworthy the resolution of the processed image. Ability to parallelize computation e.g., for channel or for level or for pixel may be an important factor [292].

Computational time is the most important variable especially when dealing with Real–Time Systems. Real–Time Systems for image processing are not the subject of this work, but we underline to the reader that the interest of researchers in computational methods for real–time applications (e.g., vehicle rear camera enhancement) cover a wide area. The reader may refer to Kim and Shim [129].

3. Number of input parameters: is often underestimated, although facilitating the use of the algorithm. Non–experts are not keen on many parameters, and thus rules of thumb and automatic adjustment are often required.

4. Perceptual attributes: according to Pedersen [222] there is a wide set of attributes that can be used to evaluate the quality of either an original image and/or one of its reproductions (see Chapter 3. After processing an image, one or more of the following attributes are usually taken into account:

   - Visibility;
   - Brightness;
   - Overall and local contrast;
   - Saturation;
   - Perceived colors;
   - Halos;
   - Noise;
   - Naturalness.

An exhaustive review of computational methods in image processing employing all the variables above is not available. Furthermore, unlike image quality metrics, experiments and statistical analysis like the ones described in Section 3.4 are seldom presented. The reader may find noteworthy information in Ebner [71, pp. 275–301], Moroney and Tastl [198] and, Rizzi and McCann [180].

According to Rizzi and McCann [180, p. 370], the so called beauty contest is a widely used method to record observers preferences in terms of perceptual attributes. As an example, Figure 5.2 shows an original image, which is followed by processed images having a short

selection of Retinex algorithms, ACE and iCAM06. In this scenario, the beauty contest might claim that Figure 5.3(e) processed by MSR has the best visibility, while Figure 5.3(i) processed by ACE has the best perceived colors. To have a valuable comparison among algorithms and identify the one with the highest performance, the beauty contest should take into account a rather large number of images and observers [180, p. 370][48, 102].

The other important three variables aforementioned should also be taken into account. Table 5.1 presents the computational complexity and the number of input parameters of the short selection of Retinex algorithms, ACE and iCAM06. Frankle and McCann, McCann99, Constrained Pseudo–Brownian Motion and PDE–Retinex show the smallest number of input parameters (with one). On the other hand, MSR, RSR, STRESS and iCAM06 show the largest number of input parameters. ACE and iCAM06 have the highest computational complexity, while Frankle and McCann has the lowest one. A comparison for computational time is not currently possible due to different implementation languages (e.g., Matlab and C++) and different hardware testing. According to the number of input parameters and the computational complexity, Frankle and McCann could be one of the best choices.

In conclusion, we want to underline with this example that the final choice of an image processing method requires a certain tradeoff among these variables pondered by the target application.

Table 5.1 Comparison of number of input parameters and computational complexity among a selection of Retinex, ACE and iCAM. Authors have not clearly depicted these variables, hence making Table 5.1 only indicative. The computational complexity and the number of input parameters were estimated by the author following the mathematical description of the algorithm and the analysis of the source code if available.

Name (Authors)	Year	Family	Number of input parameters	Computational Complexity
Frankle and McCann (Frankle and McCann [83])	1983	Scale–based	N=number of iterations	$O(n)$
MSR (Jobson et al. [123], Rahman et al. [244, 245])	1996/1997	Scale–based	r= Gaussian surround radius, N=number of levels, w= levels weights	$O(n \log n)$
McCann99 (McCann [175], Funt et al. [84])	1999	Scale-based	N=number of iterations	$O(n \log n)$
RSR (Provenzi et al. [241])	2007	Random–based	rho=spray density, N=number of iterations, M= number of samples in the spray	$O(NMn)$
STRESS (Kolås et al. [135])	2011	Random–based	rho=spray density, N=number of iterations, M= number of samples in the spray	$O(NMn)$
Constrained Pseudo–Brownian Motion (Montagna and Finlayson [191])	2011	Random–based	k=number of visits per pixel	$O(n \log n)$
PDE–Retinex (Morel et al. [195],Limare et al.)	2010/2011	Variational	t=threshold	$O(n \log n)$
ACE (Gatta et al. [86])	2002	No Retinex	slope=chromatic comparison, α=distance weight	$O(n^2)$
iCAM06 (Kuang et al. [136])	2006/2007	No Retinex	L=maximum luminance, p=contrast, F=surround adjustment	$O(n^2)$

5.4 Summary

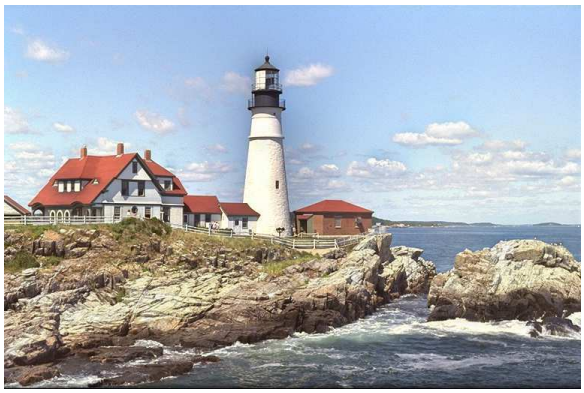
The main purpose of the material presented in this chapter was to give to the reader an insight on on Variational Image Processing (VIP) and Swarm Intelligence (SI). For VIP, after giving a general concept of variational method, we turned focus on the TV–L2 method and the corresponding Euler–Lagrange Equation. We briefly introduced methods for solving partial differential equations, and on how to apply regularization terms. As well we presented the Dirichlet and Neumann boundary conditions.

For SI, we gave instead attention only on the Ant Colony Optimization (ACO) for the Traveling Salesman Problem (TSP).

As last, we described four common variables to evaluate computational methods in image processing: computational complexity, computational time, number of input parameters and perceptual attributes.

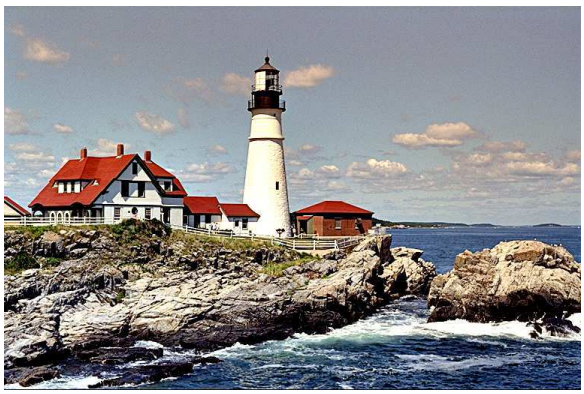
(a) Original

(b) Frankle and McCann



(c) McCann99

(d) Constrained Pseudo–Brownian Motion (Montagna and Finalyson)



(e) MSR (Jobson et al.)

(f) PDE–Retinex (Limare et al.)

(g) RSR (Provenzi et al.)

(h) STRESS (Kolås et al.)

(i) ACE (Gatta et al.)

(j) iCAM06 (Kuang et al.)

Fig. 5.2 Examples of SCA rendering for an indicative beauty contest among a selection of Retinex, ACE and iCAM. Figures (a)–(d) provided and reproduced by courtesy of Roberto Montagna. Original Lighthouse image (Figure (a)) is from the Kodak dataset [134]. Figure (e) reproduced with GIMP plug–in. Figure (f) reproduced from Limare online demo at http://www.ipol.im/pub/art/2011/lmps_rpe/. Figures (g) and (i) provided and reproduced by courtesy of Alessandro Rizzi. Figure (h) provided and reproduced by courtesy of Ivar Farup. Figure (j) reproduced with source codes at http://www.cis.rit.edu/research/mcsl2/icam06/.

Chapter 6

Summaries of included papers

6.1 Paper A: Measuring perceptual contrast in digital images

In the first paper, we present a novel contrast measure for digital images based on the previous work of Rizzi et al. [251], which presents a multilevel analysis and the Difference of Gaussians model developed by Tadmor and Tolhurst [280]. This measure, named Weighted–Level Framework, performs for each channel a pyramid subsampling of the image to various levels. A pyramidal structure is created halving the image at each iteration with prefiltering in order to avoid aliasing at low resolutions. For each level, it calculates the local contrast of each pixel using the Difference of Gaussians, thus obtaining a contrast map of each level. The overall measure of each channel is a weighted recombination of the average contrast for each level. The final global measure is given by a weighted sum of the contrast of each channel.

A psychophysical experiment in a controlled environment with 15 images and 17 observers show an improvement in correlation between measured contrast and observers perceived contrast when the variance [227] of the three color channels separately is used as weighting parameters for each local contrast map, and as final weight of each channel. Furthermore this measure keeps the same computational complexity of other state–of–the–art contrast measures but having higher performance in correlation with observers perceived contrast.

Afterward this measure is combined with Regions–of–Interest as weighting maps but results show that they do not improve the ability of contrast measures to predict perceived contrast in digital images suggesting that contrast is an intrinsic factor, which is judged by the global impression of the image.

6.2 Paper B: Measuring perceptual contrast in uncontrolled environments

In this second paper, the contrast measure previously developed is tested in uncontrolled environments. A web–based psychophysical experiment with the same 15 images plus their respective greyscale version has been carried out in order to register observers perceived contrast. An average of 170 ratings for each image was recorded. The results from the observers indicate that color images are rated higher than their respective greyscale ones indicating that contrast is influenced by color. A two–sided sign test at 95% confidence interval confirms this hypothesis.

A comparison with a lab controlled environment experiment shows that most of the images had higher mean score in the online experiment but with 95% confidence interval they are not significantly different. However, a decrease in correlation of the previously proposed measure can be noticed and a comparison of the correlation coefficients indicate that measuring contrast in uncontrolled environments can be significantly different at 99% than measuring in lab controlled environment.

6.3 Paper C: Multi–level contrast filtering in image difference metrics

In this third paper, we present a novel image difference metric to estimate the perceived difference in contrast between an original image and a reproduction.

This metric, named Weighted–Level Framework ΔE_E (WLF–DEE), implements two key features: a multilevel filtering based on the difference of Gaussians model proposed by Tadmor and Tolhurst [280] and the new Euclidean color difference formula in log–compressed OSA–UCS space proposed by Oleari et al. [213].

Analysis and comparisons with other state–of–the–art image difference metrics, such as PSNR, PSNR–HVS–M, S–CIELAB, SSIM and VSNR, are presented on two databases providing different distortions directly related to color and contrast, and on four different categories belonging to the well–known Tampere Image Database (TID–masked noise, TID–quantization, TID-image denoising and TID-contrast change).

Results show stability of WLF–DEE on all six datasets and particular strength on those databases where a change in contrast between the original and its reproduction is triggered by a change of color attributes.

6.4 Paper D: Spatio–temporal Retinex–like envelope with total variation

In this fourth paper, we present a new spatial color algorithm called STRETV (Spatio–Temporal Retinex–like Envelope with Total Variation). This algorithm derives from the approach of Kolås et al. [135] for the STRESS algorithm (Spatio–Temporal Retinex–like Envelope with Stochastic Sampling), which adjusts each pixel calculating the local reference of lighter and darker points in each channel. This is done estimating two envelope functions, the maximum and minimum envelopes, containing the image signal. STRETV replaces the stochastic sample for the calculations of the two envelope functions with the total variation method.

A first psychophysical experiment on 13 images with 20 observers confirm the efficacy of the method with a success of STRETV in comparison to the original. A second psychophysical experiment shows that the overall performance of STRESS is not significantly different at 95% confidence level but at the same time a higher preference of the observers of STRETV with respect to STRESS is reported due to a lower perception of noise.

Results of contrast enhancement and automatic color correction are presented as application of the model.

6.5 Paper E: Termite Retinex: A new implementation based on a colony of intelligent agents

In this fifth paper, we present a novel implementation of Retinex, reconsidering the idea of the paths and using an existing swarm intelligence model inspired from a biological process. This algorithm takes the name Termite Retinex (TR) and grows out from the modification of the original presentation of Retinex proposed by Land and McCann [139, 144] in conjunction with the Ant Colony System (ACS) model proposed by Dorigo et al. [67]. In this case TR does not present itself as a novel metaheuristic for optimization of some constraints but as a novel method for the exploration of the image content, tuned in particular by two parameters, α and β which weight the relative importance of the so–called "poison" and of the so–called "closeness".

Two experiments with users are presented in order to evaluate the quality of TR. A first experiment on nine images with 20 observers confirms the efficacy of the method with the images processed by TR significantly better than the original. A second experiment shows that the overall performance of TR is higher than Random Spray Retinex, an implementation with random sampling. A sign–test at 95% confidence interval confirms both statements.

Applications as color correction, low dynamic range maximization and high dynamic range tone–rendering are presented as results.

Chapter 7

Discussion

After each paper has been briefly summarized in the previous chapter, the discussion in context is here presented. Conclusions and future work will be presented in the next two chapters, respectively, However, some concluding remarks and future directions will be discussed here to drive the reader through the leading idea of this research work.

In order to explain some choices and discuss advantages and disadvantages of this research work, we will recall each of the five research questions and discuss the contribution in light of of these research questions.

7.1 Q1: Can we improve the Rizzi et al. measure of contrast using the Difference of Gaussians model instead of using a weighted 8–neighborhood?

In this research work, we present a novel measure of contrast for digital color images in **Paper A**. The measure developed takes the name Weighted–Level Framework (WLF) and is based on the following four key features:

1. The multilevel analysis as a tool to investigate complex scenes at different spatial frequencies [83, 2, 251].

2. The HVS to be more sensitive at higher frequencies [101, 230].

3. Chromatic information plays a fundamental role in contrast perception [36, 37, 228].

4. Variance has shown to be a good indicator of contrast [228].

WLF encapsulates a multilevel analysis based on the previous work of Rizzi et al. [251] in conjunction with the the Difference of Gaussians model developed by Tadmor and Tolhurst [280]. Additionally, WLF is extended in calculation to chromatic channels and performs a weighted recombination of the levels.

According to Section 2.2, local contrast measures can be categorized in biological inspired and engineering ones. Although WLF encapsulates the DOG model, which is indeed biologically inspired, should be categorized as an engineering local measure of contrast. WLF has been developed with the aim to produce a contrast number, easily interpretable by a user and hopefully being in correlation with observers' perceived contrast. A performance of 0.84 in Pearson correlation and of 0.80 in Spearman correlation is achieved in relation to observers' perceived contrast on ad–hoc database of 15 color images. On the fact, the results presented in **Paper A–Table 8** show a higher correlation of WLF compared to Peli and, Tadmor and Tolhurst measures. Confidence intervals for Pearson and Spearman correlations (Figures 7.1 and 7.2) show WLF having the same performance as Peli band–limited and Rizzi et al. (RAMM) contrast measures, and to be significantly better than Tadmor and Tolhurst (TT) measure.

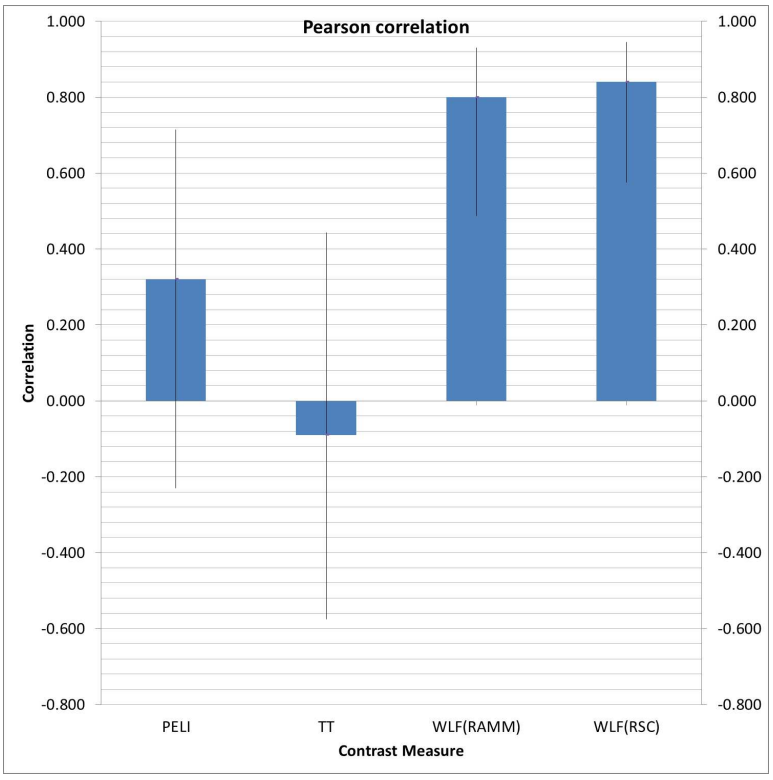


Fig. 7.1 Pearson correlation and confidence intervals for the selected contrast measures.

The weighted recombination of the levels using variance increases the correlation of WLF

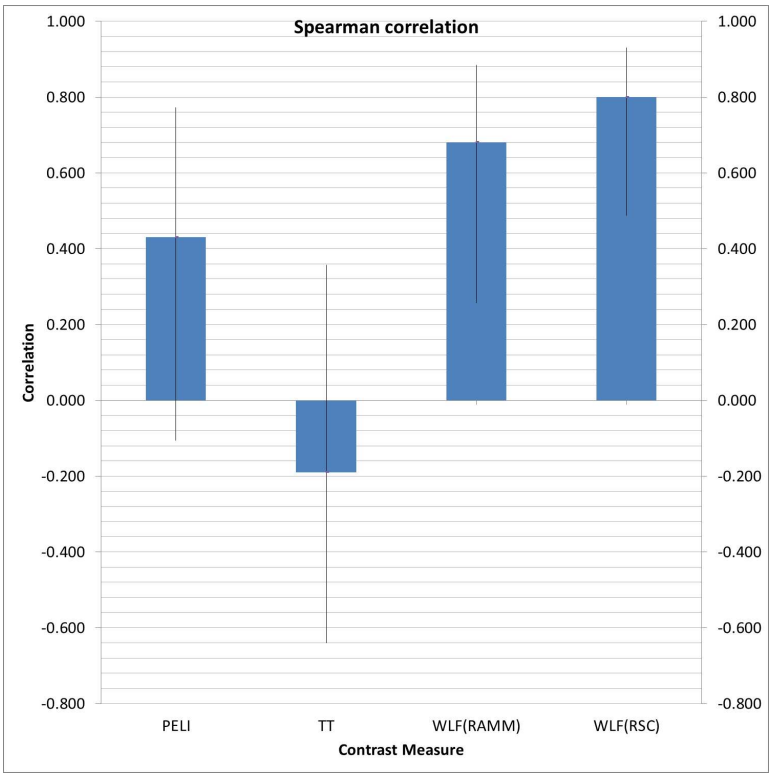


Fig. 7.2 Spearman correlation and confidence intervals for the selected contrast measures.

with observers' perceived contrast and seems to:

1. Follow the non–linearity of the HVS.
2. Be well adapted for the emulation of the complex phenomena of the HVS of giving different importance to different frequencies.

The two aforementioned statement do not find any biological proof [18], and thus WLF should be considered an engineering contrast measure. Furthermore, variance in the RGB color space, which is device dependent, does not take into account those geometrical transformations that rearrange the spatial distributions of the scene and can be deceived as shown in Figure 7.3. Nevertheless, WLF has found application in measuring the contrast of projection displays [318].

Fig. 7.3 On the left image the original, while on the right image the same image with the red channel rotated 180 deg counterclockwise, the green channel as the original, and the blue channel horizontally reflected. The three color channels of the two images are processed independently in RGB color space. Thus, the two images have the same variance for each level and as final contribution for each channel, but different perceived contrast.

Another characteristic of WLF in common with other engineering contrast measures is the discard of the use of CSFs. The use of CSFs, although in many papers not explicitly stated, have always instigated a sort of criticism. Recent studies from Haun and Peli [98] claim that:

> "... the specific sensitivity functions commonly used in standard practice to estimate perceived contrast and quality may be misapplied or inappropriate. ... summation of (suprathreshold) contrast magnitudes over the entire range of a CSF is not necessarily indicative of contrast of the image."

The literature on CSFs is wide and earlier than Peli many other works claim that the use of CSFs in many models intended to evaluate visual properties might be questionable or even inappropriate.

1. Fairchild [79, p. 29]

   > "It is important to realize that the (CSFs) functions are typical and not universal. ... These functions also depend on other variables such as luminance level, stimulus size, retinal locus."

2. Allen and Triantaphillidou [8, p. 74]

   > "Whilst the CSF for the eye expresses its ability to resolve individual spatial frequencies it gives no intuitive information about its ability to perform basic tasks or performance against the average population."

3. Shanda [264, p. 298]

   > "The spatial preprocessing uses separable convolution kernels to approximate the contrast sensitivity functions (CSF) of the human visual system. These kernels behave as bandpass functions on the "luminance" channels and low–pass functions for the chromatic opponent channels. It is important to stress that the CSF kernels chosen are tied heavily into the color space in which they are applied."

4. Daly [56]

   > "The contrast sensitivity function (CSF) describes the variations in visual sensitivity as a function of spatial frequency. ... CSF changes as a function of light adaptation, noise, color, accommodation, eccentricity, and image size."

5. Valberg [288, p. 190]

   > "Increasing age leads to reduced sensitivity for the higher spatial frequencies, while sensitivity to lower frequencies is unchanged."

All the aforementioned statements include different variables (e.g., luminance level, color and age) highlighting the high variability of individual CSFs supporting Peli's statement. A review of several measured CSFs can be found in Barten [16, pp.38–56].

Nevertheless, the use of CSFs can be still used for targeting different specific applications. For example, Pedersen and Farup [223] proposed in 2012 the use of wavelet decomposition in conjunction with CSFs filtering for the simulation of image detail visibility. The results are presented in the context of image quality.

Paper A–Table 4 presents correlation results of WLF in CIELAB and RGB color space, leaving out the CIE color appearance models. The CIE color appearance models, named

CIECAM97s and CIECAM02, are well–known, but some authors have written important papers with a deep criticism on the mathematical aspects of these models during the development of WLF [30, 33, 31].

Lee [149, pp. 447, 462] states:

1. "... current color appearance models are empirical in nature, do not account for the spatial processing involved in color vision, and the model parameters have to be tuned by trial and error."

2. "We should give a warning for engineers and scientists who are used to working on firm theoretical foundations. The color appearance model to be discussed will strike you as a mess of empirical equations, with no justification whatsoever."

3. " They represent a justifiable effort to summarize a huge amount of experimental data with a few equations. They do not look trustworthy only because it would have taken a few hundred pages to list all the relevant experimental data and to explain how the equations were derived to fit those data. In a sense, this is a desperate effort to make those data useful for people who do not have the time to go through the same derivation process, i.e., for those people who simply want to have a blackbox machine to compute the physical correlates of some of the important perceptual judgments, such as brightness, lightness, hue, chroma, and colorfulness."

The drawbacks presented by color appearance models led CIE to start a new Technical Committee, the CIE TC 8–11, for solving these mathematical inconsistencies (http://www.cie.co.at/index.php/Technical+Committees). After eight years of investigation, the CIE TC 8–11 closed down in 2015 without giving a final solution to the problem.

However, a solution formulated as non–linear optimization problem has been recently proposed by Jiang et al. [121] and can be the subject of future studies. At the same time, other authors have addressed the problem of formulating color appearance with a different approach and can be as well subject of future research [32, 211]. Waiting for a final answer, the authors have thought wisely to leave out color appearance models when WLF was developed.

Paper A–Table 4 presents different weights for λ, α, β and γ, and the reader first impression is that the results have been retrieved with a brute force search.

Wang et al. [300] claim that a training and a test datasets are required to retrieve optimal weights when using brute force like search methods. For the sake of our experiment we regret not having applied Wang. et al [300] methodology.

The relevant results from **Paper A–Table 8** have induced the authors to three further investigation:

1. Combine WLF with Regions–of–Interest (still in **Paper A**).
2. Test WLF in uncontrolled environments (**Paper B**).
3. Extend WLF to image quality (**Paper C**).

According to **Paper A–Tables 8–9** , Regions–of–Interest can be applied for contrast measuring, but they are not helpful for WLF to predict perceived contrast. In our use case, considering the ROIs extracted with saliency [115, 291] and GAFFE [246], too many important areas of the image are erased and thus weighing zero. Regarding gaze information recorded with eye tracking data [11, 12], we consider again that the recording time was too short, and thus valuable areas of the image used by the observers to judge contrast were erased.

Nevertheless, Regions–of–Interest find success in other area of research, e.g., image quality [220, 28, 248, 85].

7.2 Q2: Are there significant differences between assessing contrast in uncontrolled environments and a lab controlled environment?

In order to answer to this research question, in **Paper B**, WLF is tested in uncontrolled environments and resubmitted to validation. A web–based psychophysical experiment with the same 15 images of **Paper A** plus their respective greyscale version is presented.

According to **Paper B–Table 1**, WLF shows a substantial decrease in correlation. A comparison of the correlation coefficients indicate that measuring contrast in uncontrolled environments can be significantly different at 99% than measuring in lab controlled environment.

The invalidation of the results of **Paper A** raises different shortcomings in the online experiment. According to the guidelines given by Zuffi et al. [319].

1. Time of evaluation: unlimited in the online, only 40 seconds in the lab experiment. This variable has given to the observer a different constraint to properly inspect the whole regions and subregions of the image. A so short time in the lab experiment might have led non–experts to a “quick guess score”, without a clear evaluation of contrast.
2. No control of visual imparity or dichromatism: has left the opportunity to ineligible observers to participate. This could have been prevented with a simple color–blindness pre–test.

3. No distinction between experts and non–experts: could have helped us to differentiate again the results and understand whether there is a true difference in terms of contrast evaluation between these two categories.

   This issue was actually addressed in **Paper A–Table 2**, and according to our results expert observers have a better understanding of contrast. However, contrast studies should take into account a whole population, but with the important concern of not including observers with visual deficiencies (e.g., dichromatism).

4. No control of environment and participant conditions: time of the day, location, light in the room, age of the participant, use of eyeglasses, size of the monitor and distance from the screen might have led to improper evaluations. A questionnaire before participation might have helped to discover outliers and discard them.

These above issues raises awareness about a stricter control on certain variables, e.g., the type of participants, a visual impair subject should have been excluded in performing the experiment.

Paper B–Figure 1 indicates that observers rated the color images higher than their respective greyscale ones. A two–sided sign test at 95% confidence interval confirms this hypothesis. A detailed example is shown in Table 7.1 for images 41 and 24 of **Paper B** shown in Figure 7.4. Nine of the 10 first participants rated higher contrast for the color image with respect to his corresponding greyscale one, resulting in an average contrast of 7.9 for image 41 and 6.9 for image 24. The trend continues in the later scores, and after 150 ratings the average is 7.2 for image 41 and 5.9 for image 24, respectively.

Contrast is then influenced by color, and thus chromatic channels must be taken into calculation. Contrary to WLF, color–blind contrast measures relying only on luminance (or lightness) computation will output the same contrast number, failing in predicting observers perceived contrast.

Table 7.1 First 10 participants contrast rate for image 41 and 22, respectively. Only Participant 5, highlighted in red, gave a low score for the color image and rated slightly lower with respect to the corresponding greyscale one.

Participant	1	2	3	4	5	6	7	8	9	10
Image 41 (Color)	9	10	6	9	4	9	10	7	6	9
Image 24 (Grey)	6	9	5	7	5	5	6	5	5	8

Finally, wrapping up the results of **Paper A** and **Paper B**, the reader may question if the concept of contrast can be summarized in a unique definition. After proposing a new contrast

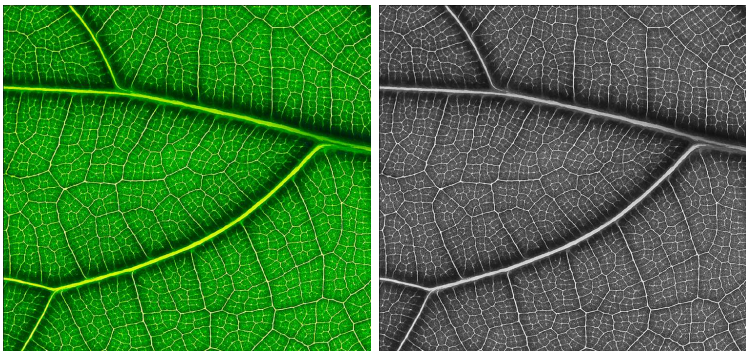


Fig. 7.4 Images 41 and 24 of **Paper B** on the left and on the right respectively.

measure taking into account all the three definition of contrast in Chapter 2 and being in relation with observers' perceived contrast, in the context of this research work, contrast can be defined with the following new proposal:

Definition 7.1. Contrast is the difference in appearance of the varying distribution of light and chromaticity between two or more parts of a scene that make objects distinguishable simultaneously.

7.3 Q3: Can we estimate the quality of a reproduction image with respect to an original image in terms of contrast using the Difference of Gaussians model?

To describe the quality between an original and a reproduction in terms of perceived magnitude of contrast, the authors have been motivated to extend WLF and develop an image difference metric in **Paper C**. An additional ingredient to encapsulate in the model was required: a color difference formula in order to perform a pixelwise difference among each level.

This metric, named Weighted–Level Framework ΔE_E (WLF–DEE), implements two key features:

1. The same multilevel filtering presented in **Paper A**, based on the difference of Gaussians model proposed by Tadmor and Tolhurst [280].

2. The new Euclidean color difference formula in log–compressed OSA–UCS space proposed by Oleari et al. [213].

According to CIE, the last recommended formula is the CIEDE2000, which has superseded the previous ones, although the ΔE^*_{ab} is still commonly used. The CIEDE2000 has several input parameters, which should have been subject of study and thus complicating the WLF–DEE investigation. Recent studies from Pant and Farup [216] shows that there is no significant difference between the Riemannized ΔE_{00} and the ΔE_E at small color difference.

The choice of the ΔE_E is from the well–known Ockham's razor that states:

> "When you have two competing theories that make exactly the same predictions, the simpler one is the better. William Ockham (1285–1349)"

According to the results presented in **Paper C–Tables 2–4**, there is no a clear winning metric among the investigated datasets. Results have been investigated using non–linear regression in order to remove any non–linearities due to the subjective experimental process and to facilitate comparison of the metrics in a common analysis space. Performances in Pearson correlation of 0.68, 0.77, 0.71 and 0.67 are achieved respectively for masked noise, quantization noise, image denoising, and contrast change categories of the TID database. Confidence intervals (Figures 7.5, 7.6, 7.7 and 7.8) show WLF–DEE having the same performance as SSIM and S–CIELAB, and to be significantly better than PSNR and VSNR on the four categories of the TID database. Performances of 0.65 and 0.49 in Pearson correlation are found on two databases (Ajagamelle and Pedersen) providing different distortions directly related to color and contrast. Confidence intervals (Figures 7.9 and 7.10) show WLF having the same performance as other state–of–the–art metrics on both databases except for S–CIELAB on Pedersen database, while the latter has slightly higher performance. Root–Mean–Square Error confirms WLF–DEE to be in line with other state–of–the–art metrics in estimating the perceived magnitude of contrast between an original and a reproduction. For the sake of completeness of WLF–DEE with respect to WLF, we have presented the confidence intervals as in Paper C (Section 10.3).

Briefly summarizing, we can say that SSIM has the highest correlation on the four categories of the TID database, PSNR on Ajagamelle database, and S–CIELAB on Pedersen database, but the confidence intervals show that WLF–DEE is not significantly different in performance with respect to the others.

However, SSIM and PSNR work only on greyscale images, whereas WLF–DEE works with color images (Figure 7.4). Furthermore, WLF–DEE has been developed with the intent of calculating image difference in terms of contrast and according to our results the metric has

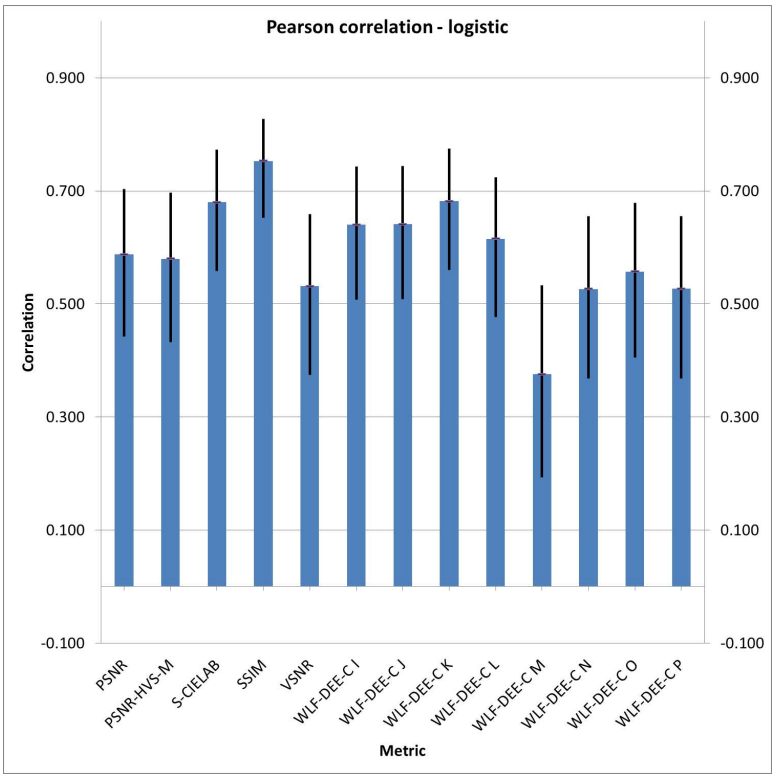


Fig. 7.5 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on TID–masked noise database.

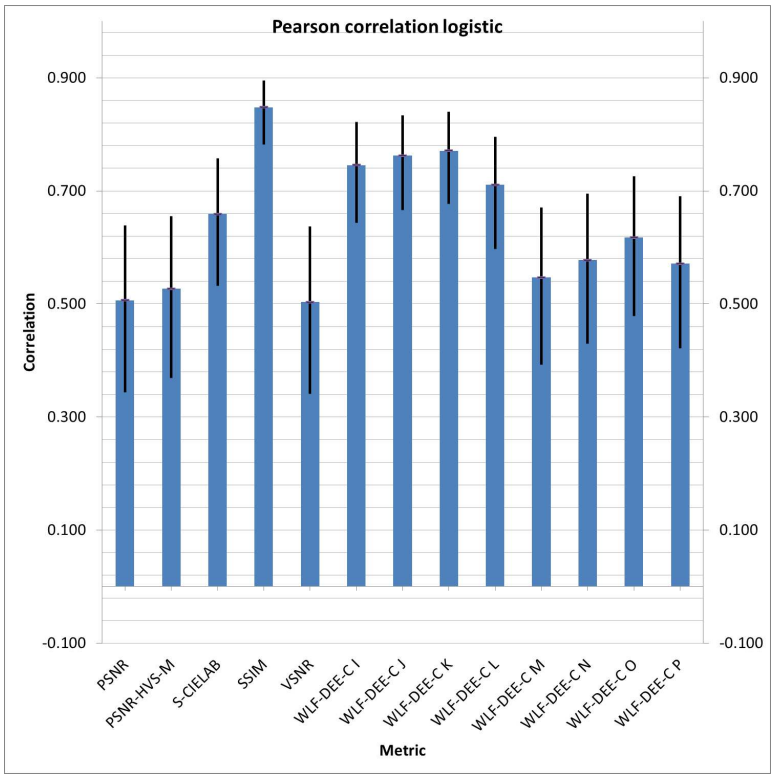


Fig. 7.6 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on TID–quantization noise database.

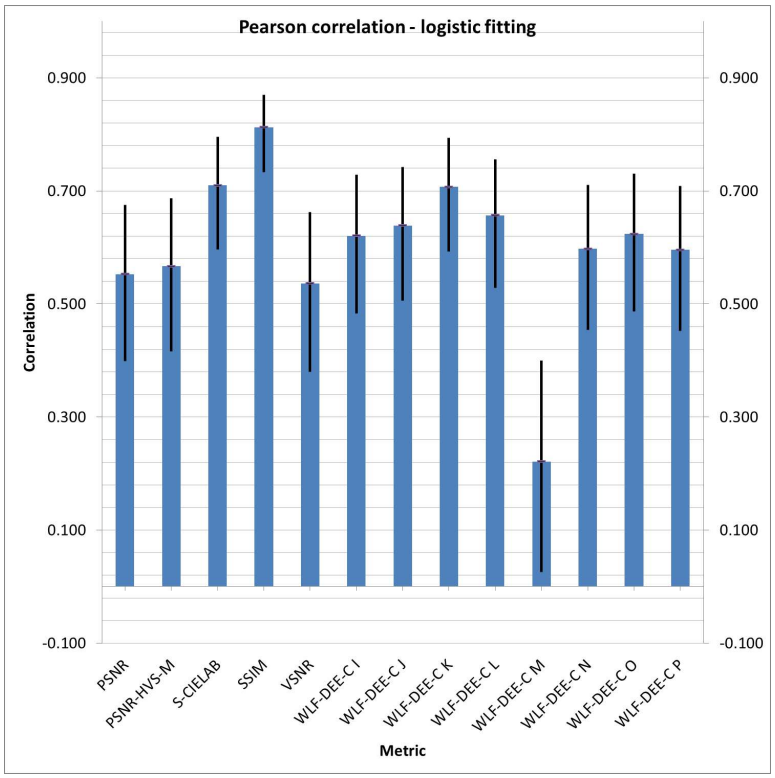


Fig. 7.7 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on TID–image denoising database.

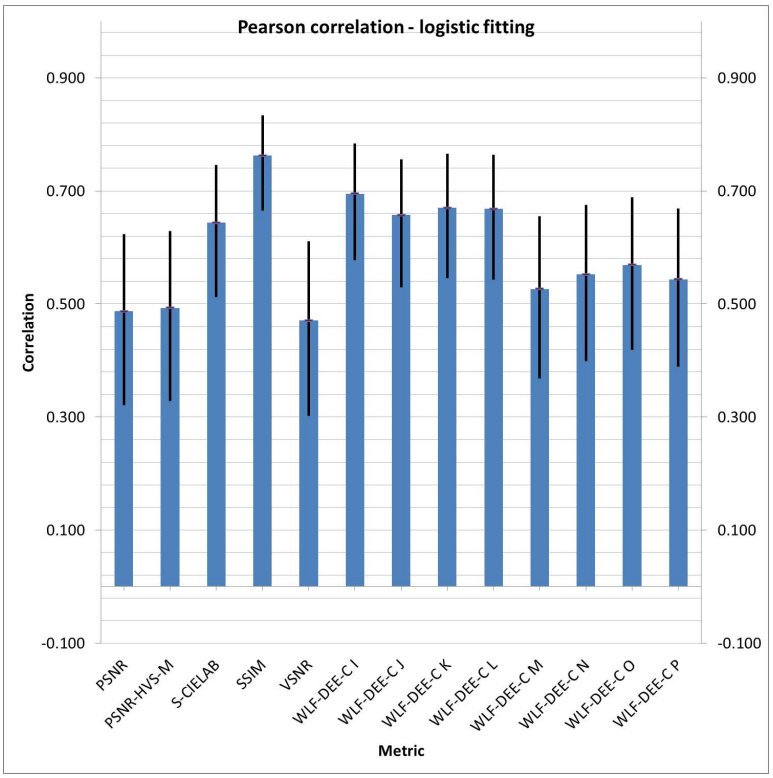


Fig. 7.8 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on TID–contrast change database.

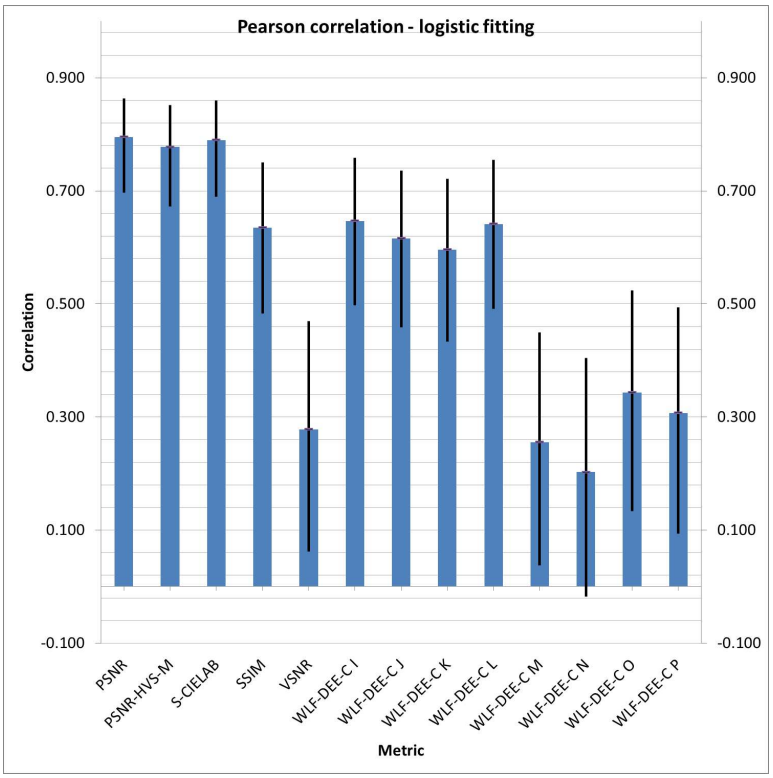


Fig. 7.9 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on Ajagamelle database.

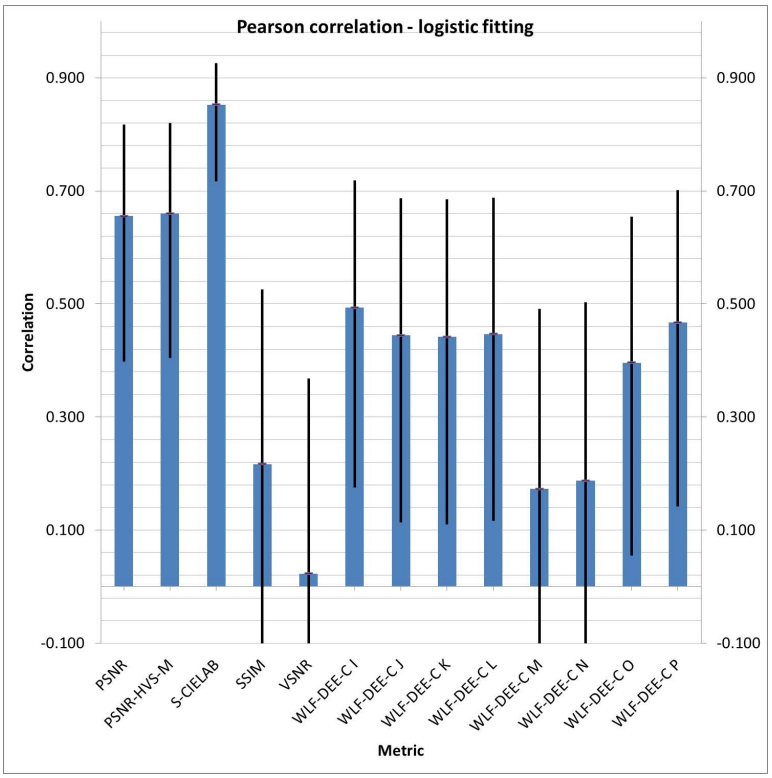


Fig. 7.10 Pearson correlation with logistic fitting and confidence intervals for the selected state–of–the–art metrics and WLF–DEE on Pedersen database.

shown to be stable in terms of changes in lightness, color and saturation, which alterate the perceived contrast.

According to the results presented in two of the related papers the DOG model is inadequate to calculate image quality in relation to perceived compression artifacts and for this reason we have decided to focus only on those categories that provide changes in contrast.

In a recent study Haun and Peli [98] state:

> "Many models intended to evaluate visual properties such as image discriminability compute perceived contrast by using contrast sensitivity functions derived from studies of human spatial vision. Such use is of questionable validity even for such applications (i.e. full–reference image quality metrics), but it is usually inappropriate for no–reference image quality measures. In this paper, we outline why the contrast sensitivity functions commonly used are not appropriate in such applications, and why weighting suprathreshold contrasts by any sensitivity function can be misleading."

Although the DOG model is not strictly comparable to a CMF or a CSF, in light of the above statement a debate on the traditional framework for image difference metrics can be opened. A framework based on pre–filtering before applying a color difference formula might be obsolete. Pre–filtering an image might remove sensitive content information while a color difference formula might be inadequate for some types of distortions and/or for reproductions under non–uniform illumination. For this reason a structural based metric (e.g., SSIM and derived ones) might be more suitable and showing higher correlation.

In conclusion, we might say that WLF–DEE should be chosen in the confined context of reproduction with only contrast changes. Contrast can be used as quality indicator, but it seems clear that is not sufficient. Quality should take into account more perceptual attributes [222, 221]. This leaves room open for future work such as extending WLF–DEE to more attributes or using it in conjunction with other metrics and/or simple statistic measures, and then judge quality as a result of multiple scores or an average of them as proposed by Zewdie et al. [313].

7.4 Q4: Can we improve the STRESS algorithm, estimating the two envelopes through the Total Variation method instead of using stochastic sampling?

As a logical consequence of **Paper A–C**, a WLF–Retinex algorithm should be presented next. However, in the current state–of–the–art, a multilevel center/surround based Retinex already exists with the name Multi–Scale Retinex (MSR) [124, 123].

This has led the authors to investigate different ways of measuring and enhancing local contrast for Retinex. A variational Retinex, named Spatio–Temporal Retinex–like Envelope with Total Variation (STRETV) is thus presented in **Paper D**. This spatial color algorithm substitutes the stochastic sampling used by the Spatio–Temporal Retinex–like Envelope with Stochastic Sampling (STRESS) [135] with the Total Variation method. The Total Variation is used as local measure of contrast and in particular for estimating two envelope functions defining the local reference white and black points of each pixel for each channel.

In this way, STRETV has different features with respect to MSR, its predecessors RSR and STRESS, and other variational Retinex algorithms. According to Section 4.2.2, the reset mechanism is a key feature of Retinex. Being a variational Retinex, STRETV does not perform a MI–Retinex spatial reset like its predecessors RSR and STRESS, which reset on the maximum of the spray. However, we may consider that STRETV performs a temporal reset as the maximum is defined by the time marching. Like its predecessors and the original Retinex, STRETV keeps a White Patch behavior.

According to Section 4.2.2 most of the variational–based Retinex (e.g., Kimmel [130]) work on the principle of separating illuminant and reflectance. In the aftermath of Horn [104] reformulation, many researchers have used Retinex as model for discounting the illuminant, assuming that appearance correlates with reflectance. Retinex, belonging to Human Vision Color Constancy and not Computer Vision Color Constancy (see Section 4.3), is indifferent to reflectance and illumination. The goal of Retinex is performing spatial comparisons treating the same both edges in illumination and in reflectance, and afterwards calculating appearance whether or not correlating with reflectance [175].

Fortunately, recent works have clarified this issue that for many years have mislead many researchers [179]. We report here an exact quotation from the results of McCann et al. [179]:

> "If a computer algorithm discounted the illumination, and succeeded in accurately calculating an object's reflectance, then that algorithm would not predict observed sensations in real life scenes with complex non uniform illumination."

Thus, STRETV does not attempt this illuminant estimation, but to calculate color appear-

ance.

As well, we underline to the reader that STRETV differs in terms of calculation from Morel et al. [195] PDE–Retinex and Kimmel et al. [130] Variational Retinex. STRETV is not directly applied to the original image content, but calculates two envelopes (maximum and minimum) containing the image signal. Afterwards, a dynamic range remapping is performed based on their estimation.

According to Section 5.3, computational time is a relevant variable for evaluating a SCA. Furthermore, in Section 5.1 different solving methods have been mentioned. Thus, using another (more efficient) method than the Euler explicit time marching scheme could be subject of future investigation. STRETV has been developed with the intent of computing the envelopes more efficiently. STRESS and STRETV were tested in the way that computational time was comparable. Results have shown that with lower computational complexity $O^{STRETV}(Nn) < O^{STRESS}(NMn)$ STRETV behaves similar to STRESS with no significantly improved performance in terms of quality. On the contrary, STRETV has shown significant better performance in terms of noise. Thus, we can affirm that STRETV is more efficient to calculate the envelopes with respect to STRESS.

Regarding the computational time, the authors are aware that more efficient solving methods are available in the literature, but **Paper D** has been presented with the main intent to be a proof of concept. Furthermore, Equation 5.5 is a highly non–linear hyperbolic equation, and while for greyscale images a lot of material is available for the process of optimization and regularization, is still subject of research for color images. Furthermore, the Euler explicit time marching scheme allowed the authors to directly apply without any reformulation the Bloomgran and Chen [23] regularization. Unlike the original Retinex and its predecessors RSR and STRESS, STRETV links the color channels resulting in a better automatic color correction.

7.5 Q5: Can we improve the original Retinex algorithm using the Ant Colony Optimization model for generating intelligent paths across the image?

The original Retinex and also the MI–Retinex family are well known algorithms for Human Vision Color Constancy with the main goal of performing local contrast enhancement and light/dark adaptation [180, pp. 324, 356]. The use of random/Brownian paths and stochastic sampling may lead to halos and chromatic noise, respectively. For this reason, after reinvestigation of the original model, a novel path–based Retinex, named Termite Retinex (TR), is

presented in **Paper E**. TR replaces the original random paths with a series of paths built upon the modification of the Ant Colony Optimization (ACO) probabilistic formula for the TSP problem. We can underline a similarity between TR probabilistic formula involving one of the 8 neighboring pixels (**Paper E–Equation 5**) and a finite difference method for calculating the gradient. Thus, **Paper E–Equation 5** becomes the local measure of contrast.

According to the results in **Paper E**, a low number of termites and short paths may lead to noise, while a low quantity of poison may lead to halos. The first impression may be that TR shares both the weaknesses of stochastic sampling and Brownian paths. On the contrary, the authors would claim that is a strong point of TR and not a drawback. The opportunity not to be vulnerable to noise and halos as counterpart of five input parameters for tuning the filter may lead a user to choose TR.

According to Section 5.3, the number of input parameters determines the ease of use of an algorithm. TR is directly inherited from Ant Colony Optimization [64] and have the same number of input parameters. Employing five parameters to tune the filtering, the authors have stressed on the rule of thumb with the aim of making TR user–friendly . Reduction of the number of input parameters and/or automatic retrieval can be subject of future work.

As TR has been derived from Ant Colony Optimization for the Traveling Salesman Problem [64], representing the content of the image on a graph could be seen as an obvious alternative choice. Montagna and Finlayson [191] follow this approach, but although the method proposed is efficient for generating pseudo–Brownian paths, some issues arise both by a general point of view and for TR. The main issue on representing the image content on a graph is how to generate the connected components (vertices and edges). Whether to use 4–neighborhood or 8–neighborhood connection is still subject of research and most of the time is target application dependent [90, pp. 591–595] [133, pp. 391–395]. As well as discussed in Montagna and Finlayson [191], the computational complexity is reduced as counterpart of a heavy memory consumption.

Graph representation of the image content could be the subject of future research for TR, but heavy memory consumption might be a considerable constraint [148, 271].

Chapter 8

Conclusion

The purpose of this research work was to contribute in the fields of image contrast, image quality and image enhancement. Five research questions were formulated and recalled here:

1. Can we improve the Rizzi et al. measure of contrast using the Difference of Gaussians model instead of using a weighted 8–neighborhood?

2. Are there significant differences between assessing contrast in uncontrolled environments and a lab controlled environment?

3. Can we estimate the quality of a reproduction image with respect to an original image in terms of contrast using the Difference of Gaussians model?

4. Can we improve the STRESS algorithm, estimating the two envelopes through the Total Variation method instead of using stochastic sampling?

5. Can we improve the original Retinex algorithm using the Ant Colony Optimization model for generating intelligent paths across the image?

Each of the five research questions were subject to two main activities: development and validation. In the first activity the problem was analyzed and then new methods were proposed, or methods from existing literature were proposed in a novel way. In the second activity psychophysical experiments and statistical tools were exploited to investigate whether the developed methods fulfilled the specific intended use.

To answer to the first research question, a measure of contrast, named WLF (Weighted–Level Framework), has been developed based on three features: multilevel analysis, Difference of Gaussians model and variance weighting. A pyramid is created halving the image at each iteration. For each level of each channel, the DOG model is computed obtaining a local

contrast map. Gaze information, saliency maps and GAFFE maps can be applied to emphasize the contrast information of a pixel and/or area. Variance is used as weight for each level of each channel and for all the three color channels. The performance of WLF were evaluated in terms of correlation on ad–hoc database of 15 color images.

Results show WLF having the same performance as Peli band–limited and Rizzi et al. (RAMM) contrast measures, and to be significantly better than Tadmor and Tolhurst (TT) measure. Regions–of–Interest, such as gaze information, saliency maps and GAFFE maps have shown, by discarding relevant information, to not improve WLF to predict perceived contrast in digital images.

To answer to the second research question, a website with the same 15 color images and the respective greyscale ones has been developed for recording perceived contrast in uncontrolled environments.

An average of 170 ratings was recorded for each image. Observers indicated that color images are rated higher than their respective greyscale ones indicating that contrast is influenced by color. Measuring contrast in uncontrolled environments can be significantly different from measuring in lab controlled environment and leading to a decrease in correlation. A comparison of the correlation coefficients at 99% confidence level confirms the above statement.

To answer to the third research question, an image difference metric, named WLF–DEE, has been developed following the same framework of WLF. For each level a pixelwise color difference between the original image and its reproduction has been computed using the Euclidean color–difference formula for small–medium color differences in log–compressed OSA–UCS space (DEE). Then, a weighted recombination of the quality map obtained for each level is performed giving the final quality score. The performance of WLF–DEE were evaluated in terms of correlation on masked noise, quantization noise, image denoising and contrast change categories of the TID database and on two databases providing different distortions directly related to color and contrast.

Results show WLF–DEE having the same performance as SSIM and S–CIELAB, and to be significantly better than PSNR and VSNR on the four categories of the TID database. WLF–DEE shows also having the same performance as other state–of–the–art metrics on both databases except for S–CIELAB on Pedersen database, while the latter has slightly higher performance. Root–Mean–Square Error confirms WLF–DEE to be in line with other state–of–the–art metrics in estimating the perceived magnitude of contrast between an original and a reproduction.

To answer to the fourth research question, a Retinex algorithm, named Spatio–Temporal Retinex–like Envelope with Total Variation (STRETV), has been developed replacing the stochastic sample of its predecessor STRESS with the Total Variation method. The corre-

sponding Euler–Lagrange equation is solved using an explicit time marching scheme and performing the Blomgren and Chan regularization at each iteration. STRETV is not directly applied to the image content but estimates two envelope functions, the maximum and minimum envelopes, containing the image signal. The two envelopes are then used to perform a dynamic range rescaling.

Two experiments show that the overall performance of STRETV is significantly better at 95% confidence level from the original image but not significantly different from its predecessor STRESS. Higher preference of STRETV with respect to STRESS is reported by observers due to a lower perception of noise. STRETV can be employed for applications such as unsupervised image enhancement, dynamic range stretching and color correction.

Finally, to answer to the fifth research question, a Retinex algorithm, named Termite Retinex (TR), has been developed replacing the original random paths of the original Retinex model with a specialized swarm of termites, derived from the Ant Colony Optimization model. TR performs an exploration of the image content tuned by two parameters, α and β, which weight the relative importance of the so–called poison and of the so–called closeness.

Two experiments show an overall higher performance of TR with respect to the original and Random Spray Retinex in terms of observers' preference and perceived noise. A sign–test at 95% confidence interval confirms the above statement. TR can be employed for applications such as unsupervised image enhancement, dynamic range stretching and color correction. Furthermore, TR can be used as unsupervised tone renderer.

Chapter 9

Future Work

In this research work, we have presented four image processing tools, a multilevel contrast measure named WLF, a multilevel contrast based image difference metric named WLF–DEE and two distinct Retinex algorithms named STRETV and TR respectively. There are several future directions of research from further implementation, optimization and testing to psychophysical experiments and analysis.

All our tools have been tested only on LDR images displayed on a digital monitor. Datasets containing other type of images (e.g., infrared, ultrasound and underwater images) and/or displayed on other media (e.g., printed images) could be subject of future evaluation.

HDR images are becoming more and more popular, and new displays able to handle more than the traditional low dynamic range are becoming available in the customer market underlining the importance of having an efficient and easy to use image processing tools for different types of images. Future work could consist in evaluating WLF in terms of correlation against other state–of–the–art contrast measures after performing a psychophysical experiment with a set of users judging contrast of HDR images displayed on native monitor. As well, a psychophysical experiment with a database of unsupervised tone rendered images, and a comparison with other state–of–the–art metrics could be subject of future work for WLF–DEE. Both STRETV and TR could be instead compared against other unsupervised tone rendering algorithms.

Recently, a solution to the inconsistencies of CIE color appearance models have been proposed. Analysis in terms of correlation with observers for both WLF and WLF–DEE could be taken into account in order to investigate whether a color appearance model is more suitable than the traditional RGB color space. The ΔE_E formula has been extended to different illuminants, and thus a dataset for color constancy studies could be considered as well.

During psychophysical experiments, observers may judge the overall quality as result of multiple image attributes scores. Thus, using WLF–DEE in conjunction with other metrics

and/or simple statistic measures, or extending WLF–DEE in calculation of more image attributes could be subject of future development and evaluation.

For a completeness of an image framework, both STRETV and TR would require exploration in applicability to other target applications (e.g., gamut mapping and dequantization). Although the authors have stressed on the rule of thumb, the automatic retrieval of the parameters to simplify the use of these frameworks should be investigated. Furthermore, experiments with users to evaluate both models in terms of perceptual attributes with respect to other image frameworks (e.g., ACE and iCAM) should be subject of analysis. For STRETV other numerical methods (e.g., FEM and Gauss–Seidel) can be subject of future investigation to speed–up the minimization of the Total Variation method.

Finally, all the tools have been designed and developed for off–line use, or in other words for processing an entire database stored on a hard drive. For real–time computation, future studies may follow two directions: reducing the computational complexity and/or re–engineering the algorithm for a dedicated hardware.

References

[1] E. H. Adelson. Perceptual organization and the judgement of brightness. *Sci*, 262:2042–2044, 1993. 27

[2] E. H. Adelson, C. H. Anderson, J. R. Bergen, P. J. Burt, and J. M. Ogden. Pyramid methods in image processing. *RCA Engineer*, 29(6), Nov 1984. 28, 30, 131

[3] A. J. Ahumada and B. L. Beard. A simple vision model for inhomogeneous image quality assessment. *SID Symposium Digest of Technical Papers*, 29(1):1109–1111, May 1998. 34

[4] S. Ajagamelle. Analysis of the difference of gaussians model in perceptual image difference metrics. Master's thesis, Gjøvik University College and Grenoble Institute of Technology, 2009. 70

[5] S. A. Ajagamelle, M. Pedersen, and G. Simone. Analysis of the difference of gaussians model in image difference metrics. In *CGIV 2010 5th European Conference on Colour in Graphics, Imaging, and Vision*, IS&T Proceedings, pages 489–496, Joensuu, Finland, Jun 2010.

[6] S. Akramullah. *Digital Video Concepts, Methods, and Metrics: Quality, Compression, Performance, and Power Trade–off Analysis*. Apress, 2014. 7, 8

[7] J. Albers. *Interaction of Color*. Yale University Press, 1975. 105

[8] E. Allen and S. Triantaphillidou. *The Manual of Photography*. Focal, 10th edition, 2011. 2, 3, 4, 8, 15, 22, 23, 26, 40, 77, 78, 79, 135

[9] ANSI/IEEE. Std p829–2007: IEEE standard for software and system test documentation, 2007. 46

[10] G. Archanjo, F. Andrijauskas, and D. Muñoz. Marvin–a tool for image processing algorithm development. pages 5–6, 2008. 83

[11] J. S. Babcock. Eye tracking observers during color image evaluation tasks. Master's thesis, Rochester Institute of Techonology, 2000. 67, 137

[12] J. S. Babcock, M. Lipps, and J. B. Pelz. How people look at pictures before, during and after scene capture: Buswell revisited. In B.E.Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging VII*, volume 4662 of *SPIE Proceedings*, pages 34–47, San Jose, CA, USA, Jan 2002. 67, 137

[13] K. Barnard, V. Cardei, and B. Funt. A comparison of computational color constancy algorithms –part I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–984, Sep 2002. 105

[14] K. Barnard and B. Funt. Analysis and improvement of multi-scale Retinex. In *CIC5 5th Color Imaging Conference: Color Science, Systems and Applications*, volume 5 of *IS&T Proceedings*, pages 221–226, Scottsdale, AZ, USA, Nov 1997. 100

[15] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms–partII: Experiments with image data. *IEEE Transactions on Image Processing*, 11(9):985–996, Sep 2002. 105

[16] P. G. J. Barten. *Contrast sensitivity of the human eye and its effect on image quality*. HV Press Knegsel, 1999. 135

[17] S. S. Bedi and R. Khandelwal. Various image enhancement techniques–a critical review. *International Journal of Advanced Research in Computer and Communication Engineering*, 2(3), Mar 2013. 84

[18] A. Beghdadi, M.-C. Larabi, A. Bouzerdoum, and K. Iftekharuddin. A survey of perceptual image processing methods. *Signal Processing: Image Communication*, 28(8):811–831, Sep 2013. 2, 134

[19] G. Beni and J. Wang. Swarm intelligence in cellular robotic systems. In *NATO Advanced Workshop on Robotics and Biological Systems*, Jun 1989. 117

[20] R. S. Berns. *Billmeyer and Saltzman's Principles of Color Technology*. John Wiley & Sons, 3rd edition, Apr 2000. 50

[21] R. Blackwell. Studies of psychophysical methods for measuring visual thresholds. *Journal of Optical Society of America*, 42(9):606–614, Sep 1952. 25

[22] P. Blanchard and E. Brüning. *Variational Methods in Mathematical Physics: A Unified Approach*. Springer–Verlag, 1992. 109

[23] P. Blomgren and T. F. Chan. Color TV: Total variation methods for restoration of vector–valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, Mar 1998. 110, 114, 149

[24] G. Boccignone, M. Ferraro, and T. Caelli. Encoding visual information using anisotropic transformations. *IEEE Trans. Pattern Anal. Mach. Intell*, 23(2):207–211, 2001. 36, 116

[25] E. Bonabeau, M. Dorigo, and G. Theraulaz. *Swarm Intelligence: From Natural to Artificial Systems*. Oxford University Press, 1999. 117

[26] N. Bonnier, F. Schmitt, H. Brettel, and S. Berche. Evaluation of spatial gamut mapping algorithms. In *CIC14 14th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*, volume 14 of *IS&T Proceedings*, pages 56–61, Scottsdale, AZ, USA, Nov 2006. 62

[27] A. Bovik. *Handbook of Image & Video Processing*. John Wiley & Sons, 2005. 24, 28, 78, 83, 90

[28] A. P. Bradley. Can region of interest coding improve overall perceived image quality? In *WDIC2003 Workshop on Digital Image Computing*, APRS Proceedings, pages 41–43, Brisbane, Australia, Feb 2003. 137

[29] D. H. Brainard and B. A. Wandell. Analysis of the Retinex theory of color vision. *Journal of the Optical Society of America A*, 3(10):1651–1661, Oct 1986. 96

[30] M. H. Brill. Irregularity in CIECAM02 and its avoidance. *Color Research and Application*, 31(2):142–145, Apr 2006. 136

[31] M. H. Brill and M. Mahy. Visualization of mathematical inconsistencies in CIECAM02. *Color Research and Application*, 38(3):188–195, Jun 2013. 136

[32] M. H. Brill and C. Oleari. Chromatic adaptation by illuminant matrix products: An alternative to sharpened von kries primaries. *Color Research And Application*, 39(3):275–278, 2014. 136

[33] M. H. Brill and S. Susstrunk. Repairing gamut problems in CIECAM02: A progress report. *Color Research and Application*, 33(5):424–426, Oct 2008. 136

[34] P. Burian. *Mastering Digital Photography and Imaging*. Sybex, 2006. 23

[35] D. A. Burkhardt, J. Gottesman, D. Kersten, and G. E. Legge. Symmetry and constancy in the perception of negative and positive luminance contrast. *Journal of the Optical Society of America A*, 1(3):309–316, Mar 1984. 2, 27

[36] A. J. Calabria and M. D. Fairchild. Perceived image contrast and observer preference: I. the effects of lightness, chroma, and sharpness manipulations on contrast perception. *Journal of Imaging Science and Technology*, 47:479–493, Nov/Dec 2003. 131

[37] A. J. Calabria and M. D. Fairchild. Perceived image contrast and observer preference II. empirical modeling of perceived image contrast and observer preference data. *Journal of Imaging Science and Technology*, 47:494–508, Nov/Dec 2003. 28, 37, 131

[38] F. W. Campbell and J. G. Robson. Application of fourier analysis to the visibility of gratings. *Journal of Physiology*, 197(3):551–566, Aug 1968. 30, 32

[39] M. E. Celebi, M. Lecca, and B. Smolka. *Color Image and Video Enhancement*. Springer International Publishing, 1st edition, 2015. 84

[40] A. Chambolle, V. Caselles, M. Novaga, D. Cremers, and T. Pock. An introduction to Total Variation for Image Analysis. Nov 2009. 113

[41] T. Chan and J. Shen. *Image Processing And Analysis: Variational, PDE, Wavelet, And Stochastic Methods*. Society for Industrial and Applied Mathematics, 2005. 90, 110, 116

[42] D. Chandler and S. Hemami. Dynamic contrast-based quantization for lossy wavelet image compression. *IEEE Transactions on Image Processing*, 14(4):397–410, Apr 2005. 64

[43] D. Chandler and S. Hemami. VSNR: A wavelet-based visual signal-to-noise ratio for natural images. *IEEE Transactions on Image Processing*, 16(9):2284–2298, Sep 2007. 63

[44] CIE. Industrial colour-difference evaluation. CIE Technical Report 116, Central Bureau of the CIE, Vienna, Austria, 1995. 51

[45] CIE. The cie 1997 interim colour appearance model (simple version), CIECAM97s. Technical Report CIE Publication 131, CIE Central Bureau, Vienna, Austria., 1998. 51

[46] CIE. Colorimetry. Technical Report 15, 2004. 36, 48

[47] CIE. A colour appearance model for colour management systems: CIECAM02. Technical Report CIE Publication 159, CIE Central Bureau, Vienna, Austria, 2004. 51

[48] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical report, CIE TC8–08, 2004. 121

[49] CIE. Colorimetry – part 4: CIE 1976 L*a*b* colour spaces. CIE draft standard ds 014-4.2/e:2006, 2006. 48

[50] CiteSeerX.edu. Cite seer x beta. 2014. 33

[51] F. J. J. Clarke, R. McDonald, and B. Rigg. Modification to the JPC79 color–difference formula. *Journal of the Society of Dyers and Colorists*, 100(4):128–132, Apr 1984. 50, 51

[52] T. J. Cooper and F. A. Baqai. Analysis and extensions of the Frankle–McCann Retinex algorithm. *Journal of Electronic Imaging*, 13(1):85–92, Jan 2004. 99

[53] T. N. Cornsweet. Visual perception. NewYork, NY, USA, 1970. Academic Press. 25

[54] Ctein. *Digital Restoration from Start to Finish*. Focal Press, 2nd edition, 2010. 3, 24

[55] L. Cui. SWVFS: a saliency weighted visual feature similarity metric for image quality assessment. *Frontiers of Computer Science*, 8(1):145–155, Feb 2014. 67

[56] S. Daly. The visible differences predictor: an algorithm for the assessment of image fidelity. In B. E. Rogowitz, editor, *Human Vision, Visual Processing, and Digital Display III*, volume 1666 of *SPIE Proceedings*, pages 2–15, San Jose, CA, USA, Feb 1992. 65, 135

[57] N. Daw. Why after–images are not seen in normal circumstances. *Nature*, 196:1143–45, 1962. 91

[58] H. de Ridder. Minkowski-metric as a combination rule for digital-image-coding impariments. In B. Rogowitz, editor, *Human Vision, Visual Processing, and Digital Display III*, volume 1666 of *SPIE Proceedings*, pages 16–26, San Jose, CA, USA, Aug 1992. 67

[59] Deneubourg, S. Aron, S. Goss, and J. M. Pasteels. The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior*, 3(2):159–168, Mar 1990. 117

[60] Dictionary.com. Dictionary.com. In *http://dictionary.reference.com/browse/contrast*, Last Visit: Sep 2014. 7, 23

[61] Dictionary.com. Dictionary.com. In *http://dictionary.reference.com/browse/pool*, Last Visit: Sep 2014. 38

[62] Dictionary.com. Dictionary.com. In *http://dictionary.reference.com/browse/quality*, Last Visit: Sep 2014. 43

[63] M. Dorigo. *Optimization, Learning and Natural Algorithms*. PhD thesis, Dipartimento di Elettronica e Informazione, Politecnico di Milano, Milano, Italy, 1992. 117

[64] M. Dorigo and L. Gambardella. Ant Colonies for the Traveling Salesman Problem. *BioSystems*, 43(2):73–81, Jul 1997. 117, 118, 150

[65] M. Dorigo and L. M. Gambardella. Ant Colony System: A cooperative learning approach to the Traveling Salesman Problem. *IEEE Transactions on Evolutionary Computation*, 1(1):53–66, Apr 1997. 117

[66] M. Dorigo, V. Maniezzo, and A. Colorni. Ant system: An autocatalytic optimizing process. Technical Report 91–016, Dipartimento di Elettronica e Informazione-Politecnico di Milano, Italy, 1991. 118

[67] M. Dorigo, V. Maniezzo, and A. Colorni. The Ant System: Optimization by a Colony of Cooperating Agents. *IEEE Transactions on Systems, Man, and Cybernetics-Part B*, 26(1):29–41, Feb 1996. 117, 130

[68] M. Dorigo and T. Stützle. Ant Colony Optimization. MIT Press, Cambridge, MA, USA, 2004. 117

[69] R. C. Eberhart, Y. Shi, and J. Kennedy. *Swarm Intelligence*. Morgan Kaufmann, 1st edition, 2001. 117

[70] F. Ebner and M. D. Fairchild. Development and testing of a color space (IPT) with improved hue uniformity. In *CIC6 6th Color Imaging Conference: Color Science, Systems and Applications*, volume 6 of *IS&T Proceedings*, pages 8–13, Scottsdale, AZ, USA, 1998. 60, 62

[71] M. Ebner. *Color Constancy*. John Wiley & Sons, 2007. 9, 104, 106, 120

[72] T. Eerola, L. Lensu, H. Kälviáinen, J.-K. Kamarainen, T. Leisti, G. Nyman, R. Halonen, and P. Oittinen. Full reference printed image quality: Measurement framework and statistical evaluation. *Journal of Imaging Science and Technology*, 54(1):10201–1–10201–13, Jan 2010. 55

[73] K. Egiazarian, J. Astola, N. Ponomarenko, V. Lukin, F. Battisti, and M. Carli. Two new full-reference quality metrics based on HVS. In *Proceedings of the 2nd International Workshop on Video Processing and Quality Metrics*, Scottsdale, AZ, USA, Jan 2006. 63

[74] P. G. Engeldrum. A framework for image quality models. *Journal of Imaging Science and Technology*, 39(4):259–270, 1995. 66

[75] P. G. Engeldrum. Image quality modeling: Where are we? In *PICS 1999 Image Processing, Image Quality, Image Capture, Systems Conference*, volume 2 of *IS&T Proceedings*, pages 251–255, Savannah, GE, USA, Apr 1999. 2, 43, 44, 46

[76] P. G. Engeldrum. A short image quality model taxonomy. *Journal of Imaging Science and Technology*, 48(2):160–165, Apr 2004. 45

[77] C. Enroth-Cugell and J. G. Robson. The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, 187:517552, 1966. 35

[78] M. D. Fairchild. Image quality measurement and modeling for digital photography. In *International Congress on Imaging Science*, pages 318–319, Tokyo, Japan, May 2002. 44

[79] M. D. Fairchild. *Color Appearance Models*. John Wiley & Sons, Chichester, UK, 2nd edition, 2005. 9, 22, 30, 135

[80] M. D. Fairchild and G. M. Johnson. The iCAM framework for image appearance, image differences, and image quality. *Journal of Electronic Imaging*, 13(1):126–138, Jan 2004. 60, 85

[81] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256, Jul 2002. 90

[82] M. Ferraro and G. Boccignone. Image contrast enhancement via entropy production. *Real-Time Imaging*, 10(4):229–238, 2004. 36, 116

[83] J. Frankle and J. J. McCann. Method and apparatus for lightness imaging, patent US 4384336 a, May 1983. 28, 98, 100, 122, 131

[84] B. Funt, F. Ciurea, and J. J. McCann. Retinex in MATLAB. *Journal of Electronic Imaging*, 13(1):48–57, Jan 2004. 99, 122

[85] J. Garcia-Alvarez, H. Führ, and G. Castellanos-Dominguez. Evaluation of region-of-interest coders using perceptual image quality assessments. *Journal of Visual Communication and Image Representation*, 24(8):1316–1327, Nov 2013. 137

[86] C. Gatta, A. Rizzi, and D. Marini. ACE: An automatic color equalization algorithm. In *CGIV 2002 1st European Conference on Color in Graphics Image and Vision*, IS&T Proceedings, pages 2–5, Poitiers, France, Apr 2002. 85, 90, 122

[87] J. R. Giles. *Introduction to the Analysis of Metric Spaces*. Cambridge University Press, 1st edition, 1987. 56

[88] E. B. Goldstein. *Blackwell Handbook of Sensation and Perception*. Blackwell Publishing, 2005. 9, 10, 11

[89] M. Gong and M. Pedersen. Spatial pooling for measuring color printing quality attributes. *Journal of Visual Communication and Image Representation*, 23(5):685–696, Jul 2012. 66, 67

[90] R. C. Gonzales and R. E. Woods. *Digital Image Processing*. Prentice Hall, 2nd edition, 2002. 83, 84, 150

[91] S. Goss, S. Aron, J. Deneubourg, and J. Pasteels. Self-organized shortcuts in the Argentine ant. *Naturwissenschaften*, 76(12):579–581, Dec 1989. 117

[92] H. Grassmann. Zur Theorie der Farbenmischung. *Annalen der Physik, 89*, 89:69–84, 1853. 47

[93] D. G. Green. The contrast sensitivity of the colour mechanisms of the human eye. *The Journal of Physiology*, 196(2):415–429, 5 1968. 30

[94] P. Green. *Color management: understanding and using ICC profiles*. John Wiley & Sons, 1st edition, 2010. 51, 83

[95] J. Guild. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London*, 230:149–187, 1931. 47

[96] R. Gury and M. Shaw. Dealing with imaginary color encodings in CIECAM02 in an ICC workflow. In *CIC 13 Thirteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 13 of *IS&T Proceedings*, pages 217–223, Scottsdale, AZ, USA, Nov 2005. 51

[97] J. Y. Hardeberg. *Acquisition and reproduction of colour images: colorimetric and multispectral approaches*. Phd thesis, Ecole Nationale Superieure des Telecommunications, 1999. 51

[98] A. M. Haun and E. Peli. Complexities of complex contrast. In R. Eschbach, G. G. Marcu, and A. Rizzi, editors, *Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications*, volume 8292 of *SPIE Proceedings*, pages 82920E–82920E–7, Burlingame, CA, USA, Jan 2012. 134, 147

[99] A. M. Haun and E. Peli. Perceived contrast in complex images. *Journal of Vision*, 13(13):1–21, Nov 2013. 14, 40

[100] E. . Hering. Outline of a theory of light sense. L. M. Hurvich and D. Jameson, Trans Cambridge, Harvard University Press 1964., 1905. 102

[101] R. F. Hess, A. Bradley, and L. Piotrowski. Contrast–coding in amblyopia I. differences in the neural basis of human amblyopia. *Proceedings of the Royal Society of London B: Biological Sciences*, 217(1208):309–330, Feb 1983. 2, 27, 28, 131

[102] J. Holm, I. Tastl, and T. Johnson. Definition & use of the ISO 12640-3 reference color gamut. In *Fourteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, pages 62–68, Scottsdale, AZ, Dec 2006. IS&T/SID. 121

[103] G. Hong and M. Luo. Perceptually based colour difference for complex images. In R. Chung and A. Rodrigues, editors, *AIC2001 9th Congress of the International Colour Association*, volume 4421 of *SPIE Proceedings*, pages 618–621, Rochester, NY, USA, Jun 2001. 60

[104] B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(4):277–299, Dec 1974. 101, 148

[105] R. Huertas, M. Melgosa, and C. Oleari. Performance of a color–difference formula based on OSA–UCS space using small-medium color differences. *Journal of the Optical Society of America A*, 23(9):2077–2084, Sep 2006. 54

[106] R. Hunt. *The Reproduction of Colour*. John Wiley & Sons, 6th edition, 2004. 50, 88

[107] A. Hurlbert. Formal connections between lightness algorithms. *Journal of the Optical Society of America A*, 3(10):1684–1693, Oct 1986. 101

[108] I3A (International Imaging Industry Association). CPIQ initiative phase 1 white paper: Fundamentals and review of considered test methods, Oct 2007. 44

[109] International Color Consortium. Icc color management. Last visit: Sep 2016. 83

[110] International Color Consortium. Information on profiles. Last visit: Sep 2016. 51

[111] R. Iordache, A. Beghdadi, and P. V. de Lesegno. Pyramidal perceptual filtering using moon and spencer contrast. In *International Conference on Image Processing*, volume 3, pages 146–149, Thessaloniki, Greece, Oct 2001. IEEE. 34, 37

[112] ISO. ISO 20462–1 photography–psychophysical experimental methods to estimate image quality–part 1: Overview of psychophysical elements, Nov 2005. 44

[113] ISO. ISO 9001:2008–quality management, 2008. 46

[114] ISO. ISO 15076–1, ICC.1:2010, image technology colour management–architecture, profile format, and data structure (profile version 4.3.0.0)., 2010. 51, 83

[115] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998. 67, 137

[116] ITU. Methods for subjective determination of transmission quality, Aug 1996. 68

[117] ITU. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, Feb 2001. 68

[118] ITU. Methodological framework for specifying accuracy and cross-calibration of video quality metrics, Feb 2004. 68

[119] ITU. Methodology for the subjective assessment of the quality of television pictures, Jan 2012. 47, 68

[120] T. Janssen. *Computational Image Quality*. PhD thesis, IPO, Center for User-System Interaction, 1999. 44

[121] J. Jiang, Z. Wang, M. R. Luo, M. Melgosa, M. H. Brill, and C. Li. Optimum solution of the CIECAM02 yellow–blue and purple problems. *Color Research and Application*, 39(5), Oct 2014. 136

[122] D. J. Jobson, Z. U. Rahman, and G. A. Woodell. Retinex image processing: Improved fidelity to direct visual observation. In *CIC4 4th Color Imaging Conference: Color Science, Systems, and Applications*, volume 4 of *IS&T Proceedings*, pages 124–125, Scottsdale, AZ, USA, Nov 1996. 99

[123] D. J. Jobson, Z. U. 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, Jul 1997. 99, 122, 148

[124] D. J. Jobson, Z. U. Rahman, and G. A. Woodell. Properties and performance of a center/surround Retinex. *IEEE Transactions on Image Processing*, 6(3):451–462, Mar 1997. 99, 148

[125] G. M. Johnson and M. D. Fairchild. Darwinism of color image difference models. In *CIC9 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, and Applications*, volume 9 of *IS&T Proceedings*, pages 108–112, Scottsdale, AZ, USA, Nov 2001. 30, 58, 60

[126] M. Kamermans and H. Spekreijse. The feedback pathway from horizontal cells to cones: A mini review with a look ahead. *Vision Research*, 39(15):2449–2468, Jul 1999. 13

[127] E. Kandel, J. Schwartz, and T. Jessell. Principles of neural science. USA, Jan 2000. McGraw–Hill. 35

[128] B. W. Keelan. Handbook of image quality – characterization and prediction. New York, NY, 2002. Marcel Dekker, Inc. 44

[129] J. Kim and H. Shin. Algorithm & soc design for automotive vision systems: For smart safe driving system. Springer Publishing Company, Incorporated, 2014. 120

[130] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel. A variational framework for Retinex. *International Journal on Computer Vision*, 52(1):7–23, Apr 2003. 102, 116, 148, 149

[131] P. E. King-Smith and J. J. Kulikowski. Pattern and flicker detection analysed by subthreshold summation. *Journal of Physiology*, 249(3):519–548, Aug 1975. 2, 27

[132] H. Kivinen, M. Nuutinen, and P. Oittinen. Comparison of colour difference methods for natural images. In *CGIV 2010 5th European Conference on Colour in Graphics, Imaging*, IS&T Proceedings, pages 510–515, Joensuu, Finland, Jun 2010. 2

[133] J. Kleinberg and E. Tardos. Algorithm design. Boston, MA, USA, 2005. Addison–Wesley Longman Publishing. 117, 119, 150

[134] Kodak. Kodak lossless true color image suite. In *http://r0k.us/graphics/kodak*, 2004. 124

[135] Ø. Kolås, I. Farup, and A. Rizzi. Spatio-temporal Retinex-inspired envelope with stochastic sampling: A framework for spatial color algorithms. *Journal of Imaging Science and Technology*, 55(4):1–10, Aug 2011. 98, 122, 129, 148

[136] J. Kuang, G. M. Johnson, and M. D. Fairchild. iCAM06: A refined image appearance model for HDR image rendering. *Journal of Visual Communication and Image Representation*, 18(5):406–414, Oct 2007. 85, 90, 122

[137] R. G. Kuehni. Threshold color differences compared to supra-threshold color differences. *Color Research & Application*, 25(3):226–229, 2000. 50

[138] P. Kumar and L. W. MacDonald. Measuring gloss by digital photography. In R. Eschbach and G. G. Marcu, editors, *Color Imaging XI: Displaying, Processing, Hardcopy, and Applications*, volume 6058 of *SPIE Proceedings*, pages 60580A–60580A–11, San Jose, CA, USA, Jan 2012. 106

[139] E. H. Land. The Retinex. *American Scientist*, 52:247–64, 1964. 89, 93, 130

[140] E. H. Land. The Retinex theory of color vision. *Scientific American*, 237:108–128, 1977. 90, 92, 93

[141] E. H. Land. Recent advances in Retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academy of Sciences of the United States of America*, 80(16):5163–5169, 1983. 94

[142] E. H. Land. An Alternative Technique for the Computation of the Designator in the Retinex Theory of Color Vision. *Proceedings of the National Academy of Sciences of the United States of America*, 83(10):3078–3080, May 1986. 94, 98

[143] E. H. Land. Recent advances in Retinex theory. *Vision Research*, 26(1):7–21, 1986. 94

[144] E. H. Land and J. J. McCann. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, Jan 1971. 85, 89, 93, 94, 130

[145] E. C. Larson and D. M. Chandler. Unveiling relationships between regions of interest and image fidelity metrics. In W. A. Pearlman, J. W. Woods, and L. Lu, editors, *Visual Communications and Image Processing*, volume 6822 of *SPIE Proceedings*, Jan 2008. 67

[146] P. Le Callet, S. Mö ller, and A. Perkis. Qualinet white paper on definitions of quality of experience (2012). In *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, number version 1.2, Lausanne, Switzerland, Mar 2013. 46, 119

[147] P. A. Le Callet. Subjective quality assessment IRCCyN/IVC database 2005. *IRCCyN*, 2005. 68

[148] M. Lecca, A. Rizzi, and G. Gianini. Energy–driven path search for Termite Retinex. *Journal of Optical Society of America A*, 33:1, 2016. 150

[149] H.-C. Lee. *Introduction to Color Imaging Science*. Cambridge University Press, 2005. 8, 27, 39, 79, 136

[150] T. C. Library. C++ template image processing toolkit. In *http://cimg.eu/*, 1999-2015. 83

[151] O. Lillesæter. Complex contrast, a definition for structured targets and backgrounds. *Journal of the Optical Society of America A*, 10(12):2453–2457, Dec 1993. 2, 27

[152] C. P. Lim, L. C. Jain, and S. Dehuri. *Innovations in Swarm Intelligence*. Springer–Verlag, 2009. 117

[153] N. Limare, A. B. Petro, C. Sbert, and J.-M. Morel. Retinex Poisson equation: a model for color perception. *Image Processing On Line*, 1, 2011. 101

[154] J. Lubin. *Digital Images and Human Vision*, chapter The use of psychophysical data and models in the analysis of display system performance, pages 163–178. MIT Press, Cambridge, MA, USA, 1993. 33

[155] J. Lubin. *A Visual Discrimination Model For Imaging System Design And Evaluation*, chapter 10, pages 245–283. Visual Models for Target Detection and Recognition. World Scientific Publishers, River Edge, NJ, USA, 1995. 2, 34

[156] J. Lubin. A human vision system model for objective picture quality measurements. In *International Broadcasting Convention*, pages 498–503, 1997. 40

[157] A. Lukin. Improved visible differences predictor using a complex cortex transform. Proceedings of GraphiCon, pages 145–150, Russia, Oct 2009. 65

[158] C. Lundstrom. Measuring digital image quality. Technical report, Center for medical image science and visualization, Linkoping University. 46

[159] H. R. W. G. Luo, M. R. The structure of the cie 1997 colour appearance model (CIECAM97s). *Color Research And Application*, 23:138–146, 1998. 51

[160] M. Luo, G. Cui, and B. Rigg. The development of the CIE 2000 colour-difference formula: CIEDE2000. *Color Research and Application*, 26(5):340–350, Oct 2001. 51

[161] M. Luo and B. Rigg. BFD(l:c) colour-difference formula: Part I - development of the formula. *Journal of the Society of Dyers and Colourists*, 103(2):86–94, Feb 1987. 50

[162] M. R. Luo, G. Cui, and C. Li. Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research & Application*, 31(4):320–330, 2006. 52

[163] M. R. Luo and B. Rigg. Chromaticity–discrimination ellipses for surface colours. *Color Research and Application*, 11(1):25–42, 1986. 50

[164] M. R. Luo and B. Rigg. BFD(l:c) color-difference formula: Part II – performance of the formula. *Journal of the Society of Dyers and Colorists*, 103(3):126–132, Mar 1987. 50

[165] D. MacAdam. Redeterminations of colors for uniform scales. *Journal of the Optical Society of America A*, 7(1):113–115, 1990. 52

[166] D. L. MacAdam. Visual sensitivities to color differences in daylight. *Journal of the Optical Society of America A*, 32(5):247–274, May 1942. 48

[167] D. L. MacAdam. Uniform color scales. *Journal of the Optical Society of America A*, 64(12):1691–1702, Dec 1974. 52

[168] L. W. MacDonald. Framework for an image sharpness management system. In *CIC7 7th Color Imaging Conference: Color Science, Systems and Applications*, volume 7 of *IS&T Proceedings*, pages 75–79, Scottsdale, AZ, USA, Jan 1999. 83

[169] R. Maini and H. Aggarwal. A comprehensive review of image enhancement techniques. *Journal of Computing*, 2(3):8–13, Mar 2010. 84

[170] J. Mannos and D. Sakrison. The effects of a visual fidelity criterion on encoding of images. *IEEE Transactions on Information Theory*, 20(4):525–536, Jul 1974. 60

[171] R. Mantiuk, S. Daly, K. Myszkowski, and H. Seidel. Predicting visible differences in high dynamic range images–model and its calibration. In S. J. D. Bernice E. Rogowitz, Thrasyvoulos N. Pappas, editor, *Human Vision and Electronic Imaging X*, volume 5666 of *SPIE Proceedings*, pages 204–214, San Jose, CA, USA, Jan 2005. 65

[172] D. Marini and A. Rizzi. Colour constancy and optical illusions: a computer simulation with Retinex theory. In *ICIAP'93 7th International Conference on Image Analysis and Processing*, pages 657–660, Monopoli, Italy, Sept 1993. 96

[173] D. Marini and A. Rizzi. A computational approach to color adaptation effects. *Image and Vision Computing*, 18(13):1005–1014, Oct 2000. 96

[174] G. Marsaglia and W. W. Tsang. The ziggurat method for generating random variables. *Journal of Statistical Software*, 5(8):1–7, Oct 2000. 97

[175] J. J. McCann. Lessons learned from mondrians applied to real images and color gamuts. *IS&T Reporter*, 14(6):1–7, Nov/Dec 1999. 99, 122, 148

[176] J. J. McCann. *Simultaneous Contrast and Color Constancy: Signatures of Human Image Processing*, chapter 6, pages 87–101. Color Perception: Philosophical, Psychological, Artistic, and Computational Perspectives. Oxford University Press, Vancouver, Canada, 2000. 27

[177] J. J. McCann. Capturing a black cat in shade: past and present of Retinex color appearance models. *Journal of Electronic Imaging*, 13(1):36–47, Jan 2004. 94

[178] J. J. McCann, S. P. McKee, and T. H. Taylor. Quantitative studies in Retinex theory a comparison between theoretical predictions and observer responses to the colormondrian experiments. *Vision Research*, 16(5):445–458, 1976. 27, 105

[179] J. J. McCann, C. Parraman, and A. Rizzi. Reflectance, illumination, and appearance in color constancy. *Frontiers in Psychology*, 5(5), Jan 2014. 105, 148

[180] J. J. McCann and A. Rizzi. *The Art and Science of HDR Imaging*. John Wiley & Sons, 2012. 4, 8, 15, 27, 85, 86, 89, 90, 91, 92, 93, 105, 120, 121, 149

[181] Merriam-Webster.com. The merriam–webster unabridged dictionary. In *http://www.merriam-webster.com/dictionary/contrast*, Last Visit: Sep 2014. 7, 23

[182] Merriam-Webster.com. The merriam–webster unabridged dictionary. In *http://www.merriam-webster.com/dictionary/pool*, Last Visit: Sep 2014. 38

[183] Merriam-Webster.com. The merriam–webster unabridged dictionary. In *http://www.merriam-webster.com/dictionary/quality*, Last Visit: Sep 2014. 43

[184] Merriam-Webster.com. The merriam–webster unabridged dictionary. In *http://www.merriam-webster.com/dictionary/qualia*, Last Visit: Sep 2014. 43

[185] Merriam-Webster.com. The merriam–webster unabridged dictionary. In *http://www.merriam-webster.com/dictionary/measure*, Last Visit: Sep 2014. 56

[186] L. Meylan and S. Süsstrunk. High dynamic range image rendering using a Retinex–based adaptive filter. *IEEE Transactions on Image Processing*, 15(9):2820–2830, Dec 2006. 100

[187] A. Michelson. Studies in optics. University of Chicago Press, 1927. 2, 27

[188] G. S. Miller and C. R. Hoffman. Illumination and reflection maps: Simulated objects in simulated and real environments. In *SIGGRAPH '84 Advanced Computer Graphics Animation seminar notes*, Jul 1984. 85

[189] A. Mittal, A. K. Moorthy, and A. C. Bovik. No–reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21(12):4695–4708, Dec 2012. 55

[190] R. Montagna. *Path–based computations in colour image processing*. PhD thesis, University of East Anglia, Mar 2011. 96

[191] R. Montagna and G. D. Finlayson. Constrained pseudo-brownian motion and its application to image enhancement. *Journal of the Optical Society of America A*, 28(8):1677–1688, Aug 2011. 96, 122, 150

[192] P. Moon and D. E. Spencer. The visual effect of non–uniform surrounds. *Journal of the Optical Society of America*, 35(3):233–247, 1945. 34

[193] A. K. Moorthy and A. C. Bovik. Perceptually significant spatial pooling techniques for image quality assessment. In B.E.Rogowitz and T. N. Pappas, editors, *Human Vision and Electronic Imaging XIV*, volume 7240 of *SPIE Proceedings*, pages 724012–724012–11, San Jose, CA, USA, Jan 2009. 66, 67

[194] A. K. Moorthy and A. C. Bovik. Visual quality assessment algorithms: What does the future hold? *Multimedia Tools and Applications*, 51(2):675–696, Jan. 2011. 55

[195] J. M. Morel, A. B. Petro, and C. Sbert. A PDE formalization of Retinex theory. *IEEE Transactions on Image Processing*, 19:2825–2837, Nov 2010. 101, 122, 149

[196] N. Moroney. Local color correction using non–linear masking. In *CIC8 8th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 8 of *IS&T Proceedings*, pages 108–11, Scottsdale, AZ, USA, Nov 2000. 60

[197] N. Moroney, M. D. Fairchild, R. W. Hunt, C. Li, M. R. Luo, and T. Newman. The CIECAM02 color appearance model. In *CIC10 Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 10 of *IS&T Proceedings*, pages 23–27, Scottsdale, AZ, USA, Nov 2002. 51

[198] N. Moroney and I. Tastl. Comparison of Retinex and iCAM for scene rendering. *Journal of Electronic Imaging*, 13(1):139–145, Jan 2004. 120

[199] B. Moulden, F. Kingdom, and L. Gatley. The standard deviation of luminance as a metric of contrast random–dot images. *Perception*, 19(1):79–101, Feb 1990. 27

[200] J. A. Movshon and L. Kiorpes. Analysis of the development of spatial contrast sensitivity in monkey and human infants. *Journal of the Optical Society of America A*, 5(12):2166–2172, Dec 1988. 30

[201] K. T. Mullen. The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. volume 359, pages 381–400, 1985. 30

[202] D. Navon. Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3):353–383, Jul 1977. 64

[203] J. A. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 7(4):308–313, Jan 1965. 72

[204] I. Newton. Opticks. London, UK, 1730. Reprinted by Dover, New York, 1979. 47

[205] N. Nill. A visual model weighted cosine transform for image compression and quality assessment. *IEEE Transactions on Communications*, 33(6):551–557, Jun 1985. 63

[206] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barbba. Does where you gaze on an image affect your perception of quality? applying visual attention to image quality metric. In *ICIP 2007 14th International Conference on Image Processing*, volume 6 of *IEEE Proceedings*, pages 169–172, San Antonio, TX, USA, Sep 2007. 67

[207] P. Nussbaum. *Colour Measurement and Print Quality Assessment in a Colour Managed Printing Workflow*. PhD thesis, University of Oslo and Gjøvik University College, 2011. 83

[208] N. Ohta and A. Robertson. *Colorimetry: Fundamentals And Applications*. John Wiley & Sons, 2005. 16, 48

[209] C. Oleari. Color opponencies in the system of the uniform color scales of the optical society of america. *Journal of the Optical Society of America A*, 21:677–682, 2004. 53

[210] C. Oleari. Hypotheses for chromatic opponency functions and their performance on classical psychophysical data. *Color Research and Application*, (30):31–41, 2005. 53

[211] C. Oleari. Corresponding color datasets and a chromatic adaptation model based on the osa-ucs system. *Journal of the Optical Society of America A*, 31(7):1502–1514, Jul 2014. 136

[212] C. Oleari. *Standard Colorimetry: Definitions, Algorithms and Software*. John Wiley & Sons, 1st edition, 2015. 9, 10, 11, 16, 17, 18, 20, 26, 106

[213] C. Oleari, M. Melgosa, and R. Huertas. Euclidean color–difference formula for small–medium color differences in log-compressed OSA–UCS space. *Journal of the Optical Society of America A*, 26(1):121–134, Jan 2009. 54, 128, 140

[214] OpticsInfoBase.com. Osa optics infobase. 2014. 33

[215] S. Pan, X. An, and H. He. Adapting iterative Retinex computation for high-dynamic-range tone mapping. *Journal of Electronic Imaging*, 22(2):023006–023006, Nov 2013. 99

[216] D. R. Pant and I. Farup. Riemannian formulation and comparison of color difference formulas. *Color Research and Application*, 37(6):429–440, Sep 2012. 140

[217] C. Parraman, J. J. McCann, and A. Rizzi. Artist's colour rendering of HDR scenes in 3D mondrian colour-constancy experiments. In R. Eschbach, G. G. Marcu, S. Tominaga, and A. Rizzi, editors, *Color Imaging XV: Displaying, Processing, Hardcopy, and Applications*, volume 7528 of *SPIE Proceedings*, page 752802, Jan 2010. 105

[218] C. Parraman, A. Rizzi, and J. J. McCann. Colour appearance and colour rendering of HDR scenes: An experiment. In R. Eschbach, G. G. Marcu, S. Tominaga, and A. Rizzi, editors, *Color Imaging XIV: Displaying, Processing, Hardcopy, and Applications*, volume 7241 of *SPIE Proceedings*, page 72410R, Jan 2009. 105

[219] M. Pavel, G. Sperling, T. Riedl, and A. Vanderbeek. Limits of visual communication - the effect of signal-to-noise ratio on the intelligibility of american-sign-language. *Journal of the Optical Society of America A*, 4(12):2355–2365, Dec 1987. 27

[220] M. Pedersen. Importance of region-of-interest on image difference metrics. Master thesis, Gjøvik University College, 2007. 137

[221] M. Pedersen. *Image quality metrics for the evaluation of printing workflows*. PhD thesis, University of Oslo and Gjøvik University College, Oct 2011. 43, 57, 58, 147

[222] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen. Attributes of image quality for color prints. *Journal of Electronic Imaging*, 19(1):011016–1–011016–13, Jan 2010. 1, 15, 37, 120, 147

[223] M. Pedersen and I. Farup. Simulation of image detail visibility using contrast sensitivity functions and wavelets. In *CIC20 20th Color and Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 20 of *IS&T Proceedings*, pages 70–75, Los Angeles, CA, USA, Nov 2012. 135

[224] M. Pedersen and J. Y. Hardeberg. Rank order and image difference metrics. In *CGIV 2008 4th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 120–125, Terrassa, Spain, Jun 2008. 69

[225] M. Pedersen and J. Y. Hardeberg. Survey of full–reference image quality metrics. Høgskolen i Gjøviks rapportserie 5, Gjøvik University College, The Norwegian Color Research Laboratory, Gjøvik, Norway, Jun 2009. 43, 60

[226] M. Pedersen and J. Y. Hardeberg. Full–reference image quality metrics: Classification and evaluation. *Foundations and Trends in Computer Graphics and Vision*, 7(1):1–80, Mar 2012. 3, 43, 55, 56, 57, 66

[227] M. Pedersen, J. Y. Hardeberg, and P. Nussbaum. Using gaze information to improve image difference metrics. In B. Rogowitz and T. Pappas, editors, *Human Vision and Electronic Imaging VIII*, volume 6806 of *SPIE Proceedings*, page 11, San Jose, CA, USA, Jan 2008. 126

[228] M. Pedersen, A. Rizzi, J. Y. Hardeberg, and G. Simone. Evaluation of contrast measures in relation to observers perceived contrast. In *CGIV 2008 4th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 253–256, Terrassa, Spain, Jun 2008. 37, 69, 131

[229] M. Pedersen, G. Simone, M. Gong, and I. Farup. A total variation based color image quality metric with perceptual contrast filtering. In *PCSPA 2011 2nd International conference on Pervasive Computing, Signal Processing and Applications*, Gjøvik, Norway, Sep 2011.

[230] E. Peli. Contrast in complex images. *Journal of the Optical Society of America A*, 7(10):2032–2040, Oct 1990. 2, 15, 27, 28, 30, 33, 40, 131

[231] E. Peli. In search of a contrast metric: Matching the perceived contrast of gabor patches at different phases and bandwidths. *Vision Research*, 37(23):3217–3224, Dec 1997. 34

[232] E. Peli. Contrast sensitivity function and image discrimination. *Journal of the Optical Society of America A*, 18(2):283–193, Feb 2001. 30, 40

[233] E. Peli. From contrast to sensitivity: Vision and image processing. In *Gjøvik Color Imaging Symposium 2009*, Gjøvik, Norway, Jun 2009. 31, 32

[234] E. Peli, L. E. Arend, G. M. Young, and R. B. Goldstein. Contrast sensitivity to patch stimuli: effects of spatial bandwidth and temporal presentation. *Spatial Vision*, 7(1):1–14, 1993. 30

[235] E. Peli and G. A. Geri. Discrimination of wide–field images as a test of a peripheral-vision model. *Journal of the Optical Society of America A*, 18(2):294–301, Feb 2001. 40

[236] M. R. Peres. *Focal encyclopedia of photography: digital imaging, theory and applications, history, and science*. Focal Press, 4th edition, 2007. 3, 4, 8, 20, 23, 24

[237] N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian, L. Jin, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo. A new color image database tid2013: Innovations and results. In J. Blanc-Talon, A. Kasinski, W. Philips, D. Popescu, and P. Scheunders, editors, *Advanced Concepts for Intelligent Vision Systems*, volume 8192 of *Lecture Notes in Computer Science*. Springer International Publishing, 2013. 82

[238] N. Ponomarenko, V. Lukin, K. Egiazarian, J. Astola, M. Carli, and F. Battisti. Color image database for evaluation of image quality metrics. In *10th International Workshop on Multimedia Signal Processing*, pages 403–408, Cairns, Queensland, Australia, Oct 2008. http://www.ponomarenko.info/tid2008.htm. 57, 68, 74

[239] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, and V. Lukin. On between-coefficient contrast masking of dct basis functions. *Proceedings of the 3rd International Workshop on Video Proceedings and Quality Metrics*, USA, 4 p.(Scottsdale, AZ, USA), Jan 2007. 63

[240] E. Provenzi, L. D. Carli, and A. Rizzi. Mathematical definition and analysis of the Retinex algorithm. *Journal of the Optical Society of America A*, 22(12):2613–2621, Dec 2005. 95

[241] E. Provenzi, M. Fierro, A. Rizzi, L. D. 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(1):162–171, Jan 2007. 97, 122

[242] E. Provenzi, C. Gatta, M. Fierro, and A. Rizzi. A spatially variant white-patch and gray-world method for color image enhancement driven by local contrast. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1757–1770, Oct 2008. 90

[243] Z. U. Rahman, D. J. Jobson, and G. A. Woodell. Multi–scale Retinex for color image enhancement. Technical report, 1996. 100

[244] Z. U. Rahman, D. J. Jobson, and G. A. Woodell. Multiscale Retinex for color rendition and dynamic range compression. In A. G. Tescher, editor, *Applications of Digital Image Processing XIX*, volume 2847 of *SPIE Proceedings*, pages 183–191, Denver, CO, USA, Aug 1996. 100, 122

[245] Z. U. Rahman, G. A. Woodell, and D. J. Jobson. A comparison of the multiscale Retinex with other image enhancement techniques. In *50th Annual Conference: A Celebration of All Imaging*, volume 50 of *IS&T Proceedings*, pages 426–431, Cambridge, MA, USA, May 1997. 122

[246] U. Rajashekar, I. van der Linde, A. C. Bovik, and L. K. Cormack. GAFFE: A gaze–attentive fixation finding engine. *IEEE Transactions on Image Processing*, 17(4):564–573, Apr 2008. 67, 137

[247] I. Ramirez. Wavelet based contrast measurement. In *VBM 2006 International Symposium Vision by Brains and Machines*, Montevideo, Uruguay, Nov 2006. 25

[248] D. V. Rao, N. Sudhakar, I. R. Babu, and L. P. Reddy. Image quality assessment complemented with visual regions of interest. In *ICCTA 2007 International Conference onComputing: Theory and Applications*, pages 681–687, March 2007. 137

[249] P. Reinagel and A. M. Zador. Natural scene statistics at the centre of gaze. *Network*, 10(4):341–350, Nov 1999. 38, 39

[250] E. Reinhard, G. Ward, S. Pattanaik, and P. Debevec. *High Dynamic Range Imaging - Acquisition, Display and Image-Based Lighting*. Morgan Kaufmann Publisher, 2005. 86, 90

[251] A. Rizzi, T. Algeri, G. Medeghini, and D. Marini. A proposal for contrast measure in digital images. In *CGIV 2004 2nd European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, Aachen, Germany, Apr 2004. 36, 37, 126, 131, 132

[252] A. Rizzi, C. Gatta, and D. Marini. From Retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalisation. *Journal of Electronic Imaging*, 13(1):75–84, Jan 2004. 85

[253] A. Rizzi and J. J. McCann. On the behavior of spatial models of color. In R. Eschbach and G. G. Marcu, editors, *Color Imaging XII: Processing, Hardcopy, and Applications*, volume 6493 of *SPIE Proceedings*, page 649302, San Jose, California, USA, Jan 2007. 85

[254] R. T. Rockafellar and R. J.-B. Wets. *Variational systems: an introduction*, volume 1091 of *Lecture Notes in Mathematics*, chapter 1, pages 1–54. Springer Berlin Heidelberg, 1984. 110

[255] R. Rodieck. Quantitative analysis of cat retinal ganglion cell response to visual stimuli. *Vision Research*, 5(12):583–601, Dec 1965. 35

[256] V. Rosselli, M.-C. Larabi, and C. Fernandez-Maloigne. Objective quality measurement based on anisotropic contrast perception. In *CGIV 2008 4th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 108–111, Terrassa, Spain, Jun 2008. 60

[257] G. S. Rubin and K. Siegel. Recognition of low-pass filtered faces and letters. *Investigative Ophthalmology and Visual Science*, (Suppl. 25):71–84, 1984. 27

[258] D. L. Ruderman, T. W. Cronin, and C.-C. Chiao. Statistics of cone responses to natural images: implications for visual coding. *Journal of the Optical Society of America A*, 15(8):2036–2045, Aug 1998. 62

[259] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D Nonlinear Phenomena*, 60(1–4):259–268, Nov 1992. 110

[260] J. C. Russ. *The image processing handbook*. CRC Press, 6th edition, 1995. 78, 115

[261] G. Sapiro and D. L. Ringach. Anisotropic diffusion of multivalued images with applications to color filtering. *IEEE Transactions on Image Processing*, 5(11):1582–1586, Nov 1996. 114

[262] S. Saponara, L. Fanucci, S. Marsi, G. Ramponi, D. Kammler, and E. Witte. Application–specific instruction–set processor for Retinex–like image and video processing. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 54(7):596–600, 2007. 101

[263] C. C. Semmelroth. Prediction of lightness and brightness on different backgrounds. *Journal of the Optical Society of America A*, 60(12):1685–1689, Dec 1970. 52

[264] J. Shanda. *Colorimetry: Understanding the CIE System*. John Wiley & Sons, 2004. 51, 53, 135

[265] R. Shapley and C. Enroth-Cugell. *Visual adaptation and retinal gain controls*, volume 3, chapter 9, pages 263–346. 1984. 35, 36

[266] A. Sharma. *Understanding Color Management*. Delmar Cengage Learning, 1st edition, 2003. 50, 54

[267] H. R. Sheikh, M. F. Sabir, and A. C. Bovik. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15(11):3440–3451, Nov 2006. 47, 72

[268] H. R. Sheikh, Z. Wang, A. C. Bovik, and L. K. Cormack. Image and video quality assessment research at live, http://live. ece.utexas.edu/research/quality, 2007. 68

[269] G. Simone, G. Audino, I. Farup, F. Albregtsen, and A. Rizzi. Termite Retinex: A new implementation based on a colony of intelligent agents. *Journal of Electronic Imaging*, 23(1):013006–1–13, Jan 2014.

[270] G. Simone, V. Caracciolo, M. Pedersen, and F. A. Cheikh. Evaluation of a difference of gaussians based image difference metric in relation to perceived compression artifacts. In *ISVC 2010 6th International Symposium Advances in Visual Computing*, Springer Proceedings, pages 491–500, Las Vegas, NV, USA, Nov 2010.

[271] G. Simone, R. Cordone, R. P. Serapioni, and M. Lecca. On edge–aware path–based color spatial sampling for Retinex: from Termite Retinex to Light Energy–driven Termite Retinex. *Journal of Electronic Imaging: Special Issue–Retinex at 50*, Oct 2016. Submitted. 150

[272] G. Simone and I. Farup. Spatio–temporal Retinex–like envelope with total variation. In *CGIV 2012 6th European Conference on Color in Graphics, Imaging and Vision*, IS&T Proceedings, pages 176–181, Amsterdam, The Netherlands, May 2012.

[273] G. Simone, C. Oleari, and I. Farup. Performance of the Euclidean color-difference formula in log-compressed OSA–UCS space applied to modified image-difference metrics. In *AIC 2009 11th Congress of the International Colour Association*, page 81, Sydney, Australia, Oct 2009. 60, 65

[274] G. Simone, M. Pedersen, I. Farup, and C. Oleari. Multi–level contrast filtering in image difference metrics. *EURASIP Journal on Image and Video Processing*, 2013(1):39, Jul 2013. 43

[275] G. Simone, M. Pedersen, and J. Y. Hardeberg. Measuring perceptual contrast in uncontrolled environments. In *EUVIP 2010 2nd European Workshop on Visual Information Processing*, pages 102–107, Paris, France, Jul 2010.

[276] G. Simone, M. Pedersen, and J. Y. Hardeberg. Measuring perceptual contrast in digital images. *Journal of Visual Communication and Image Representation*, 23(3):491–506, Apr 2012.

[277] R. Sobol. Improving the Retinex algorithm for rendering wide dynamic range photographs. *Journal of Electronic Imaging*, 13(1):65–74, Jan 2004. 99

[278] R. Sobol. Method for variable contrast mapping of digital images, patent US 20020154323a1, Oct 2006. 99

[279] M. Stokes, M. D. Fairchild, and R. S. Berns. Colorimetrically quantified visual tolerances for pictorial images. *Technical Association of the Graphic Arts*, 2:757–778, 1992. 50

[280] Y. Tadmor and D. Tolhurst. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Research*, 40(22):3145–3157, Oct 2000. 35, 126, 128, 132, 139

[281] I. Tastl, M. Bhachech, N. Moroney, and J. Holm. ICC colour management and CIECAM02. In *CIC 13 Thirteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 13 of *IS&T Proceedings*, page 318, Scottsdale, AZ, USA, Nov 2005. 51

[282] K.-H. Thung and P. Raveendran. A survey of image quality measures. In *TECHPOS 2009 International Conference for Technical Postgraduates*, IEEE Proceedings, pages 1–4, Kuala Lumpur, Malaysia, Dec 2009. 55

[283] K. Tiippana, R. Näsänen, and J. Rovamo. Contrast matching of two-dimensional compound gratings. *Vision Research*, 34(9):1157–1163, May 1994. 27

[284] A. Toet and M. Lucassen. A new universal colour image fidelity metric. *Displays*, 24(4–5):197–204, Dec 2003. 62

[285] M. Triola. Elementary statistics technology update. Pearson Education, 2011. 68, 71, 73

[286] J. Tumblin and H. Rushmeier. Tone reproduction for realistic images. *IEEE Computer Graphics and Applications*, 13(6):42–48, Nov 1993. 87

[287] S. Ullman. Analysis of visual motion by biological and computer systems. *Computer*, 14(8):57–69, Aug 1981. 35

[288] A. Valberg. *Light Vision Color*. John Wiley & Sons, 2007. 11, 14, 15, 27, 135

[289] Video Quality Experts Group. Validation of objective models of multimedia quality assessment, phase I. Technical report, International Telecommunication Union, 2008. 62, 72, 73

[290] H. Von Helmholtz. Versuch, das psychophysische Gesetz auf die Farbenunterschiede trichromatischer Augen anzuwenden. *Zeitschrift f´ur Psychologie und Physiologie der Sinnesorgane*, 3:1–20, 1892. 47

[291] D. Walther and C. Koch. Modeling attention to salient proto–objects. *Neural Networks*, 19(9):1395–1407, Nov 2006. 67, 137

[292] Y.-K. Wang and W.-B. Huang. Acceleration of the Retinex algorithm for image restoration by gpgpu/cuda. In *Parallel Processing for Imaging Applications*, number SPIE 7872, San Francisco, CA, USA, Jan 2011. 120

[293] Z. Wang and A. Bovik. A universal image quality index. *IEEE Signal Processing Letters*, 9(3):81–84, Mar 2002. 57, 61, 63

[294] Z. Wang and A. C. Bovik. *Modern Image Quality Assessment*. Morgan & Claypool, 2006. 55

[295] Z. Wang, A. C. Bovik, and L. Lu. Why is image quality assessment so difficult? In *IEEE International Conference on Acoustics, Speech, & Signal Processing*, Orlando, FL, USA, May 2002. 3, 47

[296] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, Apr 2004. 61

[297] Z. Wang and J. Y. Hardeberg. An adaptive bilateral filter for predicting color image difference. In *CIC17 17th Color Imaging Conference: Color Science and Engineering Systems, Technologies, and Applications*, volume 17 of *IS&T Proceedings*, pages 27–31, Albuquerque, NM, USA, Nov 2009. 60

[298] Z. Wang and Q. Li. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20(5):1185–1198, May 2011. 67

[299] Z. Wang and X. Shang. Spatial pooling strategies for perceptual image quality assessment. In *IEEE International Conference on Image Processing*, 2945–2948, Oct 2006. 66, 67

[300] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multi-scale structural similarity for image quality assessment. In *Proceedings of the 37th IEEE Asilomar Conference on Signals, Systems and Computers*, volume 2, pages 1398–1402, Pacific Grove, CA, Nov 2003. 62, 136

[301] G. Ward. High dynamic range imaging. In *CIC9 9th Color Imaging Conference*, volume 9 of *IS&T Proceedings*, Scottsdale, AZ, USA, Nov 2001. 86

[302] A. B. Watson. The Cortex Transform: Rapid computation of simulated neural images. *Computer Vision, Graphics, and Image Processing*, 39(3):311–327, 1987. 64

[303] Webvision. *The Organization of the Retina and Visual System*. University of Utah Health Sciences Center, 1995. 11

[304] P. Whittle. Increments and decrements: luminance discrimination. *Vision Research*, 26(10):1677–1691, 1986. 2, 27

[305] Wikipedia.org. Wikipedia: The free encyclopedia. In *https://en.wikipedia.org/wiki/Contrast_(vision)*, Last Visit: Sep 2014. 11

[306] S. Winkler. Analysis of public image and video databases for quality assessment. *IEEE Journal of Selected Topics in Signal Processing*, 6(6):616–625, Oct 2012. 68

[307] T. Wright. The photography handbook. Routledge, 1999. 23

[308] W. D. Wright. A re-determination of the trichromatic coeffcients of the spectral colours. *Transactions of the Optical Society*, 30(4):41–164, Mar 1928. 47

[309] G. Wyszecki and W. Styles. *Color Science: Concepts and Methods, Quantitative Data and Formulae*. Wiley Interscience, Derby, UK, 2nd edition, Aug 2000. 2, 14, 16, 106

[310] S. Yendrikhovskij. *Color reproduction and the naturalness constraint*. PhD thesis, Technische Universiteit Eindhoven, 1998. 44

[311] T. Young. The bakerian lecture: on the theory of light and colours. *Philosophical Transactions of the Royal Society of London*, 92:12–48, 1802. 47

[312] S. Zeki. A vision of the brain. Oxford, 1993. Blackwell Scientific Pub. 96

[313] C. Zewdie, M. Pedersen, and Z. Wang. A new pooling strategy for image quality metrics: Five number summary. In *EUVIP 2014 5th European Workshop on Visual Information Processing*, Paris, France, Dec 2014. 147

[314] L. Zhang and W. Lin. Selective visual attention: Computational models and applications. Wiley–IEEE Press, May 2013. 67

[315] L. Zhang, Y. Shen, and H. Li. VSI: A visual saliency–induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23(10):4270–4281, Oct 2014. 67

[316] X. Zhang and B. Wandell. Color image fidelity metrics evaluated using image distortion maps. *Signal Processing*, 70(3):201–214, Nov 1998. 2

[317] X. Zhang and B. A. Wandell. A spatial extension of CIELAB for digital color image reproduction. *Journal of the Society for Information Display*, 5(1):61–63, Mar 1997. 57, 60

[318] P. Zhao, M. Pedersen, J. Y. Hardeberg, and J.-B. Thomas. Measuring the relative image contrast of projection displays. *Journal of Imaging Science and Technology*, 59(3):1–13, May 2015. 2, 134

[319] S. Zuffi, E. Beltrame, and P. Scala. Visual experiments on the web: design of a web-based visual experiment management system. In S. P. Farnand and F. Gaykema, editors, *Image Quality and System Performance V*, volume 6808 of *SPIE Proceedings*, page 68080O, San Jose, CA, USA, Jan 2008. 70, 137

[320] S. Zuffi, C. Brambilla, R. Eschbach, and A. Rizzi. Controlled and uncontrolled viewing conditions in the evaluation of prints. In R. Eschbach, G. G. Marcu, and S. Tominaga, editors, *Color Imaging XIII: Processing, Hardcopy, and Applications*, volume 6807 of *SPIE Proceedings*, page 680714, San Jose, CA, USA, Jan 2008. 70

Chapter 10

Included papers in full text

10.1 Paper A: Measuring perceptual contrast in digital images

Measuring Perceptual Contrast in Digital Images

Gabriele Simone, Marius Pedersen, and Jon Yngve Hardeberg

The Norwegian Color Research Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, N-2802 Gjøvik, Norway

Abstract

In this paper we present a novel method to measure perceptual contrast in digital images. We start from a previous measure of contrast developed by Rizzi et al. (2004), which presents a multilevel analysis. In the first part of the work the study is aimed mainly at investigating the contribution of the chromatic channels and whether a more complex neighborhood calculation can improve this previous measure of contrast. Following this, we analyze in detail the contribution of each level developing a weighted multilevel framework. Finally, we perform an investigation of regions-of-interest in combination with our measure of contrast. In order to evaluate the performance of our approach, we have carried out a psychophysical experiment in a controlled environment and performed extensive statistical tests. Results show an improvement in correlation between measured contrast and observers perceived contrast when the variance of the three color channels separately is used as weighting parameters for local contrast maps. Using Regions-of-Interest as weighting maps does not improve the ability of contrast measures to predict perceived contrast in digital images. This suggests that Regions-of-Interest cannot be used to improve contrast measures, as contrast is an intrinsic factor and it is judged by the global impression of the image. This indicates that further work on contrast measures should account for the global impression of the image while preserving the local information.

Key words: Perceived Contrast, Multilevel Analysis, Difference of Gaussians, Regions-of-Interest, Psychophysical Experiment

1. Introduction

Since the initial research on image contrast, it has become clear how difficult it is to define perceptual contrast in images. One reason for this is the contextual influence on the observer task and observer experience.

A possible definition of contrast is the difference between the light and dark parts of a photograph, where less contrast gives a "flatter" picture, and more contrast gives a "deeper" picture. Many other definitions of contrast are also given, such as the difference in visual properties that makes an object distinguishable or simply the difference in color from point to point. As various definitions of contrast are given, measuring contrast is surely not a trivial task. Measuring the difference between the darkest and lightest point in an image does not predict perceived contrast since perceived contrast is influenced by the surround (viewing conditions) and the spatial arrangement of the image. Parameters as resolution, viewing distance, lighting conditions, image content, memory color etc. will affect how observers perceive contrast.

For this reason, the first approaches to this topic have confined themselves to study the phenomenon at a global level, operating in controlled situations and under the same constraints, the so-called "void conditions". After these very first experiments more complex measures have been devised, but a universal measure of contrast in images is still not clearly defined.

After introducing the state of the art of global and local contrast measures in Section 2, we present our proposal in Section 3. We start from a previous measure of contrast developed by Rizzi et al. (2004), which presents a multilevel analysis. In the first part of the work the study is aimed mainly at investigating the contribution of the chromatic channels and whether a more complex neighborhood calculation can improve this previous measure of contrast. Following this, we analyze in detail the contribution of each level developing a weighted multilevel framework. Finally we perform an investigation of regions-of-interest in combination with our measure of contrast. Section 4 describes the psychophysical experiment carried out while Section 5 presents the experimental results and it discusses how our proposal reflects perceptual contrast estimation performance. Finally in Section 6 conclusions are drawn.

2. Background

2.1. Global contrast measures

Several contrast measures have been proposed so far. The classic approaches consist of global measures, which implicitly or explicitly claim that the response of the Human Visual System (HVS) depends on the absolute luminance and not on the relation of its local variations. Global measures have been mainly developed during the second half of the 20th century and most of them take into account only the luminance in the calculation of contrast.

The very first measure of global contrast, in the case of sinusoids or other periodic patterns of symmetrical deviations ranging from the maximum luminance, L_{max}, to minimum luminance, L_{min}, is the Michelson (1927) formula:

$$C^M = \frac{L_{max} - L_{min}}{L_{max} + L_{min}}. \tag{1}$$

King-Smith and Kulikowski (1975), Burkhardt et al. (1984), and Whittle (1986) follow a similar concept replacing L_{max} or L_{min} with L_{avg}, which is the mean luminance in the image.

An alternative global measure, *Single Image Perceived Contrast*, has been developed recently by Calabria and Fairchild (2003), which modeled the following equation to measure the perceived contrast:

$$SIP_k = -1.505 + 0.131 \cdot k_C + 0.151 \cdot k_L + 666.216 \cdot k_S \tag{2}$$

where k_C, k_L, and, k_S are respectively the standard deviation of image chroma, lightness and high-passed lightness.

2.2. Local contrast measures

Global measures have some disadvantages, in fact, the study of contrast in an image at a global level provides only a measure related to the maximum global difference in lightness and in some cases chromaticity, which have been shown to be inadequate for measuring real visual configurations. Two single points of extreme brightness or darkness can determine the measure of contrast of the whole image, while the perceived contrast is clearly affected as illustrated in Figure 1.

To overcome the limits of global measures, local measures have been developed. We present here a selection of these measures focusing on those of particular relevance for our work. This selection with the key features of each measure is summarized in Table 1 at the end of this section.

Figure 1: Weakness of global measures. In all four pictures, global contrast measures would typically define contrast by the pixel values of the eyes (see arrows) as these have the highest and lowest luminance pixel values ($L_{max} = 100, L_{min} = 0$). The visual contrast of these four images are clearly different, however indicating that global measures cannot adequately predict perceived contrast.

2.2.1. Hess et al.

The issue of contrast of complex scenes at different spatial frequencies in the context of image processing and perception was addressed explicitly by Hess et al. (1983) who defined contrast in the Fourier domain as:

$$C(u,v) = \frac{2A(u,v)}{DC}, \tag{3}$$

where $A(u,v)$ is the amplitude of the Fourier transform of the image, u and v are the horizontal and vertical spatial frequency coordinates, respectively, and DC is the zero-frequency component.

2.2.2. Peli

Frankle and McCann (1983) and Adelson et al. (1984) proposed the use of the multilevel representation as an important implementation feature to mimic HVS. This consists of a set of lowpass or bandpass copies of an image, each representing pattern information at a different scale. This data structure used to represent image information is referred as "pyramid".

Peli (1990) proposed the idea of a pyramidal image-contrast structure where for each frequency band, the contrast is defined as the ratio of the bandpass-filtered image at that frequency to the lowpass image filtered to an octave below the same frequency (local luminance mean).

To define local band-limited contrast for a complex image, he obtains a band-limited version of the image in the frequency domain *A(u,v)*:

$$A(u,v) \equiv A(r,\theta) \equiv F(r,\theta)G(r), \tag{4}$$

where u and v are the respective horizontal and vertical spatial frequency coordinates, r and θ represent the respective polar spatial frequency coordinates: $r = \sqrt{u^2+v^2}$, $\theta = tan^{-1}\left(\frac{u}{v}\right)$, $F(r,\theta)$ is the Fourier transform of the image *I(x,y)*, and $G(r)$ is the band pass filter.

In the spatial domain the filtered image *a(x,y)* can be represented similarly, that is, as:

$$a(x,y) = I(x,y) * g(x,y), \tag{5}$$

where *g(x,y)* is the inverse Fourier transform of the band-pass filter *G(r)*.

In Peli's approach of measuring local contrast, the pyramid is obtained as follows:

$$A_i(u,v) \equiv A_i(r,\theta) \equiv F(r,\theta)G_i(r), \tag{6}$$

where $G_i(r)$ is a *cosine log* filter centered at frequency of $2^i cycles/picture$, expressed as:

$$G_i(r) = \frac{1}{2}\left(1 + cos\left(\pi log_2 r - \pi i\right)\right). \tag{7}$$

The resulting contrast at the band of spatial frequencies can be represented as a two-dimensional array $c_i(x,y)$:

$$c_i(x,y) = \frac{a_i(x,y)}{l_i(x,y)}, \tag{8}$$

where $a_i(x,y)$ is the corresponding local luminance mean image and $l_i(x,y)$ is a low-pass-filtered version of the image containing all energy below the band.

2.2.3. Ahumada and Beard

Ahumada and Beard (1998) proposed a method for measuring contrast which can be described as follows: the image I is first convolved with a Gaussian low pass filter F_B:

$$B(x,y) = I(x,y) * F_B(x,y), \tag{9}$$

and the blurred image B is convolved with a second Gaussian low pass filter F_L:

$$L(x,y) = B(x,y) * F_L(x,y). \tag{10}$$

Then, for every pixel of the image, the local contrast is defined as:

$$C(x,y) = \frac{B(x,y)}{L(x,y)} - 1. \tag{11}$$

2.2.4. Tremeau

Tremeau (2000) proposed a formula for measuring contrast based on a region adjacency graph defined as follows:

$$C_{ij} = \frac{4P_{ij}^2}{min\left(P_i, P_j\right) \times max\left(P_{\{i,j\}}\right)}, \tag{12}$$

where P_i and P_j are the perimeters of the regions i and j, P_{ij} is the length of the shared boundary between regions i and j, $max\left(P_{\{i,j\}}\right) = max\left\{P_{ik}/k \neq j, P_{kj}/k \neq i\right\}$ is the longest shared boundary of a region adjacent either to region i, or to region j.

The resulting weighted color contrast is given by the following equation:

$$WC_{ij} = d^2\left(i,j\right) \times C_{ij}, \tag{13}$$

where $d^2\left(i,j\right)$ is, commonly the Fisher distance, used to quantify the color distance between two regions.

2.2.5. Boccignone et al.

An alternative approach of measuring contrast is presented by Boccignone et al. (2001) and Ferraro and Boccignone (2004). They state that any image can be considered as an isolated thermodynamical system by identifying the image intensity with some thermodynamical variable. For measuring contrast they use the following formula:

$$\tilde{c}(x,y) = \int_0^\infty I(x,y,t)\sigma_{an}(x,y,t)dt, \tag{14}$$

with:

$$\sigma_{an}(x,y,t) = \sigma'_{an}(x,y,t) - \sigma''_{an}(x,y,t) = \\ = \chi(x,y)\frac{\nabla I(x,y,t)\cdot\nabla I(x,y,t)}{I(x,y,t)^2} - \chi(x,y)\frac{\nabla I(x,y,t)\cdot\nabla I^*(x,y,t)}{I(x,y,t)I^*(x,y,t)}, \quad (15)$$

where $\sigma'_{an}(x,y,t)$ represents the density of entropy production during anisotropic diffusion, whereas $\sigma''_{an}(x,y,t)$ tends to prevent entropy production, and χ is a non-negative decreasing function of the magnitude of the local image gradient, which forces convergence of the diffusion process toward some desired image representation and $*$ denotes the stationary point of the transformation.

2.2.6. Tadmor and Tolhurst

Tadmor and Tolhurst (2000) based their analysis of contrast on the Difference of Gaussians (DOG) model, which is modified and adapted to natural images.

In the conventional DOG model, the spatial sensitivity in the center component of the receptive-field is described by a bi-dimensional Gaussian with unit amplitude:

$$Center(x,y) = exp\left[-\left(\frac{x}{r_c}\right)^2 - \left(\frac{y}{r_c}\right)^2\right], \quad (16)$$

where the radius r_c represents the distance at which the sensitivity decreases to $1/e$ with respect to the peak level and (x,y) are the spatial coordinates of the receptive-field. The surround component is represented by another Gaussian curve, with a larger radius, r_s:

$$Surround(x,y) = 0.85\left(\frac{r_c}{r_s}\right)^2 exp\left[-\left(\frac{x}{r_s}\right)^2 - \left(\frac{y}{r_s}\right)^2\right]. \quad (17)$$

When the central point of the receptive-field is placed at the location (x,y), the output of the central component is calculated as:

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c}\sum_{j=y-3r_c}^{j=y+3r_c} Center(i-x, j-y)I(i,j), \quad (18)$$

while the output of the surround component is:

$$R_s(x,y) = \sum_{i=x-3r_s}^{i=x+3r_s}\sum_{j=y-3r_s}^{j=y+3r_s} Surround(i-x, j-y)I(i,j), \quad (19)$$

where in both cases *I(i,j)* is the image pixel value at position (*i,j*). The simplest case, with $r_c = 1$ and $r_s = 2$, results in a 7×7 center mask and a 13×13 surround mask.

The result of the DOG model is obtained as:

$$DOG(x,y) = R_c(x,y) - R_s(x,y). \tag{20}$$

The conventional DOG model assumes that the response of a neuron depends uniquely on the local luminance difference (ΔI) between the center and the surround. After the light adaptation process, the gain of the ganglion cells of the retina and the Lateral Geniculate Nucleus (LGN) neurons depends on the average local luminance I. Thus the model response depends on the contrast stimulus, and the DOG model must be modified by a division by the local mean luminance. They propose the following criterium for the measure of contrast:

$$c^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)} \tag{21}$$

In their experiments, using 256×256 images, the overall image contrast is calculated as the average local contrast of 1000 pixel locations taken randomly while assuring that the center and surround masks do not exceed the edges of the image:

$$C^{TT} = \frac{1}{1000} \sum_{n=1}^{1000} c_n^{TT} \tag{22}$$

The number of pixels to consider in the calculation should change according to the image size.

2.2.7. Rizzi et al.

Rizzi et al. (2004) have developed a very simple and efficient measure, able to estimate global and local components of contrast. It is based on two principles: to collect a simplified measure of difference among neighboring pixels and to do it on various frequency levels. We include this measure because there is evidence, that the use of multilevel is an important implementation feature to mimic the HVS (Frankle and McCann, 1983; Adelson et al., 1984).

In the rest of the paper we will refer to this contrast measure as RAMM. It performs a pyramid subsampling of the image to various levels in the CIELAB color space (CIE, 2004). A pyramidal structure is created by halving the image at each iteration. For each level, it calculates the local contrast in each pixel by taking the average difference between the lightness channel value of the pixel and

the surrounding eight pixels, thus obtaining a contrast map of each level. The final overall measure is a recombination of the average contrast for each level:

$$C^{RAMM} = \frac{1}{N_l} \sum_{l=1}^{N_l} \bar{c}_l, \tag{23}$$

where N_l is the number of levels and, $\bar{c}_l$ is the mean contrast in the level l:

$$\bar{c}_l = \sum_{i=1}^{i_{max}} \sum_{j=1}^{j_{max}} c_{i,j}, \tag{24}$$

where i_{max} and j_{max} indicate respectively the height and the width of the image, and $c_{i,j}$ is the contrast of each pixel calculated as:

$$c_{i,j} = \sum_{n \in N_8} \alpha_n \left| Pixel_{i,j} - Pixel_n \right| .. \tag{25}$$

The following values are used to define the weights of the neighboring pixels:

$$\alpha_n = \frac{1}{4+2\sqrt{2}} \begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix} \tag{26}$$

This measure has a computational complexity of $\Theta(NlogN)$, where N is the number of pixels, which is lower than alternative local methods, keeping a comparable level of correctness in the contrast estimate (Rizzi et al., 2004). The steps of the measure are illustrated in Figure 3(a).

Table 1: Discussed local contrast measure.

Local contrast measures	
Author	*Key features*
Hess et al. (1983)	Fourier domain
Peli (1990)	Multilevel analysis (pyramid representation), band-pass filtering
Ahumada and Beard (1998)	Gaussian filtering
Tremeau (2000)	Adjacency graph representation, Fisher distance
Boccignone et al. (2001)	Thermodynamical representation, anisotropic diffusion
Tadmor and Tolhurst (2000)	Difference of Gaussians (DOG) model
Rizzi et al. (2004)	Multilevel analysis (pyramid representation), 8-neighborhood

2.3. How many numbers for measuring contrast?

As well as concerning the notion of contrast itself, for which a clear and common definition is not found in the literature, there is an ongoing debate on how to reduce the concept of contrast from local values at each pixel location to a single number representing the whole contrast of the image.

It is possible to measure contrast in several ways, with different degrees of summarization. Four of the possible hypothesis are: one number per image, one number per color channel, one number per level using a multilevel approach or one number for each pixel. Here we want to address some of the arguments for and against each hypothesis without exhaustively discussing each detail.

Using one number per image is not sufficiently informative and it easily loses the ability to distinguish among various type of images that usually give rise to different contrast perceptions: e.g. geometric vs natural images. Furthermore several weights, such as the importance of each level and each color channel, must be chosen a priori to aggregate all the components into a single number. On the other hand it has the advantage of being a very compact measure, allowing it for instance to be used as a trigger for image dependent measures. A further advantage is that it can be easily used in comparison with perceptual experiments where observers usually provide a single number for each image.

Using one number per channel keeps chromatic information separated; depending on the color space used (e.g. luminance /chromaticity). For this reason this way of measuring it is color space dependent. Regarding observers, it is more difficult for them to express their judgment keeping channels separated.

Using one number per level allows linking the measure of contrast with the frequency scales in the image, and a posteriori recombine them with different weights. A drawback is that it is image-size dependent and it requires a vector to be stored rather than a single number. Observers can also find it extremely difficult to express a judgment separated per frequency content.

Using one number per pixel has the advantage of keeping the full locality of the information which can be further aggregated into various measures, e.g. for segmented areas or on image subsets. Moreover it can be used for pixelwise measure modification. As to disadvantages, it is obviously image-size dependent and it requires heavy memory use. Furthermore it needs to take into account masking phenomena and it is completely unsuitable for comparison with observer ratings.

As briefly discussed above, all approaches have positive and negative aspects which make none of them the clear winner.

When the desired number of values for measuring contrast is decided, the method on how to reduce them must be determined. These methods are commonly referred to as pooling. The main motivation for performing pooling is that one value is easier to relate to than hundreds of values.

The simplest pooling strategies are the arithmetic and geometrical mean but Minkowski pooling (de Ridder, 1992) is perhaps the most common and most popular for HVS based measures.

3. Proposed approaches

As presented in Section 2.2.7, the RAMM measure takes into account only the lightness channel. For this reason our investigation has focused mainly in two directions: first checking whether the use of the DOG model on the multilevel pyramid yields a better performance in considering more extended edges and gradients in the image and, second, whether the use of the chromatic channels in the computation of the perceived contrast leads to more accurate measures. Figure 2 shows an original image, an image filtered with an 8-neighborhood defined by Equation 25, and images filtered with the DOG model defined by Equation 21 with different center and surround radiuses. The superior performance of the DOG model over the 8-neighborhood for the edge detection is outstanding.

We will start proposing an extension of the RAMM measure, the so called "Retinal-like Subsampling Contrast (RSC)" measure, and then will build the so called "Weighted-Level Framework (WLF)" for measuring perceptual contrast in digital images.

3.1. Retinal-like Subsampling Contrast (RSC)

We have combined RAMM multilevel approach with Tadmor and Tolhurstï£¡s evaluation of a color stimulus (Rizzi et al., 2008). It works with the same pyramid subsampling as RAMM with the following differences: it computes in each pixel of each level the DOG contrast instead of the 8-neighborhood local contrast and it computes the DOG contrast not only for the lightness but also for the two chromatic channels. The three independent measures of each channel are then merged by a weighted linear combination. The final overall measure can be expressed by the formula:

$$C^{RSC} = \alpha \cdot C^{RSC}_{L^*} + \beta \cdot C^{RSC}_{a^*} + \gamma \cdot C^{RSC}_{b^*}, \tag{27}$$

where α, β, and γ represent the weights of each color channel. The steps of the measure are described in Figure 3(b).

(a) Original (b) 8-neighborhood

(c) DOG ($r_c = 1, r_s = 2$) (d) DOG ($r_c = 1, r_s = 3$)

(e) DOG ($r_c = 2, r_s = 3$) (f) DOG ($r_c = 2, r_s = 4$)

Figure 2: 8-neighborhood against DOGs. All images have been overboosted in luminosity to make them suitable for a printable version of the article.

The computational complexity of the RSC measure is the same as for RAMM:

$$\Theta\left(NlogN\right), \tag{28}$$

where N is the number of pixels, but with a slightly heavier multiplication constant due to the DOGs instead of the 8-neighbor difference computation.

As well as the previous presented measure, only one number of contrast is produced at the end, the averages of all the levels are averaged again among them with uniform weights.

This measure takes the name of Retinal-like Subsampling Contrast (RSC) and it derives from the fact that the DOG model has been used successfully in many studies to describe the receptive fields and responses of mammalian retinal ganglion cells and LGN neurones (Tadmor and Tolhurst, 2000) and from the way of building the pyramid structure (Rizzi et al., 2004).

3.2. The Weighted-Level Framework (WLF)

In this section we introduce the Weighted-Level Framework (WLF) and we address mainly two aspects: the subsampling method and the weights in the level recombination (Simone et al., 2009b).

An antialiasing filter is introduced in the subsampling in order to minimize distortion artifacts at low resolutions.

As demonstrated by Frankle and McCann (1983), Adelson et al. (1984) and (Peli, 1990), each level has a different contribution to contrast so we redefine Equation 23 as follows:

$$C_i = \frac{1}{N_l}\sum_{l=1}^{N_l}\lambda_l \cdot \bar{c}_l, \tag{29}$$

where N_l is the number of levels, $\bar{c}_l$ is the mean contrast in the level l and i indicates the applied color channel as before and the new parameter λ_l is the weight assigned to each level l.

The overall final measure is given by:

$$C^{WLF} = \alpha \cdot C_1 + \beta \cdot C_2 + \gamma \cdot C_3, \tag{30}$$

where α, β, and γ are the weights of each color channel.

The measure can be extended to different color spaces such as XYZ and RGB and it is not limited to CIELAB.

The general structure of our proposed measure can be seen in Figure 3(c) where the most important novelties are shown in red: the antialiasing filter in the pyramid and weighted recombination of local contrast maps.

In this framework the previously developed RAMM and RSC can be considered as just special cases with uniform weighting levels in the CIELAB color space.

3.2.1. Finding appropriate weights

In this work we have also chosen to summarize the contrast measure in a single number per image. We leave to the reader the possibility to use the proposed measure in a different way, keeping some information separated. This choice allowed us to easily calculate statistics for comparing the results of our framework with the observers perceived contrast.

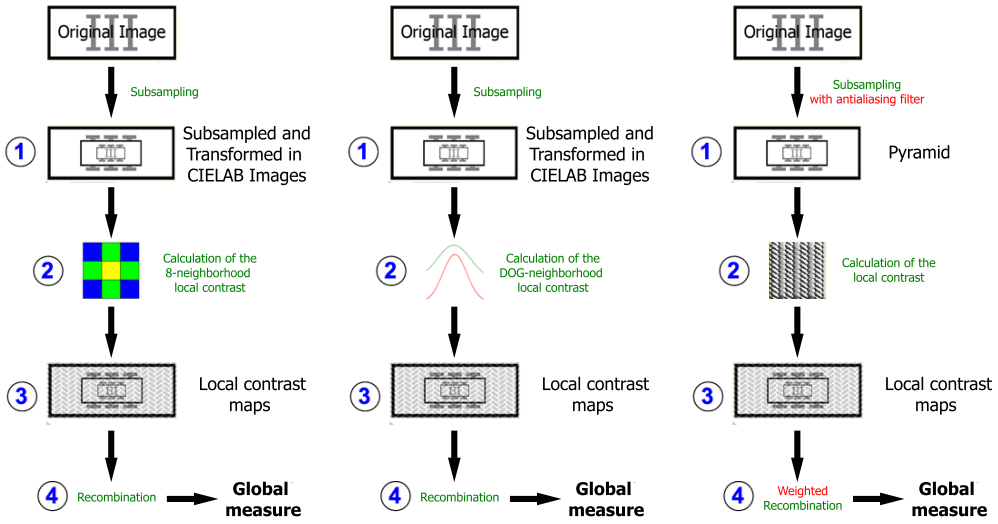
In order to obtain a single contrast number we had to properly tune α, β, γ, and λ. The proposed idea is to give them values from particular measures taken from the image itself. We have tested several alternatives for these parameters. For the parameter λ we have chosen:

- 1
- $\frac{1}{l}$, where l is the corresponding level of the pyramid
- $\frac{1}{m}$, where m is the mean of pixel values in each channel separately at each level of the pyramid
- τ as the variance of pixel values in each channel separately at each level of the pyramid

For the parameters α, β, and γ we have chosen respectively[1]:

- 1, 0, 0
- 0.333, 0.333, 0.333
- 0.5, 0.25, 0.25
- 0, 0.5, 0.5

[1]Note that the weights in 1, 3, and 4 above only applies to lightness-chromaticity color space such as in our case CIELAB.



(a) RAMM (b) RSC (c) WLF

Figure 3: Workflow for RAMM, RSC, and WLF. The difference between RAMM and RSC is found in the neighborhood calculation of local contrast, where RAMM uses a 8-neighborhood while RSC uses DOG-neighborhood. RAMM and RSC are just special cases with uniform weighting levels in CIELAB color space of WLF, which implements an antialiasing filter in the subsampling, a weighted recombination of the local contrast maps and it is extended also to RGB and XYZ color space.

- $\frac{1}{v}$, where v is the mean of pixel values in each channel separately
- ω as variance of pixel values in each channel separately

3.3. Applying Regions-of-Interest weighting to the WLF

Eye tracking has been used in a number of different color imaging research projects with great success, allowing researchers to obtain information on where observers gaze. Gaze information has been used by researches to improve image quality metrics (Pedersen and Hardeberg, 2008; Pedersen, 2007), using region-of-interest as a weighting map for the metrics. The motivation for incorporating eye tracking as a region-of-interest weighting is to take advantage of the fact that some areas in an image are more important than others.

We use a similar approach and apply gaze information as a weighting map for the contrast measures. Regions that draw attention will be weighted higher than regions that observers do not look at or pay attention to.

From the eye tracking data a number of different maps have been calculated, among them the time used at one pixel multiplied with the number of times the observer fixated on this pixel, the number of fixations at the same pixel, the mean time at each pixel, and the total time for each pixel. All of these have been normalized by the maximum value in the map, and a Gaussian filter corresponding to the 2-degree visual field of the human eye was applied to the map to even out differences (Bai et al., 2006; Babcock, 2000), and to simulate that we look at an area rather than one particular pixel (Henderson et al., 2003).

Gathering gaze information is time consuming and resource demanding, and because of this we have investigated other ways to obtain similar information. One possibility is saliency map, which is a map that represents visual saliency of a corresponding visual scene. One proposed model was introduced by Walther and Koch (2006) for bottom-up attention to salient objects, and this has been adopted for the saliency maps used in this study. The saliency map has been computed at level 1 (i.e. the size of the saliency map is equal to original images) and 7 fixations (i.e. giving the seven most salient regions in the image), for the other parameters standard values in the SaliencyToolbox (Walther and Koch, 2006) have been used.

Rajashekar et al. (2008) proposed Gaze-Attentive Fixation Finding Engine (GAFFE) based on statistical analysis of image features for fixation selection in natural scenes. The GAFFE uses four foveated low-level image features: luminance, contrast, luminance-bandpass and contrast-bandpass to compute the simulated fixations of a human observer. The GAFFE maps have been computed for 10, 15, and 20 fixations, where the first fixation has been removed since this always is

placed in the center resulting in a total of 9, 14, and 19 fixations. A Gaussian filter corresponding to the 2-degree visual field of the human eye was applied to simulate that we look at an area rather than at one single point, and a larger filter (approximately 7-degree visual field) was also tested.

In this work each contrast map of each level is weighted pixelwise with its relative gaze information, or saliency map, or gaze-attentive fixation finding engine as shown in Figure 4.

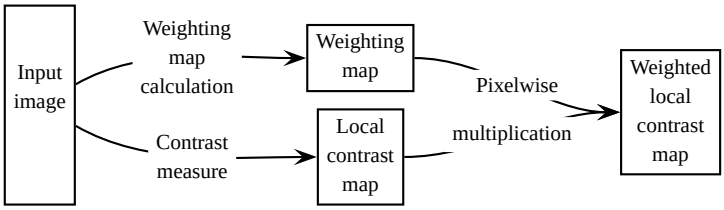


Figure 4: Framework for using weighting maps with contrast measures. A weighting map is multiplied pixelwise with the relative local contrast map in order to generate a weighted contrast map. As weighting maps we have used gaze maps, saliency maps, and GAFFE maps.

We will test the following maps: fixations only, time only, mean time, fixations multiplied with time, saliency, 10 fixations GAFFE map, 15 fixations GAFFE map, 10 fixations big Gaussian GAFFE map, 15 fixations big Gaussian GAFFE map, 20 fixations GAFFE map 20 fixations big Gaussian GAFFE map and six combinations of gaze maps and GAFFE maps.

4. Experimental methodology

In order to evaluate the performance of the measures in relation to perceived contrast a psychophysical experiment has been carried out. In this paper we quantify performance by estimating the correlation between the measure score and the observer score.

Several guidelines have been given in the literature for the selection of images for psychophysical experiment, in the context of investigating image quality issues. Holm et al. (2006) recommend the use of a broad range of natural images as well as test charts to reveal the quality issues. The Commission Internationale de l' Èclairage CIE (156:2004) suggests to include images with the following characteristics: high-key, low-key, low lightness contrast, leaves and sky, no neutrals,

no white point, no black point, heavy cast, few hues, business graphic, and flesh tones. Büring et al. (2004) propose to use natural images, as well as saturated colors.

We have chosen 15 different images (Figure 5), each representing different characteristics, such as different levels of lightness and colorfulness, following the recommendations from above.

17 observers were asked to rate the contrast of these 15 images. Nine of the observers were considered experts, i.e. had experience in color science, image processing, photography or similar and eight were non-experts or with little experience in these fields. All observers were recruited from Gjøvik University College, both students and staff in the ages between 18-45. Observers rated contrast from 1 to 100, where 1 was the lowest contrast and 100 maximum contrast. The observers were told to rate the contrast as they comprehended contrast, i.e. no definition of contrast was given by the researchers before commencing the experiment in order not to influence the observers. All observers had normal or corrected to normal vision. Each image was shown for 40 seconds with the surrounding screen black [2], and the observers stated the perceived contrast within this time limit. The experiment was carried out on a calibrated CRT monitor, LaCIE electron 22 blue II, in a gray room. The observers were seated approximately 80 cm from the monitor, and the lights were dimmed and measured to approximately 17 lux.

4.1. Statistical analysis

In order to compare the scores of each measure, the perceived contrast score was recorded for each observer for each image and then two types of correlation coefficients (CC) were computed (Kendall et al., 1991): the Pearson-product-moment CC and the Spearman-rank CC. The Pearson CC assumes that the variables are ordinal and evaluate the linear relationship between two variables. The Spearman CC is a non-parametric measure of correlation and it is used as a measure of linear relationship between two sets of ranked data, instead of the actual values. This describes the relationship between variables with no assumptions on the frequency distribution of the variables and on how tightly the ranked data clusters are around a straight line.

For both correlation coefficients the relative p-value has been calculated. With

[2] The black surround was provided to the observer in order to have as reference the black point (darkest color reproducible).

(a) Image 1.

(b) Image 2.

(c) Image 3.

(d) Image 4.

(e) Image 5.

(f) Image 6.

(g) Image 7.

(h) Image 8.

(i) Image 9.

(j) Image 10.

(k) Image 11.

(l) Image 12.

(m) Image 13.

(n) Image 14.

(o) Image 15.

Figure 5: Images used in the experiment. Images 1 and 2 are provided by Ole Jakob Bøe Skattum, image 10 is provided by CIE, images 8 and 9 from the ISO 12640-2 standard, images 3, 5, 6, and 7 from Kodak PhotoCD, images 4, 11, 12, 13, 14, and 15 from the ECI Visual Print Reference.

a *p*-value of 5% (or 0.05) there is only a 5% chance that results you are seeing would have come up in a random distribution.

4.1.1. Statistical methods for weighting maps

After a short investigation of the results we found that the data cannot be assumed to be normally distributed, and therefore a special care must be given to the statistical analysis. One common method for statistical analysis is the Z-score (Engeldrum, 2000), this requires the data to be normally distributed, and in our case this analysis would not give valid results. Simply using the mean opinion score would also result in problems, since the dataset cannot be assumed to be normally distributed. Because of this we use the rank from each observer to carry out a Wilcoxon signed rank test, a non-parametric statistical hypothesis test. This test does not make any assumption on the distribution, and it's therefore an appropriate statistical tool for analyzing this data set.

In order to test the performance of the contrast measures with different weighting maps and parameters an extensive statistical analysis has been carried out. First the images with high and low perceived contrast has gone through a rank sum test for all 15 images, the *p* values from this test indicates the ability of the contrast measure to differentiate between the images with perceived low and high contrast. These *p* values have been used as a basis for a sign test for all parameters for each map and the same for each parameter for all maps. The results from this analysis indicate whether a weighting map is significantly different from no map (standard method), or if a set of parameters is significantly different from other sets of parameters. In the case of a significant difference additional analysis is carried out to indicate whether the performance is better or worse for the tested weighting map or set of parameters.

5. Results and analysis

5.1. Observer rating

Table 2 shows the mean perceived contrast and standard deviations for all observers, experts and non-experts for each image.

Figure 6 shows the perceived contrast stated by all the observers with a 95% confidence interval. The image rated with the highest mean by the observers is image 15, but it can not be differentiated from many of the other images due to the confidence intervals. The image with the lowest rated contrast is image 13, but this cannot be differentiated from a number of other images as well. This indicates that the perceived contrast of many of the images are statistically similar. Image 12

Table 2: Perceived contrast results for all observers, experts and non-experts. Considering all the observers, image 15 has the highest mean value , while image 13 has the lowest mean value. Image 15 also has the lowest standard deviation. Considering the experts, image 5 has the highest mean value while image 13 has the lowest mean value. Considering the non-experts, image 15 has the highest mean value, and also the lowest standard deviation while image 4 has the lowest mean value. Highest values are highlighted in cyan while lowest values are highlighted in red.

	All observers		Experts		Non-experts	
Image	Mean value	Standard deviation	Mean value	Standard deviation	Mean value	Standard deviation
1	58.7	19.2	65.0	13.8	51.9	22.8
2	57.1	15.4	63.3	15.4	50.0	12.8
3	61.8	14.3	70.6	8.5	51.9	13.1
4	50.3	23.1	58.9	18.3	40.6	25.1
5	70.5	18.7	79.8	10.5	60.0	20.9
6	53.9	19.1	62.4	15.8	44.4	18.6
7	63.8	16.4	67.8	13.0	59.4	19.5
8	57.7	19.1	61.7	15.0	53.1	23.1
9	65.0	22.6	74.1	12.7	54.8	27.5
10	57.7	20.7	67.3	12.0	46.9	23.8
11	59.7	18.6	66.7	17.0	51.9	18.1
12	57.7	24.0	59.8	17.5	55.4	30.9
13	49.0	17.5	46.9	20.9	51.3	13.8
14	61.5	21.7	66.7	16.9	55.6	26.0
15	71.7	10.2	71.7	9.4	71.6	11.6

has the highest standard deviation value, indicating the largest difference between the answers from the observers.

Figure 7 shows the perceived contrast stated by the experts with a 95% confidence interval. Image 13 has been rated as the image with the lowest contrast, while image 5 has the highest contrast according to the experts. The experts agree most upon image 3, while the highest standard deviation is found in image 13, indicating high deviation in the contrast score for this image.

Figure 8 shows the perceived contrast stated by the non-experts with a 95% confidence interval. Image 4 is given the lowest contrast, this is also the darkest image i.e. having the lowest mean L^* value. Image 15 is given the highest contrast, this image also has the lowest standard deviation. Image 12 has the highest standard deviation, indicating a high degree of disagreement among the observers.

There is a clear difference between the experts and non-experts, in 14 of the 15 images the experts have a higher mean value than the non-experts. For image 10 the difference in mean perceived contrast is 20.5 (experts=67.3, non-experts=46.9), resulting in a large difference in perceived contrast between the two groups. Image 13 is the only image where the non-experts have a higher

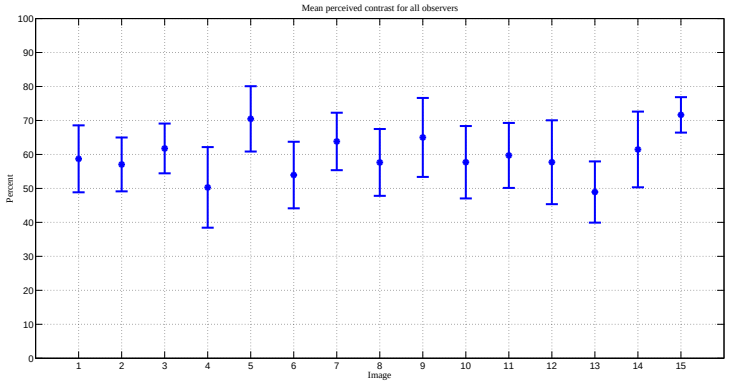


Figure 6: Mean perceived contrast score for all observers. Image 15 has the highest mean value, while image 13 has the lowest mean value. Image 15 also has the lowest standard deviation, indicating that observers agree about the high contrast in this image.

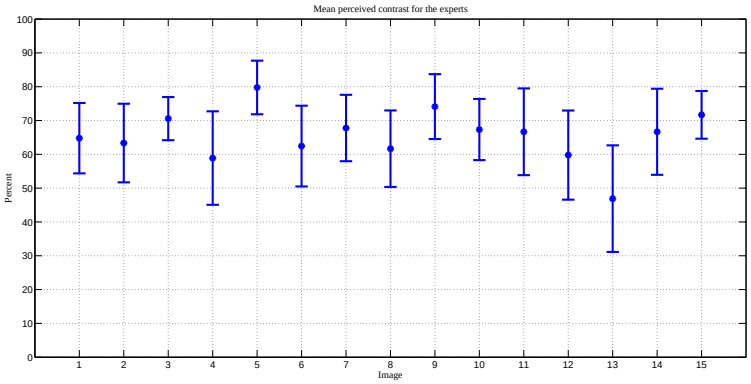


Figure 7: Mean perceived contrast score for the experts. Image 5 has the highest mean value. Image 13 has the lowest mean value, and also has the highest standard deviation.

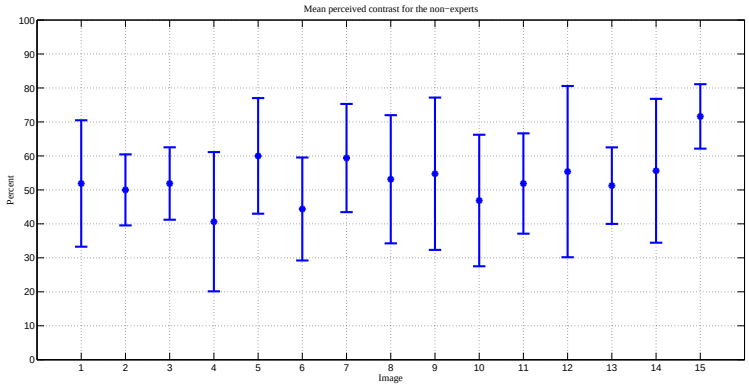


Figure 8: Mean perceived contrast score for the non-experts. Image 15 has the highest mean value, and also the lowest standard deviation. Image 4 has the lowest mean value.

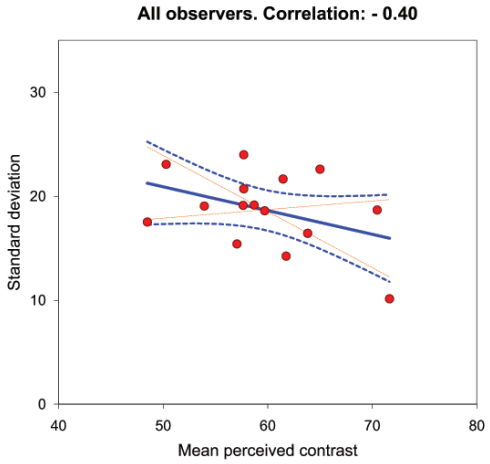
mean than the experts, but the difference is low, only 4.4 (experts=46.9, non-experts=51.3). The non-experts have used more of the scale than experts, i.e. they have a larger mean difference between the maximum value and minimum value. Experts rate the contrast to be higher then non-experts in most scenes. Experts also agree more upon the contrast in the image than non-experts. We can thus conclude that the perceived contrast in an image depends on the background of the observer.

We have also investigated the connection between mean perceived contrast and standard deviation. For all observers there is a correlation of -0.40 (Figure 9(a)), while for non-experts only -0.23 (Figure 9(c)). For the expert observers we have a correlation of -0.83, indicating a high consensus among the experts for the images with higher perceived contrast, and as the perceived contrast decreases the standard deviation increases (Figure 9(b)).

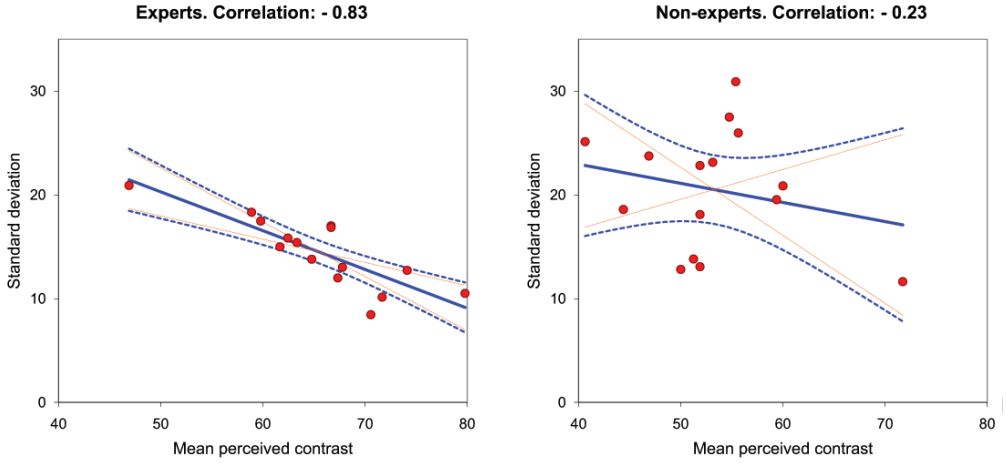
5.2. Performance of contrast measures

We have tested four different contrast measures with the following abbreviations, PELI (Peli, 1990), TT (Tadmor and Tolhurst, 2000), WLF(RAMM) and WLF(RSC) (Simone et al., 2009b). We remind that in the framework RAMM and RSC previously developed can be considered just special cases with uniform weighting levels.

Around 10000 variations have been calculated in the framework developed by Simone et al. (2009b), so given the huge amount of data we show here only results with Pearson correlation greater than 0.75 for WLF(RAMM) and greater than 0.8 for WLF(RSC), in addition to default parameters.



(a) All observers.



(b) Expert observers.

(c) Non-expert observers.

Figure 9: Correlation between the mean perceived contrast and standard deviation for all, expert, and non-expert observers. For experts high correlation is found, indicating that when the mean perceived contrast decreases the standard deviation increases. This indicates that expert observers agree on the rating of images with high contrast, but when the contrast decreases the observers rate the images more differently. The solid blue line is the linear regression line, the dashed red lines indicate the 95% confidence interval (CI) of the slope of the regression line, and the dotted blue curves indicate the total 95% CI of the linear regression estimate.

Table 3: Pearson and Spearman correlation for WLF using RAMM neighborhood. Yellow cells indicate the standard values. Grey cells indicate the highest Pearson and Spearman correlation. All numbers are rounded to two decimals. λ equal to τ indicates that each level of the pyramid is weighted with the variance of the three channels separately. $\frac{1}{v}$ indicates that the considered channel is weighted with the reciprocal of its mean, while ω indicates that it is weighted with its variance.

WLF(RAMM)								
Color space	λ	α	β	γ	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
CIELAB	1	1	0	0	0.41	0.13	0.40	0.14
CIELAB	1	0.33	0.33	0.33	0.54	0.04	0.47	0.08
RGB	τ	0.33	0.33	0.33	0.78	0.01	0.71	0.00
RGB	τ	ω	ω	ω	0.78	0.00	0.71	0.00
RGB	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.78	0.00	0.69	0.00
CIELAB	τ	0.33	0.33	0.33	0.78	0.00	0.58	0.03
CIELAB	τ	0.50	0.25	0.25	0.79	0.00	0.65	0.02
CIELAB	τ	ω	ω	ω	0.80	0.00	0.68	0.03

Table 4: Pearson and Spearman correlation for WLF using RSC neighborhood. Yellow cells indicate the standard values. Grey cells indicate the highest Pearson and Spearman correlation. Cyan row indicates the best result achieved by the framework. All numbers are rounded to two decimals. λ equal to τ indicates that each level of the pyramid is weighted with the variance of the three channels separately. $\frac{1}{v}$ indicates that the considered channel is weighted with the reciprocal of its mean, while ω indicates that it is weighted with its variance.

WLF(RSC)										
Color space	r_c	r_s	λ	α	β	γ	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
CIELAB	1	2	1	0.333	0.333	0.333	-0.23	0.42	-0.21	0.44
CIELAB	1	2	1	0.50	0.25	0.25	-0.16	0.57	-0.20	0.48
CIELAB	1	2	1	1	0	0	0.02	0.95	-0.21	0.45
RGB	1	2	τ	0.33	0.33	0.33	0.82	0.00	0.68	0.01
RGB	2	4	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.82	0.00	0.72	0.00
CIELAB	3	4	τ	0.50	0.25	0.25	0.82	0.00	0.72	0.00
RGB	2	3	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.83	0.00	0.77	0.00
CIELAB	2	3	τ	0.50	0.25	0.25	0.83	0.00	0.73	0.00
CIELAB	2	4	τ	0.50	0.25	0.25	0.83	0.00	0.82	0.00
RGB	2	4	τ	0.33	0.33	0.33	0.84	0.00	0.79	0.00
RGB	2	4	τ	ω	ω	ω	0.84	0.00	0.80	0.00

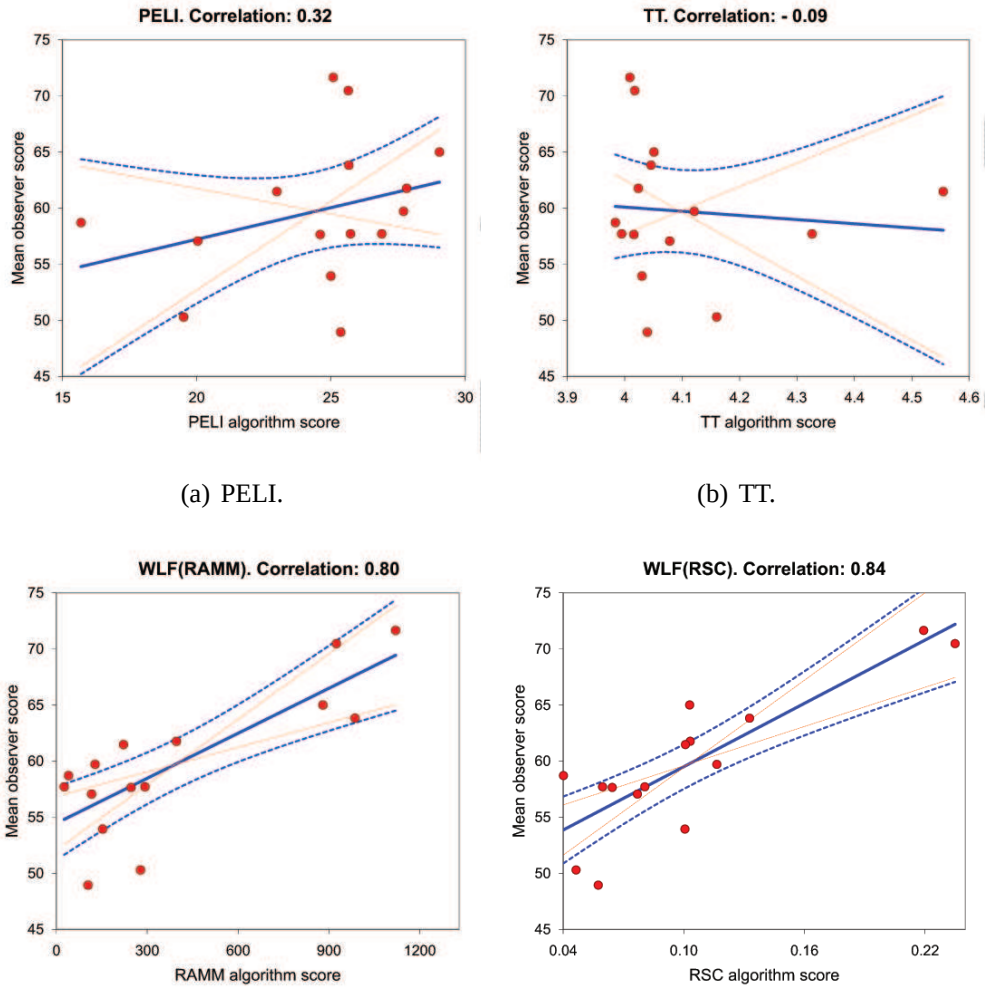
Table 3 and Table 4 show respectively the results using RAMM neighborhood and RSC neighborhood. First we can notice from both tables the absence of results with XYZ color space, which seems to be inappropriate for contrast measure. RGB and CIELAB appear to be equal as results show a difference not greater than ± 0.03. We can point out that the parameter λ, which gives a weight to each level of the pyramid, is the key of improving contrast measure especially if λ is equal to τ or, to be more precise equal, to the variance of each channel separately (setting 4 in the list above). Furthermore, unlike previous studies, α, β, and γ lose their importance as we can see that results do not differ so much. It has also to be underlined that the RSC measure works better with greater values of r_c and r_s compared to the standard values given by Tadmor and Tolhurst (2000) in their previous studies. The same discussion can be done for the Spearman correlation which is always lower than Pearson correlation. For both Pearson and Spearman correlation very low p-values confirm the usefulness of the measures.

The Weighted-Level Framework using the RSC neighborhood (Figure 10(d)) reaches higher performance than using RAMM neighborhood (Figure 10(c)), stating that a DOG neighborhood perform a better evaluation of perceived contrast than a simple 8-neighborhood.

For the PELI measure the size of image must be a power of two, because of this the images have been resized to 512×512. The PELI measure rates the image with the highest and lowest perceived contrast to have approximately the same score, an image with a medium perceived contrast is rated as the image with the lowest contrast by PELI. This results in a Pearson correlation of 0.32 (Table 5 and Figure 10(a)). The PELI measure has a scattering of the data points, where the image rated with the highest contrast and lowest contrast by the observers receive similar PELI scores. This results in a low Pearson correlation (Table 5). For the Spearman correlation the PELI measure shows an increased correlation, where the correlation is 0.43 (Table 5). This is due by the fact that the Spearman CC looks at the rank of the results, and even though PELI has a spread scatter plot, the order (rank) of the results is more similar to the observers. In this sense the distribution of Peli contributes to a low Pearson correlation, but when we discard the distribution and only look at the rank it performs better. However, it has lower performance than WLF.

The TT measure has a low correlation between the perceived contrast and predicted contrast (Table 5 and Figure 10(b)). Four of the images with the highest perceived contrast have been rated in the lower half of the TT scale, resulting in a low performance.

The analysis in this section needs to be put in perspective with the variations



(a) PELI. (b) TT.

(c) WLF(RAMM). (d) WLF(RSC).

Figure 10: Pearson correlation between observer mean score and PELI, TT, WLF(RAMM), and WLF(RSC) measure score respectively. The solid blue line is the linear regression line, the dashed red lines indicate the 95% confidence interval (CI) of the slope of the regression line, and the dotted blue curves indicate the total 95% CI of the linear regression estimate.

Table 5: Pearson and Spearman correlation of the four tested measures. The TT measure has the lowest correlation between the perceived contrast and predicted contrast while WLF(RSC) has the highest.

measure	Pearson correlation	Pearson *p*-value	Spearman correlation	Spearman *p*-value
PELI	0.32	0.01	0.43	0.01
TT	-0.09	0.73	-0.19	0.51
WLF(RAMM)	0.80	0.001	0.68	0.03
WLF(RSC)	0.84	0.00	0.80	0.00

in the observer results.

Table 6: Statistical significance of using different maps with WLF (RAMM). If the weighting maps are significantly different than using no weighting map then the value in the table is below 0.05. With these results we can see that the different weighting maps have the same performance as no map at a 5% significance level, indicating that weighting RAMM with maps does not improve predicted contrast.

Map	fixation only	fixation x time	saliency	GAFFE10	GAFFEBG20	no map
fixation only	-					
fixation x time	1.00	-				
saliency	0.63	1.00	-			
GAFFE10	0.25	0.25	0.25	-		
GAFFEBG20	0.13	0.38	1.00	0.06	-	
no map	0.50	0.50	0.63	1.00	0.06	-

5.3. Results with gaze information

The perceived contrast for the 15 images (Figure 5) from 17 observers were gathered. The 15 images have been grouped into three groups based on the Wilcoxon signed rank test: high, medium and low contrast. From the signed rank test observers can differentiate between the images with high and low contrast, but not between high/low and medium contrast. Images 5, 9, and 15 have high contrast while images 4, 6, 8, and 13 have low contrast. This is further used to analyze the performance of the different contrast measures and weighting maps.

We have tested many different weighting maps, and due to page limitations we cannot show all results. We will show results for fixations only, fixations multiplied with time, saliency, 10 fixation GAFFE map (GAFFE10), 20 fixations big Gaussian GAFFE map (GAFFEBG20) and no map (normally unweighted contrast measure). The maps that were excluded are time only, mean time, 15 fixations GAFFE map, 20 fixations GAFFE map, 10 fixations big Gaussian GAFFE map, 15 fixations big Gaussian GAFFE map, and 6 combinations of gaze maps and GAFFE maps. The maps that have been excluded show no significant difference from no map, or have a lower performance than no map.

As we can see from Tables 6 and 7, using maps is not significantly different from not using them as they have the same performance at a 5% significance level. We can only see a difference between saliency maps and gaze maps (fixation only and fixations $\times$ time), but since these are not significantly different from no map they do not increase the contrast measure's ability to predict perceived contrast. The contrast measures with the use of maps have been tested in the framework developed by Simone et al. (2009a) with different settings shown in Tables 8 and 9. For RAMM the standard parameters (LAB-1-1-0-0-0 and LAB-1-0.33-0.33-0.33) perform significantly worse than the other parameters in the table, having a p-value greater than 0.05. For RSC we noticed that three parameters are

Table 7: Statistical significance of using different maps with WLF(RSC). If the weighting maps are significantly different than using no weighting map then the value in the table is below 0.05. We can see that no maps are significantly different than no map, indicating that the have the same performance at a 5% significance level. We can see a difference between saliency maps and gaze maps (fixation only and fixation x time), but since these are not significantly different from no map they do not increase the contrast measure's ability to predict perceived contrast. Gray cells indicate significant difference at a 5% significance level.

Map	fixation only	fixation x time	saliency	GAFFE10	GAFFEBG20	no map
fixation only	-					
fixation x time	1.00	-				
saliency	0.02	0.03	-			
GAFFE10	0.29	0.51	1.00	-		
GAFFEBG20	0.23	0.227	0.73	0.69	-	
no map	0.50	1.00	0.13	0.73	0.34	-

significantly different from the standard parameters (LAB-1-2-1-0.33-0.33-0.33 and LAB-1-2-1-0.50-0.25-0.25), but after further analysis of the underlying data these perform worse than the standard parameters.

We can see from Figure 11 that using a saliency map for weighting discards relevant information used by the observer to judge perceived contrast since contrast is a complex feature and it is judged by the global impression of the image.

Furthermore, in order to validate the results from the other dataset we have carried out the same analysis for 25 images each with four contrast levels from the TID2008 database (Ponomarenko et al., 2008). The score from the two contrast measures have been computed for all 100 images, and a similar statistical analysis is carried out as above but for four groups (very low contrast, low, high and very high contrast). The results from this analysis supports the findings from the first dataset.

In conclusion using weighting maps from gaze information, saliency maps, and GAFFE maps do not improve the WLF to predict perceived contrast in digital images.

6. Conclusions

We have developed a Weighted-Level Framework (WLF) to measure perceptual contrast. It performs for each channel a pyramid subsampling of the image to various levels. A pyramidal structure is created halving the image at each iteration with prefiltering in order to avoid aliasing at low resolutions. For each level, it calculates the local contrast of each pixel using the Difference of Gaussians, thus obtaining a contrast map of each level. The overall measure of each channel is a

Table 8: Statistical significance in terms of *p*-values of using different parameters with WLF(RAMM). Gray cells indicate significant difference at a 5% (0.05) significance level between the results of the two configurations. RAMM parameters are the following: color space (CIELAB or RGB), pyramid weight, and the three last parameters are channel weights. "var" indicates the variance.

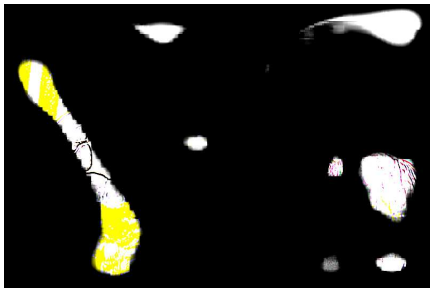
Parameters	LAB-1-1-0-0	LAB-1-0.33-0.33-0.33	RGB-4-var1-var2-var3	LAB-4-0.33-0.33-0.33	LAB-4-0.50-0.25-0.25	LAB-4-var1-var2-var3
LAB-1-1-0-0	-					
LAB-1-0.33-0.33-0.33	0.09	-				
RGB-4-var1-var2-var3	0.00	0.01	-			
LAB-4-0.33-0.33-0.33	0.00	0.01	1.00	-		
LAB-4-0.50-0.25-0.25	0.00	0.00	0.50	1.00	-	
LAB-4-var1-var2-var3	0.00	0.00	0.50	1.00	1.00	-

Table 9: Statistical significance in terms of *p*-values of using different parameters with WLF(RSC). Gray cells indicate significant difference at a 5% (0.05) significance level between the results of the two configurations. RSC parameters are the following: color space (CIELAB or RGB), radius of the center Gaussian, radius of the surround Gaussian, pyramid weight, and the three last parameters are channel weights. "m" indicates the mean.

Parameters	LAB-1-2-1-0.33-0.33-0.33	LAB-1-2-1-0.5-0.25-0.25	LAB-1-2-1-1-0-0	RGB-1-2-4-0.33-0.33-0.33	RGB-2-4-4-m1-m2-m3	RGB-2-3-4-m1-m2-m3	LAB-2-3-4-0.50-0.25-0.25
LAB-1-2-1-0.33-0.33-0.33	-						
LAB-1-2-1-0.5-0.25-0.25	1.00	-					
LAB-1-2-1-1-0-0	0.00	0.00	-				
RGB-1-2-4-0.33-0.33-0.33	0.45	0.45	0.00	-			
RGB-2-4-4-m1-m2-m3	0.00	0.00	0.58	0.00	-		
RGB-2-3-4-m1-m2-m3	0.00	0.00	0.77	0.00	0.22	-	
LAB-2-3-4-0.50-0.25-0.25	0.29	0.29	0.00	0.00	0.00	0.00	-

(a) Original

(b) Relative local contrast map



(c) Saliency weighted local contrast map

Figure 11: The original, the relative local contrast map and saliency weighted local contrast map. Subfigure 11(c) clearly shows how saliency maps discard relevant information used by the observer to judge perceived contrast.

weighted recombination of the average contrast for each level. The final global measure is given by a weighted sum of the contrast of each channel.

In conclusion, after testing the framework, we propose a measure of contrast, based on RSC neighborhood computation, that is called $C^{WLF(RSC)}$ and defined as follows:

$$C^{WLF(RSC)} = \omega_R \cdot C_R^{RSC} + \omega_G \cdot C_G^{RSC} + \omega_B \cdot C_B^{RSC}, \tag{31}$$

where

$$\begin{aligned} C_R^{RSC} &= \tfrac{1}{N_l}\textstyle\sum_l^{N_l} \tau_{R,l} \cdot \bar{c}_{R,l}, \\ C_G^{RSC} &= \tfrac{1}{N_l}\textstyle\sum_l^{N_l} \tau_{G,l} \cdot \bar{c}_{G,l}, \\ C_B^{RSC} &= \tfrac{1}{N_l}\textstyle\sum_l^{N_l} \tau_{B,l} \cdot \bar{c}_{B,l}, \end{aligned} \tag{32}$$

where N_l is the number of levels, $\bar{c}_{R,l}$, $\bar{c}_{G,l}$, and $\bar{c}_{B,l}$ are respectively the mean contrast in the level l for R, G and B channel, $\tau_{R,l}$, $\tau_{G,l}$, and $\tau_{B,l}$ are respectively the variance of pixel values in the level l for R, G, and B channel, ω_R, ω_G, and ω_B are respectively the variance of pixel values for R, G, and B channel used for recombining the overall contrast of each channel C_R^{RSC}, C_G^{RSC} and C_B^{RSC}. A performance of 0.84 in Pearson correlation and 0.80 in Spearman correlation is achieved in relation to observers perceived contrast.

We have tried to improve the WLF using weighting maps, from gaze information, saliency maps, and GAFFE maps. None of them improve the WLF to predict perceived contrast in digital images as relevant information used by the observer to judge perceived contrast are mainly discarded. This suggests that regions-of-interest cannot be used to improve contrast measures as contrast is an intrinsic factor and is judged by a global impression of the image. This indicates that further work on contrast measures should be carried out accounting for the global impression of the image while preserving the local information. Future work can also follow different experimental methodologies to record perceived contrast by the observers: one opportunity is to reperform exactly the same experiment in uncontrolled environments while a second one is to use pairwise comparison to rank the the images.

7. Acknowledgement

The authors would like to thank Alessandro Rizzi (University of Milano, Italy), Ivar Farup (Gjøvik University College, Norway) and Fritz Albregtsen (University of Oslo, Norway) for their interesting discussions and feedback.

References

Adelson, E. H., Anderson, C. H., Bergen, J. R., Burt, P. J., Ogden, J. M., 1984. Pyramid methods in image processing. RCA Engineer 29 (6).

Ahumada, A. J., Beard, B. L., 1998. A simple vision model for inhomogeneous image quality assessment. In: SID International Digest of Technical Papers.

Babcock, J. S., 2000. Eye tracking observers during color image evaluation tasks. Master's thesis, Rochester Institute of Techonology.

Bai, J., Nakaguchi, T., Tsumura, N., Miyake, Y., 2006. Evaluation of image corrected by retinex method based on S-CIELAB and gazing information. IEICE trans. on Fundamentals of Electronics, Communications and Computer Sciences Vol.E89-A (11), 2955–2961.

Boccignone, G., Ferraro, M., Caelli, T., 2001. Encoding visual information using anisotropic transformations. IEEE Trans. Pattern Anal. Mach. Intell 23 (2), 207–211.
URL http://www.computer.org/tpami/tp2001/i0207abs.htm

Büring, H., Herzog, P. G., Jung, E., 2004. Evaluation of current color management tools: Image quality assessments. In: 2nd European Conference on Color in Graphics, Imaging and Vision (CGIV). Aachen, Germany, pp. 459–462.

Burkhardt, D. A., Gottesman, J., Kersten, D., Legge, G. E., 1984. Symmetry and constancy in the perception of negative and positive luminance contrast. J. Opt. Soc. Am. A 1 (3), 309.
URL http://josaa.osa.org/abstract.cfm?URI=josaa-1-3-309

Calabria, A. J., Fairchild, M. D., November/December 2003. Perceived image contrast and observer preference II. empirical modeling of perceived image contrast and observer preference data. The Journal of imaging science and technology 47, 494–508.

CIE, 156:2004. Guidelines for the evaluation of gamut mapping algorithms. Tech. Rep. ISBN: 3-901-906-26-6, CIE TC8-08.

CIE, 2004. Colorimetry. Tech. Rep. 15.

de Ridder, H., February 1992. Minkowski-metric as a combination rule for digital-image-coding impariments. In: Rogowitz, B. (Ed.), Human Vision, Visual Processing, and Digital Display III. Vol. 1666. SPIE/ IS&T, San Jose, CA, pp. 16–26.

Engeldrum, P. G., 2000. Psychometric Scaling. Imcotek Press, Winchester, MA, USA.

Ferraro, M., Boccignone, G., 2004. Image contrast enhancement via entropy production. Real-Time Imaging 10 (4), 229–238.

Frankle, J., McCann, J., 1983. Method and apparatus for lightness imaging United States Patent No. 4,384,336.

Henderson, J. M., Williams, C. C., Castelhano, M. S., Falk, R. J., 2003. Eye movements and picture processing during recognition. Perception & Psychophysics 65, 725–734.

Hess, R. F., Bradley, A., Piotrowski, L., 1983. Contrast-coding in amblyopia. i. differences in the neural basis of human amblyopia. Proc. R. Soc. London Ser. B (217), 309–330.

Holm, J., Tastl, I., Johnson, T., 2006. Definition & use of the iso 12640-3 reference color gamut. In: Fourteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications. IS&T/SID, Scottsdale, AZ, pp. 62–68.

Kendall, M. G., Stuart, A., Ord, J. K., 1991. Kendall's Advanced Theory of Statistics: Classical inference and relationship, 5th Edition. Vol. 2. A Hodder Arnold Publication.

King-Smith, P. E., Kulikowski, J. J., August 1975. Pattern and flicker detection analysed by subthreshold summation. J Physiol 249 (3), 519–548.
URL http://view.ncbi.nlm.nih.gov/pubmed/1177103

Michelson, A., 1927. Studies in Optics. University of Chicago Press.

Pedersen, M., 2007. Importance of region-of-interest on image difference metrics. Master thesis, Gjøvik University College.
URL http://www.hig.no/content/download/9052/122117/file/Pedersen%20-%20Import

Pedersen, M., Hardeberg, J. Y., June 2008. Rank order and image difference metrics.

Peli, E., 1990. Contrast in complex images. Journal of the Optical Society of America 7, 2032–2040.

Ponomarenko, N., Lukin, V., Egiazarian, K., Astola, J., Carli, M., Battisti, F., October 2008. Color image database for evaluation of image quality metrics, 403–408http://www.ponomarenko.info/tid2008.htm.

Rajashekar, U., van der Linde, I., Bovik, A. C., Cormack, L. K., apr 2008. Gaffe: A gaze-attentive fixation finding engine. In: IEEE transactions on image processing. Vol. 17. pp. 564–573.
URL http://live.ece.utexas.edu/research/gaffe

Rizzi, A., Algeri, T., Medeghini, G., Marini, D., April 2004. A proposal for contrast measure in digital images. In: CGIV 2004 - Second European Conference on Color in Graphics, Imaging and Vision. Aachen, Germany.

Rizzi, A., Simone, G., Cordone, R., June 2008. A modified algorithm for perceived contrast in digital images. In: CGIV 2008 - Fourth European Conference on Color in Graphics, Imaging and Vision. IS&T, Terrassa, Spain, pp. 249–252.

Simone, G., Pedersen, M., Hardeberg, J. Y., Farup, I., June 2009a. On the use of gaze information and saliency maps for measuring perceptual contrast. In: in Computer Science, L. N. (Ed.), 16th Scandinavian Conference on Image Analysis. Vol. 5575. Oslo, Norway, pp. 597–606,.

Simone, G., Pedersen, M., Hardeberg, J. Y., Rizzi, A., January 2009b. Measuring perceptual contrast in a multilevel framework. In: Rogowitz, B., Pappas, T. (Eds.), Human Vision and Electronic Imaging XIV. Vol. 7240. SPIE, pp. 72400Q–72400Q–9.

Tadmor, Y., Tolhurst, D., 2000. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. Vision Research 40, 3145–3157.

Tremeau, A., Apr. 2000. Color contrast parameters for analysing image differences. In: Colour, Institute, I. (Eds.), Proc. of the Univ. of Derby Color Image Science 2000 Conference. Derby, pp. 11–23.

Walther, D., Koch, C., 2006. Modeling attention to salient proto-objects. Neural Networks 19, 1395–1407.

Whittle, P., 1986. Increments and decrements: luminance discrimination. Vision Research (26), 1677–1691.

10.2 Paper B: Measuring perceptual contrast in uncontrolled environments

MEASURING PERCEPTUAL CONTRAST IN UNCONTROLLED ENVIRONMENTS

Gabriele Simone, Marius Pedersen and Jon Yngve Hardeberg.

Gjøvik University College, Gjøvik, Norway.

ABSTRACT

Contrast is one of the most relevant perceptual and quality factors in digital images and measuring it is not a trivial task. We have carried out an online psychophysical experiment to register perceived contrast. The results from the observers indicate that color images are rated higher than their respective greyscale one indicating that contrast is influenced color. A two-sided sign test at 5% level confirms this hypothesis.

A comparison with a lab controlled environment experiment has been carried out in order to investigate possible differences. A statistical analysis of the two experiments indicate that the mean ratings of the observers are not significantly different. A decrease in correlation of previously developed contrast measures can be noticed and a comparison of the correlation coefficients indicate that measuring contrast in uncontrolled environments can be significantly different than measuring in a lab controlled environment.

Index Terms— Perceived Contrast, Psychophysical Experiment, Uncontrolled Environments, Contrast Measures.

1. INTRODUCTION

Since the first studies on contrast in images, it has emerged how hard it could be to give a definition of perceptual contrast and, moreover, how subjective and related to the observation task or observer experience this definition could turn out to be.

A possible definition of contrast is the difference between the light and dark parts of a photograph, where less contrast gives a flatter picture, and more a deeper picture. Many other definitions of contrast are also given, such as the difference in visual properties that makes an object distinguishable or just the difference in color from point to point. As various definitions of contrast are given, measuring contrast is very difficult. Measuring the difference between the darkest and lightest point in an image does not predict perceived contrast since perceived contrast is influenced by the surround and the spatial arrangement of the image. Parameters such as resolution, viewing distance, lighting conditions, image content, memory color might affect how observers perceive contrast.

For this reason, the first approaches to this topic have confined themselves to study the phenomenon at a global level, operating in controlled situations and under same constraints, the so-called "void conditions". After these first experiments more complex measures have been devised, but a universal measure of contrast in images is still not clearly defined.

Recent studies has shown that contrast is one of the most relevant perceptual and quality factors in printing [1]. Zuffi et al. [2, 3] showed that evaluating quality of prints in controlled and uncontrolled environment is not significantly different. For this reason, in this paper we perform an evaluation of contrast measures in relation to observers perceived contrast in uncontrolled environments.

The rest of this paper will be organized as follows: in section 2 we will provide an insight into the state-of-the-art of image contrast measures. Section 3, describes the psychophysical experiment, and section 4 presents the experimental results and discusses how the results from contrast measures reflects their perceptual contrast estimation performance. Finally in section 5 conclusions are drawn.

2. STATE OF THE ART

The very first measure of global contrast, in the case of sinusoids or other periodic patterns of symmetrical deviations ranging from the maximum luminance L_{max} to minimum luminance L_{min} is the Michelson [4] formula:

$$C^M = (L_{max} - L_{min})/(L_{max} + L_{min}).$$

King-Smith and Kulikowski [5], and Burkhardt [6] and Whittle [7] follow a similar concept replacing L_{max} or L_{min} with L_{avg}, which is the mean luminance of the image.

In 1990 Peli [8] proposed a local contrast measure based on a multilevel structure of band-pass filters, the so called pyramid, which better estimates apparent perceived contrast [9]. The multilevel approach has been proved to be an important implementation feature to mimic HVS [10, 11] and during the decades it has been model for other local contrast measures as: Winkler and Vandergheynst [12], who define an isotropic measure using directional wavelets; Iordache et al. [13] in cmobination with the Moon and Spencer model [14]; Dauphin et al. [15] taking into account frequency and directional selectivity of the subbands.

In 2004 Rizzi et al. [16] developed a very simple and efficient measure, able to estimate global and local components of contrast. In details, this measure performs a pyramid subsampling of the image to various levels in the CIELAB color

space. Then a calculation of the local contrast in each pixel is carried out by taking the average of the absolute value difference between the lightness channel value of the pixel and the surrounding eight pixels. A contrast map is thereby produced for each level separately. The final overall measure is a recombination of the average contrast for each level:

$$C^{RAMMG} = \frac{1}{N_l}\sum_{l=1}^{N_l}\bar{c}_l, \quad (1)$$

where N_l is the number of levels and $\bar{c}_l$ is the mean contrast in the level *l*:

$$\bar{c}_l = \sum_{i=1}^{i_{max}}\sum_{j=1}^{j_{max}} c_{i,j}, \quad (2)$$

where i_{max} and j_{max} indicate respectively the height and the width of the image and $c_{i,j}$ is the contrast of each pixel.

Following a similar approach as introduced in RAMMG, Rizzi et al. proposed in 2008 the RSC algorithm [17]. The RSC works with the same pyramid subsampling as RAMMG but it computes for each pixel of each channel the Difference of Gaussians (DOG) contrast calculation, and this computation is performed separately for the lightness and the chromatic channels instead of only for lightness. The three measures are then combined with different weights. The final overall measure is expressed as:

$$C^{RSC} = \alpha \cdot C^{RSC}_{L^*} + \beta \cdot C^{RSC}_{a^*} + \gamma \cdot C^{RSC}_{b^*}, \quad (3)$$

where α, β, and γ represent the weights of each channel.

The DOG model as local contrast measure was modified and adapted to natural images by Tadmor and Tolhurst in 2000 [18]. In the conventional model, the spatial sensitivity in the center of receptive-fields is described by a bi-dimensional Gaussian with unit amplitude:

$$Centre(x,y) = exp[-(x/r_c)^2 - (y/r_c)^2],$$

where the radius r_c represents the distance at which the sensitivity decreases following $1/e$ with respect to the peak level and (x,y) are spatial coordinates where the receptive-field is placed. The surround component is represented by another Gaussian curve, with a larger radius, r_s:

$$Surround(x,y) = 0.85(r_c/r_s)^2 exp[-(x/r_s)^2 - (y/r_s)^2].$$

When the central point of the receptive-field is placed at the location (*x,y*), the output of central component is calculated as:

$$R_c(x,y) = \sum_i \sum_j Centre(i-x, j-y)I(i,j),$$

while the output of the surround component is:

$$R_s(x,y) = \sum_i \sum_j Surround(i-x, j-y)I(i,j),$$

where in both cases *I(i,j)* is the image pixel value at position (*i,j*). The result of the DOG model is obtained as:

$$DOG(x,y) = R_c(x,y)/R_s(x,y).$$

The conventional DOG model assumes that the response of a neuron depends uniquely on the local luminance difference (ΔI) between the center and the surround. After the light adaptation process, the gain of the ganglion cells of the retina and the Lateral Geniculate Nucleus (LGN) neurons depends on the average local luminance I. Thus the model response depends on the contrast stimulus. They propose the following criterium for the measure of contrast:

$$c^{TT}(x,y) = (R_c(x,y) - R_s(x,y)) \,/\, (R_c(x,y) + R_s(x,y))$$

Simone et al. [19] developed shortly afterwards a weighted level framework (WLF) as an evolution of the previous contrast measures from Rizzi et al. The main improvements are the use of antialiasing filtering in the pyramid construction combined with a weighted recombination of the local contrast maps:

$$C_i = \frac{1}{N_l}\sum_{l=1}^{N_l}\lambda_l \cdot \bar{c}_l, \quad (4)$$

where N_l is the number of levels, $\bar{c}_l$ is the mean contrast in the level *l*, λ_l is the weight assigned to each level *l* and *i* indicates the applied channel. The overall final measure is given by equation:

$$C^{WLF} = \alpha \cdot C_1 + \beta \cdot C_2 + \gamma \cdot C_3, \quad (5)$$

where α, β, and γ are the weights of each color channel. The measure can be extended to different color spaces and it is not limited to CIELAB as the RAMMG and RSC. Having the weighting parameters λ_l, α, β, and γ assuming values from measures taken from the image itself makes the WLF contrast measure independent by many custom parameters. A psychophysical experiment clearly shows an improvement of the contrast measures when the variance of image pixels is used as parameter for λ [19].

Alternative approaches have been also developed as Tremeau [20] based on a region adjaceny graph and Boccignone et al. [21] and Ferraro et al. [22], where the image can be considered as an isolated thermodynamical system.

3. PSYCHOPHYSICAL EXPERIMENT

A website was developed with the purpose of analyzing and comparing observers perceived contrast in uncontrolled environments against a lab controlled experiment [23] in order to see if contrast can be considered a feature of the image independent of the environment. The test set is composed of 41 images divided in the following categories:

- 15 natural color images presented and analyzed by Pedersen et al. [23],
- the same previous 15 images presented in greyscale,
- 11 artificial images, divided in five artistic and six computer graphic.

Before allowing a user to commence the experiment, a check of a minimum screen resolution of 1024×768 was performed. Afterwards the following instructions, also translated in other three languages, were provided:

Your task is to rate one or several images according to your perceived sensation of contrast. No definition of contrast is given, and you are encouraged to use your own idea of contrast. The experiment is the following, you rate the image by the perceived contrast, from 1 to 10.

The images were shown randomly on a neutral grey background and in order to avoid "faulty" answers, the observer had the option to skip images considered too hard to judge. No limitations were given on the number of images to rate, i.e. an observer could stop whenever he wanted. Participants were recruited from all over the world among friends, colleagues and worldwide universities. A total of more than 3000 ratings were recorded, resulting in an average of 170 ratings for each image.

4. RESULTS

Figure 1 shows all the images used in the online experiment sorted from lowest (up-left) to highest (down-right) contrast according to the average rating given by the observers. Figure 2 shows the mean perceived contrast score and the 95% confidence interval for all images divided in 10 categories. The scores range from "category 4" to "category 9", judging all the images from medium-low to very high contrast. It is interesting to see that all the artistic articial images are ranked between "category 4" and "category 6", which might be due by the absence of strong edges. Artificial computer graphics and greyscale images ranks are spread in all the categories used by the observers.

We have focused our analysis in particular on the 15 natural color images. All the images have a mean rank from "category 6" to "category 9", indicating to have high contrast by observer judgement. Besides, none of them has been rated to have lower contrast than their respective greyscale version indicating that contrast is influenced color. A two-sided sign test at 5% level confirms that rating color images is significantly different of rating greyscale images.

Following we have made a comparison with the lab controlled experiment performed by Pedersen et al. [23]. First we can notice that in the lab experiment the observers used only a part of the scale while in the online the observers used all the scale with a tendancy of using high scores. Figure 3 shows that 12 of 15 images had a higher mean value, but with a 95% confidence interval they are not significantly different. Only image 21, 23, and 37 had higher mean value in lab controlled environment than in the online, but do not differentiate with 95% confidence interval. In the uncontrolled experiment, the ratings are spread as in the controlled and with 95% confidence interval we cannot claim that uncontrolled contrast rating is significantly different. In lab controlled environment the confidence intervals are wider due to the reduced number of observers.

As contrast measure we have used the weighted level framework (WLF) developed by Simone et al. [19] briefly described in Section 2. Table 1 shows the Pearson correlation for controlled and uncontrolled environment for different settings of the framework. We can notice a decrease in correlation for the online experiment due to the difference of the mean values of the observers in the two experiments (Figure 3).

A comparison of the correlation coefficients has also been performed as proposed by Gibert et al. [24]. Measuring contrast in uncontrolled environments can be significantly different at 99% than measuring in lab controlled environment. Furthermore analysis of the Spearman correlation is similar.

5. CONCLUSION

An online psychophysical experiment to register perceived contrast has been carried out. An average of 170 ratings for 41 image were recorded. The results from the observers indicate that color images are rated higher than their respective greyscale one indicating that contrast is influenced color. A two-sided sign test at 5% level confirms this hypothesis.

A comparison with a lab controlled environment experiment has been carried out. Most of the images had higher mean score in the online experiment but with 95% confidence interval they are not significantly different. However, this results in a decrease in correlation of previously developed contrast measures can be noticed and a comparison of the correlation coefficients indicate that measuring contrast in uncontrolled environments can be significantly different at 99% than measuring in lab controlled environment.

A future work could be performing an evaluation of contrast measures and a comparison of the online in relation to observers perceived contrast in uncontrolled environments of the same dataset of printed images.

6. ACKNOWLEDGEMENT

The authors would like to thank Alessandro Rizzi (University of Milano, Italy) and Ivar Farup (Gjøvik University College, Norway) for their interesting discussions and feedback.

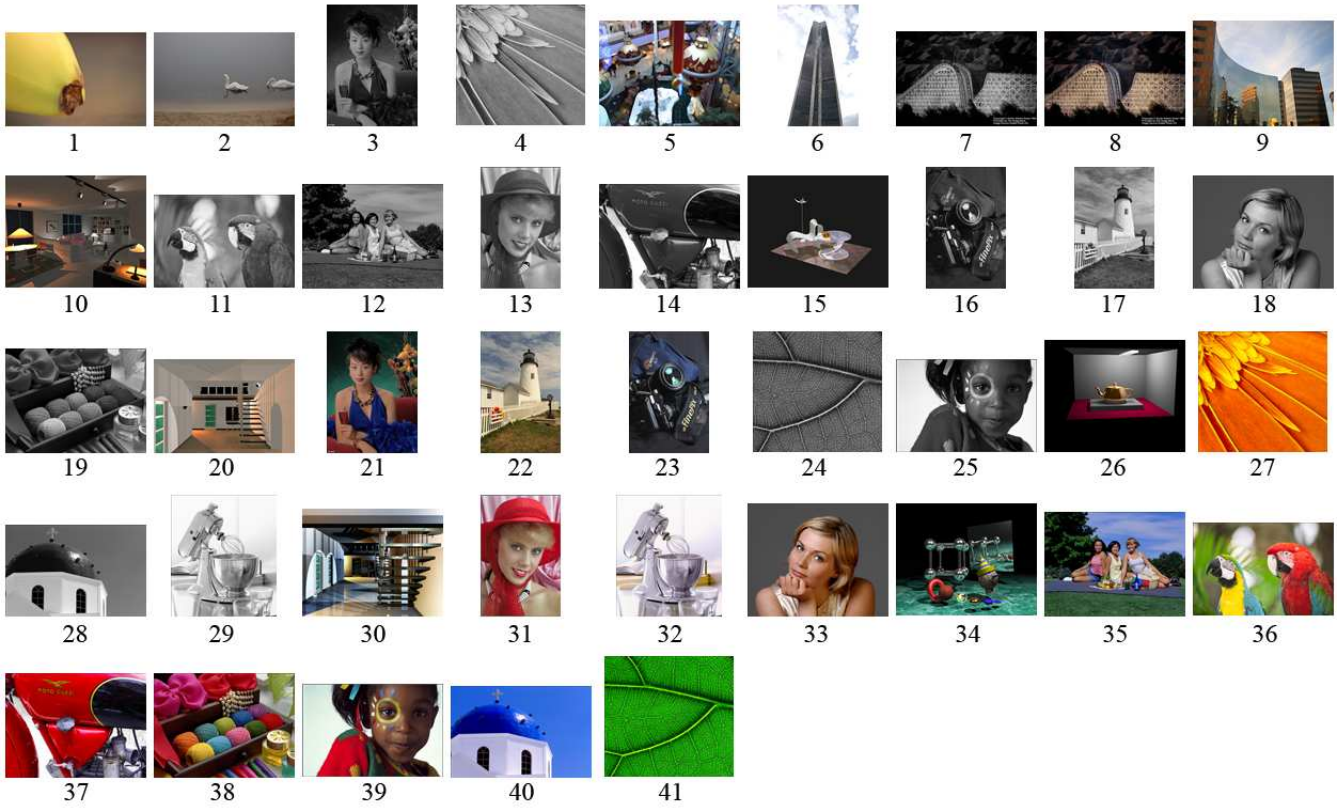


Fig. 1. Images used in the experiment sorted from lowest (up-left) to highest (down-right) contrast according to the average rating given by the observers.

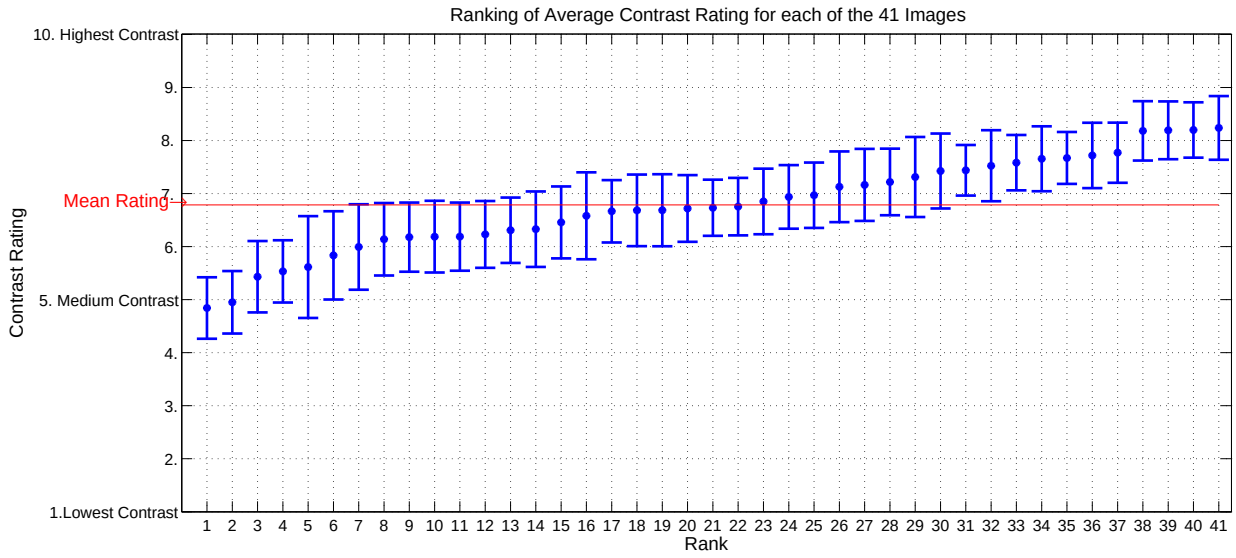


Fig. 2. Mean perceived contrast score and the standard deviation for all images divided in 10 categories. The mean scores range from ”category 4” to ”category 9”.

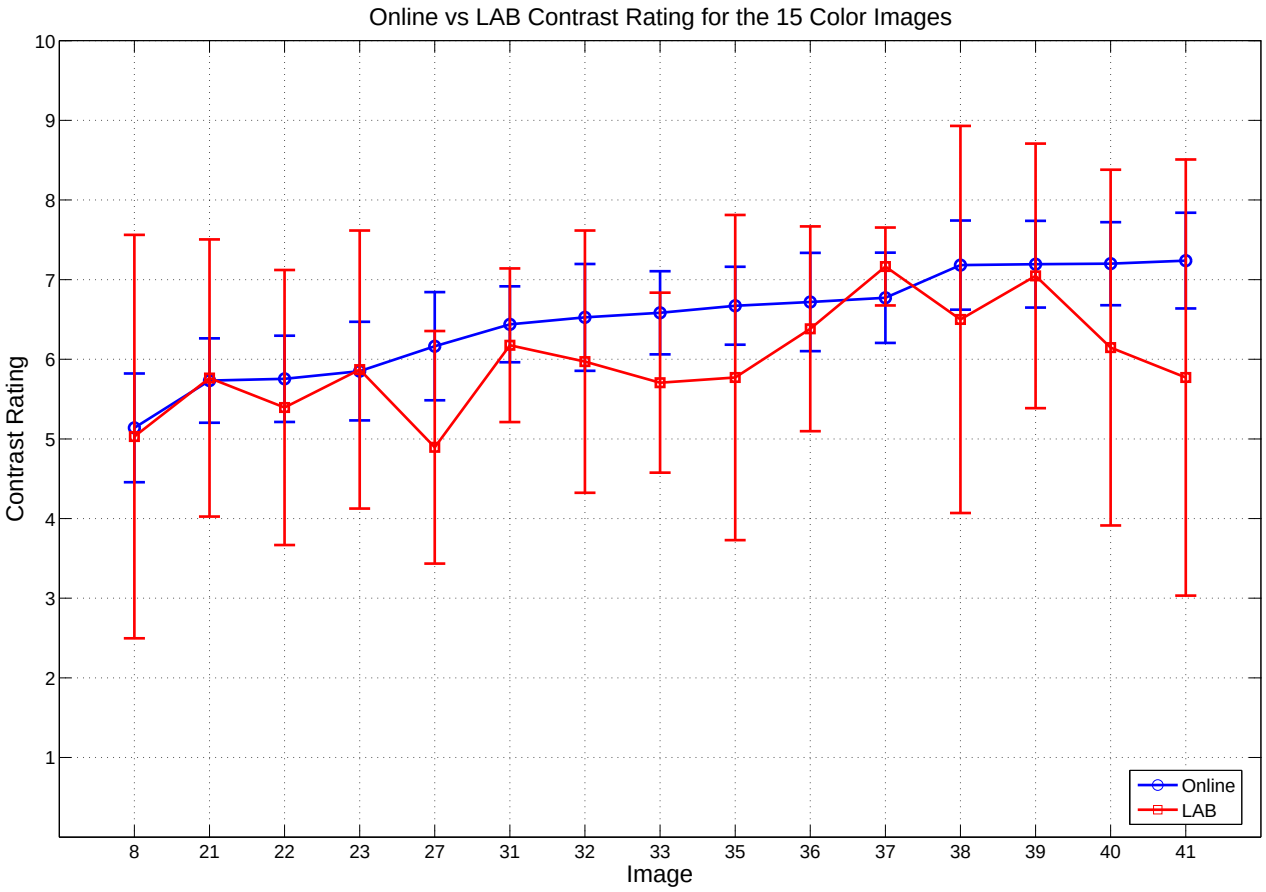


Fig. 3. Observer score in lab controlled and uncotrolled environment. 12 of 15 images are rated higher in an uncontrolled environments. Only image 21, 23, and 37 were rated higher in lab controlled environment than in the online. However, with a 95% confidence interval uncontrolled contrast rating is not significantly different than lab controlled rating for any of the images.

Table 1. Pearson correlation for controlled and uncontrolled environments for different settings of the framework. Yellow cells indicate the standard values while grey cells indicate the highest Pearson correlation. All numbers are round to three decimals. λ equal to τ indicates that each level of the pyramid is weighted with the variance of the three channels separately. $\frac{1}{v}$ indicates that the considered channel is weighted with the reciprocal of its mean while ω indicates that is weighted with its variance. A substantial decrease in correlation can be noticed in uncontrolled environments.

C^{WLF}											
							UNCONTROLLED		CONTROLLED		
ColorSpace	r_c	r_s	λ	α	β	γ	Pearson correlation	Pearson p-value	Pearson correlation	Pearson p-value	Comparison
CIELAB	1	2	1	0.333	0.333	0.333	0.331	0.228	-0.226	0.418	Significantly different at confidence level <50%
CIELAB	1	2	1	0.5	0.25	0.25	0.351	0.199	-0.159	0.571	Significantly different at confidence level <50%
CIELAB	1	2	1	1	0	0	0.362	0.184	0.016	0.954	Significantly different at confidence level 65%
RGB	1	2	τ	0.333	0.333	0.333	0.534	0.04	0.818	0.000	Significantly different at confidence level 99%
RGB	2	4	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.441	0.099	0.819	0.000	Significantly different ta confidence level 99%
CIELAB	3	4	τ	0.5	0.25	0.25	0.489	0.064	0.822	0.000	Significantly different at confidence level 99%
RGB	2	3	τ	$\frac{1}{v}$	$\frac{1}{v}$	$\frac{1}{v}$	0.463	0.082	0.826	0.000	Significantly different ta confidence level 99%
CIELAB	2	3	τ	0.5	0.25	0.25	0.516	0.049	0.829	0.000	Significantly different at confidence level 99%
CIELAB	2	4	τ	0.5	0.25	0.25	0.563	0.029	0.830	0.000	Significantly different at confidence level 99%
RGB	2	4	τ	ω	ω	ω	0.506	0.054	0.844	0.000	Significantly different at confidence level 99%
RGB	2	4	τ	0.333	0.333	0.333	0.508	0.053	0.845	0.000	Significantly different at confidence level 99%

7. REFERENCES

[1] M. Pedersen, N. Bonnier, J. Y. Hardeberg, and F. Albregtsen, "Attributes of image quality for color prints," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011016–1 – 011016–13, Jan 2010.

[2] S. Zuffi, E. Beltrame, and P. Scala, "Visual experiments on the web: design of a web-based visual experiment management system," in *Image Quality and System Performance V*, San Jose, CA, USA, 2008, vol. 6808, p. 68080O, SPIE.

[3] S. Zuffi, C. Brambilla, R. Eschbach, and A. Rizzi, "Controlled and uncontrolled viewing conditions in the evaluation of prints," in *Color Imaging XIII: Processing, Hardcopy, and Applications*, San Jose, CA, USA, 2008, vol. 6807, p. 680714, SPIE.

[4] A. Michelson, *Studies in Optics*, University of Chicago Press, 1927.

[5] P. E. King-Smith and J. J. Kulikowski, "Pattern and flicker detection analysed by subthreshold summation.," *J Physiol*, vol. 249, no. 3, pp. 519–548, August 1975.

[6] D. A. Burkhardt, J. Gottesman, D. Kersten, and G. E. Legge, "Symmetry and constancy in the perception of negative and positive luminance contrast," *J. Opt. Soc. Am. A*, vol. 1, no. 3, pp. 309, 1984.

[7] P. Whittle, "Increments and decrements: luminance discrimination," *Vision Research*, , no. 26, pp. 1677–1691, 1986.

[8] E. Peli, "Contrast in complex images," *J. Opt. Soc. Am. A*, vol. 7, pp. 2032–2040, 1990.

[9] E. Peli, "In search of a contrast metric: Matching the perceived contrast of gabor patches at different phases and bandwidths," *Vision Research*, vol. 37, no. 23, pp. 3217 – 3224, 1997.

[10] J. Frankle and J. McCann, "Method and apparatus for lightness imaging," vol. United States Patent No. 4,384,336, 1983.

[11] E. H. Adelson, C. H. Anderson, J. R. Bergen, Peter J. Burt, and J. M. Ogden, "Pyramid methods in image processing," *RCA Engineer*, vol. 29, no. 6, 1984.

[12] S. Winkler and P. Vandergheynst, "Computing isotropic local contrast from oriented pyramid decompositions," in *in Proc. ICIP*, Kobe, Japan, Oct. 25–28, 1999, vol. 4, pp. 420–424.

[13] R. Iordache, A. Beghdadi, and P. V. de Lesegno, "Pyramidal perceptual filtering using moon and spencer contrast," in *International Conference on Image Processing*, Thessaloniki, Greece, Oct 2001, vol. 3, pp. 146–149, IEEE.

[14] P. Moon and D. E. Spencer, "The visual effect of non-uniform surrounds," *J. Opt. Soc. Am.*, vol. 35, no. 3, pp. 233–247, 1945.

[15] G. Dauphin, A. Beghdadi, and P.V. de Lesegno, "A local directional bandlimited contrast," in *Seventh International Symposium on Signal Processing and Its Applications*, Paris, France, Jul 2003, vol. 2, pp. 197 – 200.

[16] A. Rizzi, T. Algeri, G. Medeghini, and D. Marini, "A proposal for contrast measure in digital images," in *CGIV 2004 - Second European Conference on Color in Graphics, Imaging and Vision*, Aachen, Germany, April 2004.

[17] A. Rizzi, G. Simone, and R. Cordone, "A modified algorithm for perceived contrast in digital images," in *CGIV 2008 - Fourth European Conference on Color in Graphics, Imaging and Vision*, Terrassa, Spain, Jun 2008, pp. 249–252, IS&T.

[18] Y. Tadmor and D.J. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes," *Vision Research*, vol. 40, pp. 3145–3157, 2000.

[19] G. Simone, M. Pedersen, J. Y. Hardeberg, and A. Rizzi, "Measuring perceptual contrast in a multilevel framework," in *Human Vision and Electronic Imaging XIV*. Jan 2009, vol. 7240, SPIE.

[20] A. Tremeau, "Color contrast parameters for analysing image differences," in *Proc. of the Univ. of Derby Color Image Science 2000 Conference*, Colour and I. Institute, Eds., Derby, Apr. 2000, pp. 11–23.

[21] G. Boccignone, M. Ferraro, and T. Caelli, "Encoding visual information using anisotropic transformations," *IEEE Trans. Pattern Anal. Mach. Intell*, vol. 23, no. 2, pp. 207–211, 2001.

[22] M. Ferraro and G. Boccignone, "Image contrast enhancement via entropy production," *Real-Time Imaging*, vol. 10, no. 4, pp. 229–238, 2004.

[23] M. Pedersen, J. Y. Hardeberg, and P. Nussbaum, "Using gaze information to improve image difference metrics," in *Human Vision and Electronic Imaging VIII*, San Jose, USA, Jan 2008, vol. 6806, pp. 680611–1–680611–12, SPIE.

[24] J. M. Gibert, J. M. Dagà, E. J. Gilabert, and J. Valldeperas, "Evaluation of colour difference formulae," *Coloration Technology*, vol. 121, no. 3, pp. 147–152, 2005.

10.3 Paper C: Multi–level contrast filtering in image difference metrics

EURASIP Journal on Image and Video Processing
a SpringerOpen Journal

RESEARCH Open Access

Multi-level contrast filtering in image difference metrics

Gabriele Simone[1*], Marius Pedersen[1], Ivar Farup[1] and Claudio Oleari[2]

Abstract

In this paper, we present a new metric to estimate the perceived difference in contrast between an original image and a reproduction. This metric, named weighted-level framework ΔE_E (WLF-DEE), implements a multilevel filtering based on the difference of Gaussians model proposed by Tadmor and Tolhurst (2000) and the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. (2009). Extensive tests and analysis are presented on four different categories belonging to the well-known Tampere Image Database and on two databases developed at our institution, providing different distortions directly related to color and contrast. Comparisons in performance with other state-of-the-art metrics are also pointed out. Results promote WLF-DEE as a new stable metric for estimating the perceived magnitude of contrast between an original and a reproduction.

1 Introduction

The quality of image reproduction depends on many factors that cannot be completely distinguished since they are mutually dependent. It is generally considered that the quality of color in an image reproduction may be quantified by measuring how accurate the reproduction of color is to the original when viewed by a human observer. Color discrimination is determined by many factors, including the spatial pattern of the image and the visual processing, which starts with cone activation and is followed by adaptation to the illumination. Images are constituted by spatial color patterns, which are different from the uniform color patches considered in the colorant factories. Color discrimination and appearance in images are a function of spatial pattern, and color differences are harder to detect at higher frequencies, where contrast plays an important role. Moreover, color opponencies have a role in color discrimination, especially differences along the blue-yellow color direction [1]. As a consequence, for measuring the perceptual difference between an original image and its reproduction, a perceptual image difference metric is needed. Generally, this metric is obtained by extending the color difference formulae to the color of a complex image. The history of the color metrics is more than one century long.

Over the years, many image difference metrics have been proposed [2], some for measuring general image quality and image difference and some for detecting specific distortions. However, at the moment, no universal image difference metric exists. Image difference metrics are based on a number of different ideas but are usually following a general framework. The image and its reproduction are first transformed into a suitable color space, preferably a perceptually uniform one. Then, a simulation of the human visual system (HVS) is carried out, from simplistic methods as smoothing of the image by a local neighborhood to more complex methods, e.g. using contrast sensitivity functions (CSFs). Finally, the difference is calculated usually by a color difference formula. Thus, an efficient image difference metric needs a filter which is able to mimic the HVS and a suitable color difference formula. This work will mainly focus on the filtering side, particularly taking into account that contrast is an important image attribute playing an important role in image quality discrimination [3]. However, recent improvements of color difference formula and image processing methods will be considered for the development of our metrics.

After introducing the improvements of color difference formulae in Section 2 and image difference metrics in Section 3, we present our proposal in Section 4. We propose a new image difference metrics based on multilevel contrast filtering using the difference of Gaussians

*Correspondence: gabriele.simone@hig.no
[1]The Norwegian Colour and Visual Computing Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, Gjøvik 2802, Norway
Full list of author information is available at the end of the article

(DOG) model proposed by Tadmor and Tolhurst [4]. This metric will use also the new Euclidean color difference formula in log-compressed OSA-UCS color space proposed by Oleari et al. [5] instead of the traditional ΔE^*_{ab} that most of the state-of-the-art metrics generally use. Section 5 presents the description of three state-of-the-art databases used to evaluate the performance of the new metric with the experimental results and discusses how our proposal reflects perceptual image difference estimation performance. Finally, in Section 6 conclusions are drawn.

2 Color difference metrics

The first studies on color difference metrics date back to the end of the nineteenth century [6]. In color metrics history, a milestone is represented by the CIELAB color system [7,8] proposed by the International Commission on Illumination (CIE) in 1976 as a uniform color space, in which the color difference between two colors is represented by the Euclidean distance ΔE^*_{ab}.

Since 1976, the main effort was to improve the CIELAB color difference formula, and almost all the proposals are based on the CIELAB space. Then, the structure of such a space has strongly conditioned the shape and the quality of almost all other proposals. The empirical color difference data used for evaluating and improving the color difference formulae were the supra-threshold Bradford University (BFD) ellipses [9], here represented on the $a^*_{10}b^*_{10}$ coordinates of the CIELAB space (Figure 1).

In 1984, the British 'Colour-Measurement Committee' (CMC) of the 'Society of Dyers and Colourist - UK' recommended a color difference formula, ΔE_{CMC}, that has been integrated into some ISO standards [10]. The ΔE_{CMC} formula introduces a local metric tensor in the CIELAB space with the consequence that the color tolerances in the CIELAB space are represented by ellipsoids with semi-axis lengths depending on the point in the space and with one axis oriented as the lightness, one as the chroma and one as the hue, i.e. as a function of the differences of ΔL^*, ΔC^* and ΔH^*. Moreover, the weighting factors of the formula are hue-dependent. The CMC formula has acceptance in industrial color control applications.

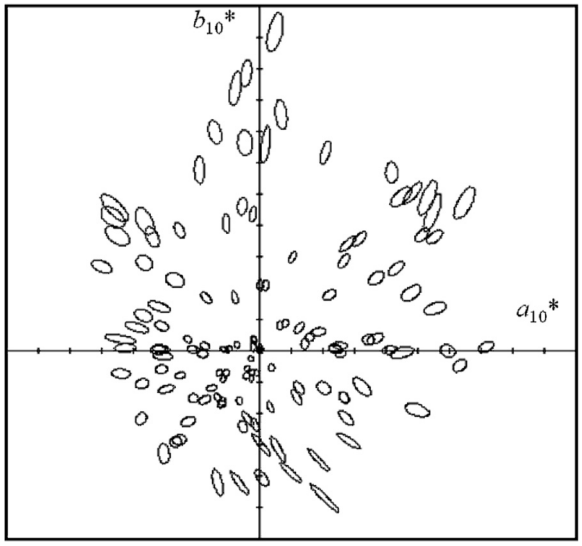


Figure 1 BFD ellipses represented on the $a^*_{10}b^*_{10}$. The ellipses show only approximate regularity, with exclusion of the blue-magenta hues, where the ellipses are distorted. The CIEDE2000 formula is particularly complicated for regularizing such a distortion.

In 1987, Luo and Rigg [11,12] proposed the BFD color difference formula providing a correction of the CMC formula in the blue region. In 1994, CIE proposed the non-Euclidean formula ΔE_{94} [13] by introducing a local metric tensor in the CIELAB space in analogy with the CMC formula. In 2000, CIE proposed its last color difference formula, named CIEDE2000 and denoted by ΔE_{00} [14], which is based on an enlarged dataset of empirical color differences, known as the COM dataset [14], which is added to the original BFD color difference data. This formula is only apparently based on the CIELAB space because a coordinate transformation is made in order to correct hue distortions typical of this space. The CIEDE2000 formula represents distorted ellipsoids. Although enlarged, the empirical dataset shows an evident noise, and the data related to different laboratories are not completely in agreement. The quality of these datasets combined with the large number of parameters used in their fitting induces us to suppose a risk of over-fitting.

These color difference formulae have been also applied to natural images, and evaluation of some of them for the measurement of color image reproduction quality can be found in [15-18].

In 2009, a color difference formula for small-medium color differences not based on CIELAB was published. This formula, termed ΔE_E, is Euclidean [5]. The empirical color difference data plotted in the OSA-UCS space show a higher regularity than what has been seen before (Figures 2 and 3), inducing the authors to state a hue-independence hypothesis. On the basis on this hypothesis, the OSA-UCS space has been chroma log-compressed, and in this new space, the color difference formula assumes a Euclidean shape.

Since this formula is used in this work, let us recall the formula completely from Oleari et al. [5]. The transformation from the tristimulus space (X_{10}, Y_{10}, Z_{10}), where Y_{10} is the percentage luminance factor, and the OSA-UCS space (L_{OSA}, G, J) is the following: the lightness is the same as defined by the OSA-UCS committee:

$$L_{\mathrm{OSA}} \equiv 5.9 \left\{ \left[Y_0^{1/3} - \frac{2}{3} + 0.042(Y_0 - 30)^{1/3} \right] - 14.4 \right\} \frac{1}{\sqrt{2}}, \tag{1}$$

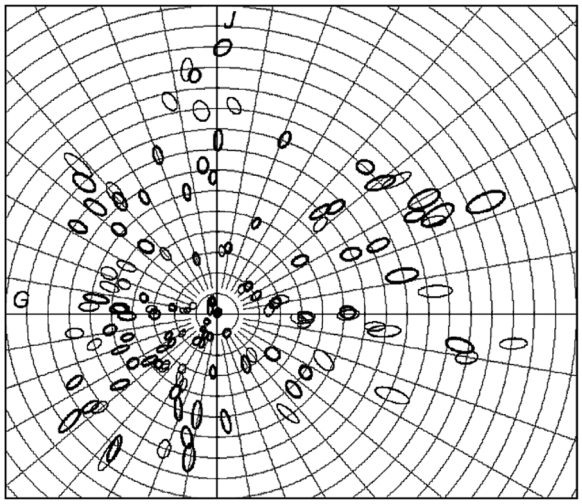


Figure 2 BFD ellipses on the coordinates *JG* of the OSA-UCS space. The ellipses show a general regularity that suggests the hue-independence hypothesis and consequently a simple ellipsoidal color difference formula [5,21], which is based on all the existing data (COM dataset [14]). The next step towards the Euclidean formula [21] is the logarithmic compression of the chroma (Figure 3).

with

$$Y_0 \equiv Y_{10}(4.4934x_{10}^2 + 4.3034y_{10}^2 - 4.276x_{10}y_{10} - 1.3744x_{10} - 2.5643y_{10} + 1.8103), \quad (2)$$

the tristimulus values in the main reference frame [19,20] are

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} 0.6597 & 0.4492 & -0.1089 \\ -0.3053 & 1.2126 & 0.0927 \\ -0.0374 & 0.4795 & 0.5579 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (3)$$

(this is an active transformation that takes into account the adaptation to the visual situation used to define the OSA-UCS system), and the OSA-UCS coordinates (g, j) obtained by computation and denoted by (G, J) are

$$\begin{pmatrix} J \\ G \end{pmatrix} = \begin{bmatrix} 2(0.5735L_{\mathrm{OSA}} + 7.0892) & 0 \\ 0 & -2(0.7640L_{\mathrm{OSA}} + 9.2521) \end{bmatrix} \begin{bmatrix} 0.1792 & 0.9837 \\ 0.9482 & -0.3175 \end{bmatrix} \begin{pmatrix} \ln\left(\frac{A/B}{0.9366}\right) \\ \ln\left(\frac{B/C}{0.9807}\right) \end{pmatrix} \quad (4)$$

$$h = \arctan\left(-\frac{J}{G}\right) \text{ and the chroma } C_{\mathrm{OSA}} = \sqrt{G^2 + J^2}. \quad (5)$$

Then, the logarithmic compression that transforms $(L_{\mathrm{OSA}}, C_{\mathrm{OSA}}, h)$ into (L_E, G_E, J_E) is

$$L_E = \frac{1}{b_L} \ln\left[1 + \frac{b_L}{a_L}(10L_{\mathrm{OSA}})\right] \text{ with} \quad (6a)$$

$$a_L = 2.890, \quad (6b)$$

$$b_L = 0.015, \quad (6c)$$

$$\begin{cases} G_E = -C_E \cos(h), \\ J_E = C_E \sin(h) \end{cases} \text{ with} \quad (7a)$$

$$C_E = \frac{1}{b_C} \ln\left[1 + \frac{b_C}{a_C}(10C_{\mathrm{OSA}})\right], \quad (8a)$$

$$a_C = 1.256, \quad (8b)$$

$$b_C = 0.050. \quad (8c)$$

Finally, the Euclidean color difference formula is

$$\Delta E_E = \sqrt{(\Delta L_E)^2 + (\Delta G_E)^2 + (\Delta J_E)^2}. \quad (9)$$

This formula is as equally good as CIEDE2000 in the prediction of many available empirical datasets, but with higher simplicity [21]. The transformation between the tristimulus space and the OSA-UCS psychometric space is simple and shows a clear relationship with visual processing. No hue distortion, as in CIELAB, exists. The chroma log compression is required in the passage from large to small color differences. Because the image quality is evaluated in the visual situation of the small-medium color differences, we think this formula is a good candidate for the color image metrics [22,23].

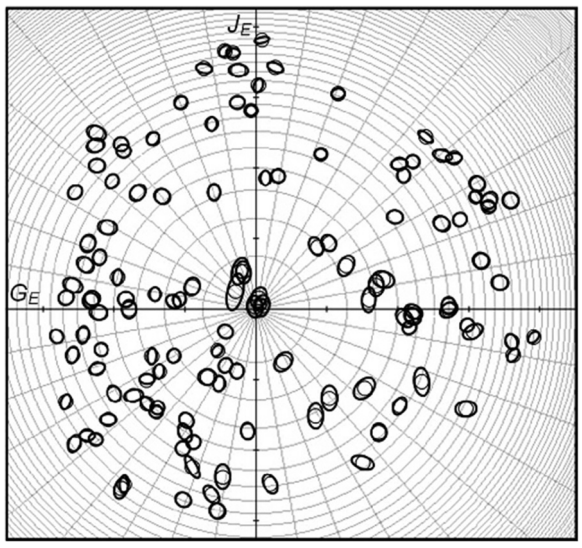


Figure 3 OSA-UCS constant lightness plane with log-compression of the chroma (*cfr.* Figure 2). The BFD ellipses are compared with circles with unitary radius, showing the goodness of the Euclidean color difference formula.

3 Image difference metrics

Nowadays, more than 100 image difference metrics are available in the literature [2], some for general image quality and image difference and some for specific distortion detection, and it would be impossible to report all of

them. For this reason, in this paper we discuss a small selection of important image difference metrics, focusing on those based on contrast filtering or which have been shown to have high correlation with observer perceived difference [2,24]. In this work, we will consider only full-reference image difference metrics, where the original and the reproduction are both available.

In 1997, Zhang and Wandell [25] proposed a spatial extension to the CIELAB color metric. This metric, named Spatial-CIELAB (S-CIELAB), should fulfill two goals: a spatial filtering to simulate the blurring of the HVS and a consistency with the basic CIELAB calculation for large uniform areas. The image is separated into an opponent color space, and each opponent color image is spatially convolved with a kernel defined on the spatial visual sensitivity for the color opponencies. Finally, the filtered image is transformed into the CIELAB system, and ΔE^*_{ab} is used for calculation of the color differences in each pixel; these are averaged to obtain a single quality value for the whole image.

The structural similarity (SSIM) index, proposed by Wang et al. in 2004 [26], attempts to quantify the visible difference between a distorted image and a reference image. This index is based on the UIQ index of Wang and Bovik [27]. The algorithm defines the structural information in an image as those attributes that represent the structure of the objects in the scene, independent of the average luminance and contrast. The index is based on a combination of luminance, contrast and structure comparison. The comparisons are done for local windows in the image, and the overall image quality is the mean of all these local windows.

In 2006, Egiazarian et al. [28] proposed the peak signal-to-noise ratio (PSNR)-HVS metric based on the HVS and PSNR, which is a measure of the peak error between the original and the reproduction. The metric uses a scanning window to remove mean shift and contrast stretching similar to UIQ [27]. PSNR-HVS is then calculated on the scanned images by using PSNR, where MSE is calculated as described by Nill [29]. An extension of this metric that incorporates contrast masking, PSNR-HVS-M, was given by Ponomarenko et al. [30].

In 2007, Chandler and Hemami [31] proposed a new metric, termed visual signal-to-noise ratio (VSNR), based on near-threshold and supra-threshold properties of the human visual system, incorporating both low-level and mid-level features. Low-level features are related to contrast sensitivity and mid-level features to global precedence, which states that the HVS visually integrates image edges in a coarse-to-fine-scale (global-to-local) fashion. The metric consists of two stages: First, contrast thresholds are used to detect visible distortions in the image, which is done in the wavelet domain by computing the contrast signal-to-noise ratio (CSNR). Then, the contrast detection threshold is computed based on the CSNR, which is done for each octave band. The contrast is then compared with the detection threshold, and if above, the distortion is assumed supra-threshold (visible). In this case, a second stage is carried out, where a model of global precedence is proposed to account for mid-level properties of the HVS. The global precedence takes into account that contrasts of distortions should be proportioned across spatial frequency. The final metric is computed as the combination of perceived contrast of the distortion and disruption of global precedence. The VSNR is an interesting metric since it is based on contrast thresholds and does not modulate the entire image as the CSFs in, for example, S-CIELAB.

Spatial-DEE (S-DEE) developed in 2009 is the first metric using the Euclidean color difference formula in log-compressed OSA-UCS space (see Section 2), and it has been proposed by Simone et al. [32]. This metric is based on the modified S-CIELAB workflow developed by Johnson and Fairchild [33] in 2001. The metric is obtained by substituting ΔE^*_{ab} with ΔE_E in the last step of the workflow.

4 From contrast to image difference: WLF-DEE

The history of contrast is over a century long, and a lot of effort has been concerned on developing accurate contrast measures which are able to predict observer perceived contrast in natural images [34]. Furthermore, recent studies have shown that contrast is an important image quality attribute that falls under the umbrella of image quality [3].

In this work, we will propose a new image difference metric based on the work on contrast from Tadmor and Tolhurst [4] with two key features:

1. It uses a multi-level approach (or the so-called pyramidal image structure) as first suggested by Frankle and McCann in 1983 [35] and then Adelson et al. in 1984 [36] in order to perform a full investigation of all the frequencies in the image.
2. It uses the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. [5] and described in detail in Section 2.

In 2000, Tadmor and Tolhurst [4] based their analysis of contrast on the DOG model, which is modified and adapted to natural images. Since this model is used in this work, let us recall it completely:

In the conventional model, the spatial sensitivity in the center of receptive fields is described by a bi-dimensional Gaussian function with unit amplitude:

$$\text{Center}(x,y) = \exp\left[-\left(\frac{x}{r_c}\right)^2 - \left(\frac{y}{r_c}\right)^2\right], \tag{10}$$

where the radius r_c represents the distance at which the sensitivity decreases to $1/e$ and (x,y) are the spatial coordinates of the receptive field. The surround component is represented by another Gaussian curve, with a larger radius, r_s:

$$\text{Surround}(x,y) = \rho \left(\frac{r_c}{r_s}\right)^2 \exp\left[-\left(\frac{x}{r_s}\right)^2 - \left(\frac{y}{r_s}\right)^2\right], \tag{11}$$

where the scaling factor $\rho = 0.85$ sets the integrated sensitivity of the surround component to be 85% of that of the center. This scaling factor used by Tadmor and Tolhurst [4] is representative of the values reported for retinal ganglion cells and lateral geniculate nucleus (LGN) neurons of cat and monkey [37,38]. When the central point of the receptive field is placed at the location (x,y), the output of the central component is calculated as

$$R_c(x,y) = \sum_{i=x-3r_c}^{i=x+3r_c} \sum_{j=y-3r_c}^{j=y+3r_c} \text{Center}(i-x, j-y) I(i,j), \tag{12}$$

while the output of the surround component is

$$R_s(x,y) = \sum_{i=x-3r_s}^{i=x+3r_s} \sum_{j=y-3r_s}^{j=y+3r_s} \text{Surround}(i-x, j-y) I(i,j), \tag{13}$$

where in both cases $I(i,j)$ is the image pixel value at position (i,j).

The result of the DOG model is given by

$$\text{DOG}(x,y) = R_c(x,y) - R_s(x,y). \tag{14}$$

The conventional DOG model [39] assumes that the response of a neuron depends uniquely on the local luminance difference (ΔI) between the center and the surround. After the light adaptation process, the gains of the ganglion cells of the retina and the neurons of the LGN depend on the average local luminance I. Thus, the model response depends on the contrast stimulus. After testing three different outputs for contrast

$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y)} \text{(center-only scheme)}, \tag{15a}$$

$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_s(x,y)} \text{(surround-only scheme)}, \tag{15b}$$

$$C^{TT}(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)} \text{(center-plus-surround scheme)}, \tag{15c}$$

they propose the criterion in Equation 15c for the contrast measuring, which is similar with the Michelson definition of grating contrast [40].

In this approach, a new metric is proposed and referred to as weighted-level framework ΔE_E (WLF-DEE). This metric is analogous to the S-CIELAB approach, but the original S-CIELAB spatial filtering is replaced with a multi-level DOG calculation, while ΔE^*_{ab} is replaced with the ΔE_E color difference formula. The general workflow of the metric is the following: The images are subsampled to various levels. The undersampling is simple since the images are reduced to half, and antialiasing filtering avoids artifacts at low resolutions. A pixelwise neighborhood contrast calculation is executed at each level by using the DOG on the three channels separately. Thus, local contrast maps for each level and each channel are obtained. An example is shown in Figure 4. Local contrast differences are computed by using the ΔE_E, described in Section 2. A weighted recombination of the local contrast maps is computed and represented by a global image difference metric. Once local contrast maps are generated for each level, how to reduce the concept of contrast from local values at each pixel location to a single number representing the global image difference is still subject to debate. The simplest strategy is taking the mean of each level and averaging all together. This new metric performs a weighted recombination of the levels, given by the

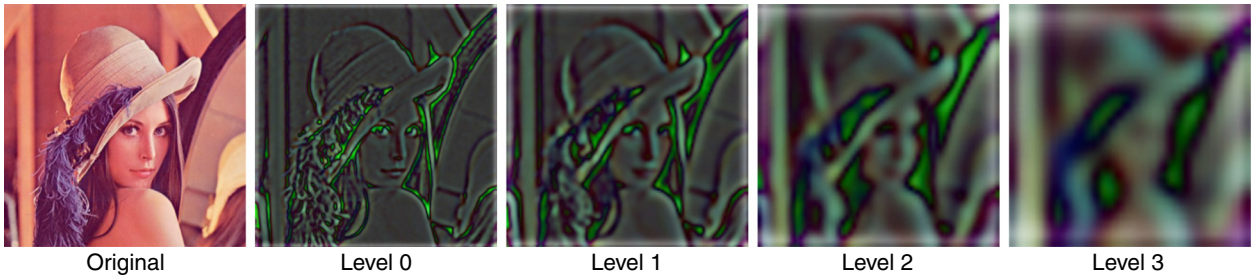


Figure 4 Original and local contrast maps generated by WLF-DEE with $r_c = 2$ and $r_s = 4$. Here the first four filtered levels are shown. The total number of levels is image size dependent. The brightness of all images have been increased to make them suitable for a printable version of the article.

following equation, which leads to the final global image difference:

$$\text{WLF-DEE} = \frac{1}{N_l}\sum_{l=1}^{N_l}\lambda_l \cdot \overline{c}_l, \tag{16}$$

where N_l is the number of levels, $\overline{c}_l$ is the mean contrast in the level l, and λ_l is the weight assigned to each level l. So with the final measure, WLF-DEE aims at predicting the perceived magnitude of contrast between an original and a reproduction. This value can be used also as a quality indicator as contrast is one of the main quality attributes [3].

4.1 WLF-DEE characteristics

4.1.1 Parameter tuning

As we can see from the previous Section, the DOG model is affected by the radius of the center component r_c, the radius of the surround component r_s and the scaling factor of the surround component ρ, in addition to the three different schemes described in Equation 15. With the weighting parameters λ, WLF-DEE is subject to a total of four parameters. Since this metric comes out as a direct application of the research on perceptual contrast of Simone et al. [34] and performing the calculation of all possible values for these four parameters would be computationally exhausting, we have followed their rules of thumb for the choice. Table 1 reports our selection of tested configurations.

We have chosen four different configurations of r_c and r_s, with the particular choice of $r_c = 2$ and $r_s = 4$, which can be considered as the most effective in the identification of edges and blocks in relation to contrast [34]. Using larger radiuses would reduce the identification of local contrast. For the scaling factor of the surround component, in addition to the value $\rho = 0.85$ suggested by Tadmor and Tolhurst [4], we have chosen $\rho = 1.0$ justified by the fact that when the central and surround components are placed on a completely uniform area, the DOG model would give an output of contrast even though no contrast would be perceived by an observer. For the weighting level, we have chosen uniform weighting ($\lambda = 1$), which is the simplest strategy, and the variance of the level, which means that λ assumes values taken from the image itself. Thus, λ becomes level and image dependent. This strategy has shown benefits in correlation with observer perceived contrast [34]. For the three different schemes of the DOG model proposed, we have decided to test all of them. In conclusion, WLF-DEE will be tested for a total of 48 different configurations.

Table 1 Selected values of parameters for WLF-DEE

Configuration	r_c	r_s	ρ	λ
A	1	2	0.85	Uniform
B	2	3	0.85	Uniform
C	3	4	0.85	Uniform
D	2	4	0.85	Uniform
E	1	2	0.85	Variance
F	2	3	0.85	Variance
G	3	4	0.85	Variance
H	2	4	0.85	Variance
I	1	2	1.00	Uniform
J	2	3	1.00	Uniform
K	3	4	1.00	Uniform
L	2	4	1.00	Uniform
M	1	2	1.00	Variance
N	2	3	1.00	Variance
O	3	4	1.00	Variance
P	2	4	1.00	Variance

4.1.2 Computational complexity

The computational complexity of WLF-DEE follows other state-of-the-art image difference metrics using multi-level approach, which is

$$\Theta\left(N \log N\right), \tag{17}$$

where N is the number of pixels in the image.

5 Tests and results

5.1 Databases

Many different databases have been used for evaluating the image difference metrics, but in order to extensively test WLF-DEE, we have chosen three databases: the public Tampere Image Database 2008 (TID2008) and two databases developed at 'Norwegian Colour and Visual Computing Laboratory'.

The first database, the TID2008 database [24], contains a total of 1,700 images, with 25 reference images (Figure 5) and 17 types of distortions over four distortion levels. The mean opinion scores are the results of 654 observers attending the experiments. No viewing distance is stated in the TID database; therefore, we have used a standard viewing distance of 50 cm for the metrics requiring this setting. The authors have decided to include the images where distortions provide directly or indirectly a change in contrast, narrowing the TID2008 to a total of 400 images equally divided in the following four categories: masked noise, quantization noise, denoising and contrast change.

The second database, proposed by Pedersen et al. [41], contains four original images (Figure 6), three portraits and one illustration. The originals were altered in lightness, where each image had four versions with global lightness differences and four versions with local lightness changes. The lightness changes were 3 and 5 ΔE^*_{ab}. Four

Figure 5 TID2008 database [24]. The 25 reference images.

versions were brighter than the original, and four darker, for a total of 32 modified images. The psychophysical experiment was done on a calibrated CRT monitor, LaCie electron 22 blue II (LaCie, Basel, Switzerland), in a grey room . The observers were seated approximately 80 cm from the screen. The light was measured to approximately 17 lux in front of the monitor. A total of 25 observers were recruited for the experiment, and they were asked in a pairwise comparison experiment to choose the image most similar to the original. This database is

Figure 6 Pedersen database [41]. The four reference images changed globally and locally in lightness.

particularly of our interest because contrast is directly related to change in luminance [40], which is related to lightness [7,8].

The third database from Ajagamelle [42] contains a total of 10 original images covering a wide range of characteristics and scenes Figure 7. The images were modified using Adobe Photoshop software on a global scale with separate and simultaneous variations of contrast, lightness and saturation, resulting in a total number of 80 test images. The experiment was carried out as a category-judgment experiment with 14 observers. Each pair of images was displayed on an Eizo ColorEdge CG241W digital LCD display (Eizo Corporation, Ishikawa, Japan). The monitor was calibrated and profiled using Gretag-Macbeth Eye-One Match 3. The settings on the monitor were *sRGB* with a resolution of $1,600 \times 1,200$ pixels. The experiment took place in a windowless room with neutral grey walls, ceiling and floor. The ceiling lights in the room was set to provide a level of ambient illumination around 40 lux, which is below the upper threshold of 64 recommended by the CIE [43]. The white point was set to the D65, the gamma to 2.2 and the luminance level to 80 cd/m^2. The display was placed at a viewing distance of 70 cm. The images presented were 750×499 pixels or 499×750 pixels, which subtended roughly 20° of the visual angle when viewed at this distance.

5.2 Performance measures

Two types of correlation coefficients are computed [45] in order to evaluate the performance of WLF-DEE:

1. *The Pearson product-moment.* It assumes that the variables are ordinal, and it evaluates the linear relationship between two variables. This is a performance measure relating to the prediction accuracy of the metric [46].
2. *Spearman rank.* It is a non-parametric measure of correlation, and it is used as a measure of linear relationship between two sets of ranked data, instead of the actual values. This describes the relationship between variables with no assumptions on the frequency distribution of the variables and on how tightly the ranked data clusters are around a straight line. This is a performance measure relating to the prediction monotonicity of the metric [46].

The relationships between the metrics and the observers are not necessarily linear. In order to remove any non-linearities due to the subjective experimental process and to facilitate comparison of the metrics in a common analysis space, we investigate the relationship between the metrics and observers by using non-linear regression [46]. In this work, we apply the same mapping function as that of Sheikh et al. [47]:

$$f(x) = \theta_1 \left(\frac{1}{2} - \frac{1}{1 + e^{\theta_2 (x - \theta_3)}} \right) + \theta_4 X + \theta_5, \tag{18}$$

where θ_i, $i = 1, 2, 3, 4, 5$, are parameters to be be fitted. The 95% confidence intervals for the correlation values are calculated using Fisher's Z transformation as described by the Video Quality Expert Group [48].

In order to have a complete analysis the following coefficients are also presented:

- *Root-mean-square error (RMSE)* [48]. It is a measure of the differences between the values predicted by the metric and the scores actually given by the observers.
- *Significance of the difference between the Pearson correlation coefficients (t-value)* [48]. This measure assumes that a good fit for observers' quality score is given by the normal distribution. It uses the H_0 hypothesis that assumes that there is no significant difference between correlation coefficients and the H_1 hypothesis, which considers that the difference is significant, although not specifying better or worse.

Figure 7 Ajagamelle database [42]. Nine images were captured and provided by an independent photographer, and one image was selected from a standard natural image set provided by the CIE [44].

Table 2 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on all databases

Metric	TID (masked noise)		TID (quantization noise)		TID (image denoising)		TID (contrast change)		Ajagamelle		Pedersen	
	Linear	Logistic	Linear	Logistic	Linear	Logistic	Linear	Logistic	Linear	Logistic	Linear	Logistic
PSNR	0.57	0.59	0.50	0.51	0.55	0.55	0.47	0.49	0.73[a]	0.80[a]	0.66	0.66
PSNR-HVS-M	0.57	0.58	0.52	0.53	0.56	0.57	0.49	0.49	0.72	0.78	0.66	0.66
S-CIELAB	0.62	0.68	0.63	0.66	0.66	0.71	0.60	0.64	0.67	0.79	0.80[a]	0.85[a]
SSIM	0.71[a]	0.75[a]	0.83[a]	0.85[a]	0.79[a]	0.81[a]	0.73	0.76	0.64	0.64	0.22	0.22
VSNR	0.52	0.53	0.45	0.50	0.53	0.54	0.47	0.47	0.16	0.28	0.02	0.02
WLF-DEE-C I	0.60	0.64	0.72	0.75	0.62	0.62	0.65[b]	0.69[b]	0.39[b]	0.65[b]	0.27[b]	0.49[b]
WLF-DEE-C J	0.61[b]	0.64	0.74	0.76	0.64	0.64	0.62	0.66	0.39	0.62	0.27[b]	0.44
WLF-DEE-C K	0.61[b]	0.68[b]	0.75[b]	0.77[b]	0.66[b]	0.71[b]	0.63	0.67	0.38	0.60	0.27[b]	0.44
WLF-DEE-C L	0.57	0.62	0.68	0.71	0.58	0.66	0.59	0.67	0.40	0.64	0.27[b]	0.45
WLF-DEE-C M	0.35	0.38	0.25	0.55	0.22	0.22	0.25	0.53	0.21	0.26	0.16	0.17
WLF-DEE-C N	0.38	0.53	0.33	0.58	0.23	0.60	0.29	0.55	0.20	0.20	0.16	0.19
WLF-DEE-C O	0.39	0.56	0.38	0.62	0.24	0.62	0.34	0.57	0.20	0.34	0.17	0.40
WLF-DEE-C P	0.32	0.53	0.26	0.57	0.19	0.60	0.24	0.54	0.21	0.31	0.17	0.47

[a]The most performant metric. [b]The highest performance of WLF-DEE.

Table 3 Spearman correlation for WLF-DEE and the selected state-of-the-art metrics on all databases

Metric	TID (masked noise)	TID (quantization noise)	TID (image denoising)	TID (contrast change)	Ajagamelle	Pedersen
	Linear	Linear	Linear	Linear	Linear	Linear
PSNR	0.55	0.47	0.53	0.45	0.71	0.72
PSNR-HVS-M	0.53	0.46	0.53	0.45	0.70	0.72
S-CIELAB	0.66	0.63	0.69	0.62	0.72	0.77[a]
SSIM	0.74[a]	0.83[a]	0.78[a]	0.74[a]	0.72[a]	0.49
VSNR	0.48	0.39	0.49	0.43	0.14	0.02
WLF-DEE-C I	0.60[b]	0.71[b]	0.66[b]	0.63[b]	0.61[b]	0.43
WLF-DEE-C J	0.59	0.69	0.65	0.59	0.59	0.42
WLF-DEE-C K	0.59	0.70	0.66[b]	0.60	0.57	0.44[b]
WLF-DEE-C L	0.57	0.65	0.62	0.59	0.60	0.43
WLF-DEE-C M	0.43	0.48	0.50	0.46	0.56	0.37
WLF-DEE-C N	0.50	0.56	0.56	0.48	0.55	0.38
WLF-DEE-C O	0.53	0.59	0.58	0.48	0.53	0.38
WLF-DEE-C P	0.48	0.51	0.53	0.47	0.56	0.37

[a]The most performant metric. [b]The highest performance of WLF-DEE.

Table 4 RMSE for WLF-DEE and the selected state-of-the-art metrics on all databases

Metric	TID (masked noise)	TID (quantization noise)	TID (image denoising)	TID (contrast change)	Ajagamelle	Pedersen
PSNR	14.73	15.38	15.72	15.91	1.49	0.40
PSNR-HVS-M	14.84	15.15	15.54	15.85	1.54	0.40
S-CIELAB	13.35	13.41	13.29	13.94	0.88	0.28[a]
SSIM	11.99[a]	9.45[a]	10.99[a]	11.80[a]	1.75	0.53
VSNR	15.42	15.41	15.92	16.07	1.03	0.53
WLF-DEE-C I	13.99	11.88	14.79	13.11[b]	1.01	0.46
WLF-DEE-C J	13.98	11.54	14.51	13.73	1.01	0.47
WLF-DEE-C K	13.33[b]	11.36[b]	13.33[b]	13.53	1.01	0.48
WLF-DEE-C L	14.36	12.54	14.22	13.56	1.02	0.47
WLF-DEE-C M	16.88	14.93	18.39	15.49	0.94	0.52
WLF-DEE-C N	15.48	14.55	15.12	15.19	0.78[a,b]	0.52
WLF-DEE-C O	15.12	14.03	14.74	14.98	1.06	0.49
WLF-DEE-C P	15.48	14.64	15.15	15.29	1.43	0.47

[a]The most performant metric. [b]The highest performance of WLF-DEE.

Table 5 Significance of the difference between Pearson correlation coefficients on TID-masked noise database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	0.08	-											
S-CIELAB	−1.08	−1.16	-										
SSIM	−2.13[a]	−2.21[a]	−1.04	-									
VSNR	0.57	0.49	1.65	2.69[a]	-								
WLF-DEE-C I	−0.59	−0.67	0.50	1.54	−1.16	-							
WLF-DEE-C J	−0.60	−0.68	0.48	1.53	−1.17	−0.01	-						
WLF-DEE-C K	−1.10	−1.19	−0.02	1.02	−1.67	−0.52	−0.50	-					
WLF-DEE-C L	−0.30	−0.38	0.78	1.82	−0.87	0.28	0.30	0.80	-				
WLF-DEE-C M	1.94	1.86	3.02[a]	4.07[a]	1.37	2.53[a]	2.54[a]	3.05[a]	2.24[a]	-			
WLF-DEE-C N	0.62	0.54	1.70	2.75[a]	0.05	1.21	1.22	1.72	0.92	−1.32	-		
WLF-DEE-C O	0.32	0.23	1.40	2.44[a]	−0.25	0.90	0.92	1.42	0.62	−1.63	−0.30	-	
WLF-DEE-C P	0.62	0.53	1.70	2.74[a]	0.05	1.20	1.21	1.72	0.92	−1.33	0.00	0.30	-

[a]Significantly different at 5% significance level (1.96).

Table 6 Significance of the difference between Pearson correlation coefficients on TID-quantization noise database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	−0.20	-											
S-CIELAB	−1.63	−1.43	-										
SSIM	−4.81[a]	−4.62[a]	−3.19[a]	-									
VSNR	0.03	0.22	1.66	4.84[a]	-								
WLF-DEE-C I	−2.82[a]	−2.63[a]	−1.19	1.99[a]	−2.85[a]	-							
WLF-DEE-C J	−3.10[a]	−2.90[a]	−1.47	1.72	−3.12[a]	−0.27	-						
WLF-DEE-C K	−3.23[a]	−3.04[a]	−1.61	1.58	−3.26[a]	−0.41	−0.14	-					
WLF-DEE-C L	−2.30[a]	−2.11[a]	−0.67	2.51[a]	−2.33[a]	0.52	0.79	0.93	-				
WLF-DEE-C M	−0.39	−0.19	1.24	4.43[a]	−0.42	2.43[a]	2.71[a]	2.85[a]	1.91	-			
WLF-DEE-C N	−0.71	−0.51	0.92	4.11[a]	−0.73	2.12[a]	2.39[a]	2.53[a]	1.60	−0.32	-		
WLF-DEE-C O	−1.13	−0.94	0.49	3.68[a]	−1.16	1.69	1.96	2.10[a]	1.17	−0.75	−0.43	-	
WLF-DEE-C P	−0.64	−0.44	0.99	4.18[a]	−0.66	2.19[a]	2.46[a]	2.60[a]	1.67	−0.25	0.07	0.50	-

[a]Significantly different at 5% significance level (1.96).

Table 7 Significance of the difference between Pearson correlation coefficients on TID-image denoising database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	−0.14	-											
S-CIELAB	−1.84	−1.70	-										
SSIM	−3.57[a]	−3.43[a]	−1.73	-									
VSNR	0.16	0.30	2.00[a]	3.73[a]	-								
WLF-DEE-C I	−0.73	−0.58	1.12	2.85[a]	−0.89	-							
WLF-DEE-C J	−0.94	−0.79	0.91	2.64[a]	−1.10	−0.21	-						
WLF-DEE-C K	−1.81	−1.66	0.04	1.76	−1.97	−1.08	−0.87	-					
WLF-DEE-C L	−1.15	−1.01	0.69	2.42[a]	−1.31	−0.43	−0.22	0.66	-				
WLF-DEE-C M	2.77[a]	2.91[a]	4.61[a]	6.34[a]	2.61[a]	3.49[a]	3.70[a]	4.57[a]	3.92[a]	-			
WLF-DEE-C N	−0.47	−0.33	1.37	3.10[a]	−0.63	0.25	0.46	1.34	0.68	−3.24[a]	-		
WLF-DEE-C O	−0.76	−0.62	1.08	2.81[a]	−0.92	−0.04	0.17	1.05	0.39	−3.53[a]	−0.29	-	
WLF-DEE-C P	−0.45	−0.30	1.39	3.12[a]	−0.61	0.28	0.49	1.36	0.70	−3.21[a]	0.02	0.31	-

[a]Significantly different at 5% significance level (1.96).

Table 8 Significance of the difference between Pearson correlation coefficients on TID-contrast change database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	−0.06	-											
S-CIELAB	−1.62	−1.56	-										
SSIM	−3.27[a]	−3.21[a]	−1.65	-									
VSNR	0.15	0.20	1.77	3.41[a]	-								
WLF-DEE-C I	−2.26[a]	−2.21[a]	−0.64	1.01	−2.41[a]	-							
WLF-DEE-C J	−1.78	−1.73	−0.16	1.49	−1.93	0.48	-						
WLF-DEE-C K	−1.94	−1.88	−0.32	1.33	−2.08[a]	0.32	−0.16	-					
WLF-DEE-C L	−1.92	−1.86	−0.30	1.35	−2.06[a]	0.34	−0.14	0.02	-				
WLF-DEE-C M	−0.37	−0.31	1.25	2.90[a]	−0.51	1.89	1.41	1.57	1.55	-			
WLF-DEE-C N	−0.62	−0.57	1.00	2.65[a]	−0.77	1.64	1.16	1.32	1.30	−0.25	-		
WLF-DEE-C O	−0.79	−0.74	0.83	2.48[a]	−0.94	1.47	0.99	1.15	1.13	−0.42	−0.17	-	
WLF-DEE-C P	−0.54	−0.48	1.08	2.73[a]	−0.68	1.72	1.24	1.40	1.38	−0.17	0.08	0.25	-

[a]Significantly different at 5% significance level (1.96).

Table 9 Significance of the difference between Pearson correlation coefficients on Ajagamelle database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	0.29	-											
S-CIELAB	0.09	−0.20	-										
SSIM	2.08[a]	1.79	1.99[a]	-									
VSNR	4.96[a]	4.67[a]	4.87[a]	2.88[a]	-								
WLF-DEE-C I	1.96	1.67	1.87	−0.12	−3.00[a]	-							
WLF-DEE-C J	2.28[a]	1.99[a]	2.19[a]	0.20	−2.68[a]	0.32	-						
WLF-DEE-C K	2.47[a]	2.18[a]	2.38[a]	0.39	−2.49[a]	0.51	0.20	-					
WLF-DEE-C L	2.01[a]	1.73	1.93	−0.07	−2.95[a]	0.06	−0.26	−0.46	-				
WLF-DEE-C M	5.11[a]	4.83[a]	5.02[a]	3.03[a]	0.15	3.16[a]	2.84[a]	2.64[a]	3.10[a]	-			
WLF-DEE-C N	5.46[a]	5.17[a]	5.37[a]	3.38[a]	0.50	3.50[a]	3.18[a]	2.99[a]	3.44[a]	0.35	-		
WLF-DEE-C O	4.51[a]	4.23[a]	4.42[a]	2.43[a]	−0.45	2.56[a]	2.24[a]	2.04[a]	2.50[a]	−0.60	−0.95	-	
WLF-DEE-C P	4.76[a]	4.48[a]	4.67[a]	2.68[a]	−0.20	2.81[a]	2.49[a]	2.29[a]	2.75[a]	−0.35	−0.70	0.25	-

[a]Significantly different at 5% significance level (1.96).

Table 10 Significance of the difference between Pearson correlation coefficients on Pedersen database

Metric	PSNR	PSNR-HVS-M	S-CIELAB	SSIM	VSNR	WLF-DEE-C I	WLF-DEE-C J	WLF-DEE-C K	WLF-DEE-C L	WLF-DEE-C M	WLF-DEE-C N	WLF-DEE-C O	WLF-DEE-C P
PSNR	-												
PSNR-HVS-M	−0.03	-											
S-CIELAB	−1.83	−1.80	-										
SSIM	2.15[a]	2.18[a]	3.98[a]	-									
VSNR	2.90[a]	2.93[a]	4.73[a]	0.75	-								
WLF-DEE-C I	0.93	0.96	2.76[a]	−1.22	−1.97[a]	-							
WLF-DEE-C J	1.17	1.20	3.00[a]	−0.98	−1.73	0.24	-						
WLF-DEE-C K	1.18	1.21	3.01[a]	−0.97	−1.72	0.25	0.01	-					
WLF-DEE-C L	1.16	1.19	2.99[a]	−0.99	−1.74	0.23	−0.01	−0.02	-				
WLF-DEE-C M	2.33[a]	2.35[a]	4.15[a]	0.18	−0.58	1.39	1.16	1.14	1.17	-			
WLF-DEE-C N	2.27[a]	2.30[a]	4.10[a]	0.12	−0.63	1.34	1.10	1.09	1.11	−0.06	-		
WLF-DEE-C O	1.39	1.42	3.22[a]	−0.76	−1.51	0.46	0.22	0.21	0.23	−0.93	−0.87	-	
WLF-DEE-C P	1.06	1.09	2.89[a]	−1.09	−1.84	0.13	−0.11	−0.12	−0.10	−1.26	−1.21	−0.33	-

[a]Significantly different at 5% significance level (1.96).

5.3 Results

As mentioned in Section 4.1.1, WLF-DEE has been tested for a total of 48 configurations, but in order to give a more readable and understandable presentation of the results, we will present only a selection of them. As WLF-DEE using the two DOG schemes in Equations 15a and 15b have lower performance in correlation with respect to WLF-DEE using the DOG scheme in Equation 15(c), these results will be excluded. This confirms also the statement of Tadmor and Tolhurst that the DOG model in analogy with the Michelson formula has better performance for contrast assessment [4]. On the same way, all configurations with $\rho = 0.85$ in Equation 13 will be excluded, as they show lower performance in correlation with respect to those configurations with $\rho = 1.00$. This will end in a presentation of a total of only eight results of WLF-DEE shown in Table 2 for Pearson correlation, Table 3 for Spearman correlation and Table 4 for RMSE. Significance of the difference between the Pearson correlation coefficients are presented for each database in Tables 5, 6, 7, 8, 9 and 10.

Considering the TID database, SSIM has the higher Pearson correlation in all four categories. In the masked noise category, SSIM is followed by WLF-DEE K, while in the quantization noise, it is followed by S-CIELAB and then WLF-DEE K. In the category denoising and contrast change, SSIM instead is followed by WLF-DEE I and then WLF-DEE K. For all the four categories, it is possible to notice that all the metrics give higher correlation with perceived observer difference using the logistic fitting. Furthermore, as the confidence intervals (Figures 8, 9, 10 and 11) of WLF-DEE K overlap with the confidence intervals of SSIM, we can claim to have the same performance. Overall, for the four categories of the TID database, WLF-DEE K shows to be significantly better than PSNR and VSNR and to have the same performance with SSIM and S-CIELAB.

For the Ajagamelle database, PSNR shows the higher Pearson correlation, followed by PSNR-HVS-M, SCIELAB, SSIM and then four configurations (I, J, K, L) of WLF-DEE, which have very close results. Since in this case, the confidence intervals (Figure 12) of WLF-DEE K overlaps with the confidence interval of those metrics with slightly higher correlation, we can claim to have the same performance. For the Pedersen database, S-CIELAB shows the higher Pearson correlation, followed by PSNR-HVS-M, PSNR, four configurations (I, J, K, L) of WLF-DEE with very close results and then SSIM. In this database instead, confidence intervals (Figure 13) show that WLF-DEE-K has a slightly lower performance than S-CIELAB, but not with respect to SSIM. Also for these two databases, it holds true that all the metrics give higher correlation using logistic fitting.

Considering all the six database sets examined, WLF-DEE gives higher correlation using configurations I, J, K, L with respect to configurations M, N, O, P, and in particular, WLF-DEE K most agrees with observer perceived difference, indicating that large radiuses of the Gaussians and uniform weighting of the levels should be used for the estimation of perceived difference. Furthermore, it is possible to notice that WLF-DEE K with logistic fitting shows a stable trend among the six datasets having an average performance in correlation of 0.65. This holds true also for other tested metrics such as S-CIELAB and PSNR-HVS-M, but not for SSIM and VSNR which show very high correlation in one dataset and very low in an another one.

Analysis with the Spearman correlation follows the same discussion with the Pearson correlation except for the Ajagamelle database, where the highest correlation is shown by S-CIELAB, but not outperforming most of the other metrics. The results are presented only with linear fitting as no improvements can be found in any metric using the logistic fitting. Also with the Spearman correlation, WLF-DEE K shows its stability with an average performance of 0.59.

Analysis with root-mean-square error shows that for all the four categories of the TID database, SSIM has the lowest RMSE. As the confidence intervals of SSIM overlap with two configurations of WLF-DEE-C (J, K) (Figures 14, 15, 16 and 17), it cannot be claimed that the two metrics are significantly different in performance.

For Ajagamelle database instead, WLF-DEE-C N shows the lowest RMSE, followed by S-CIELAB and then by other several configurations of WLF-DEE-C (M, I, J, K, L) and VSNR. Confidence intervals (Figure 18) shows that these three metrics are not significantly different among each other but they outperform other tested metrics such as SSIM, PSNR and PSNR-HVS-M. For Pedersen database, S-CIELAB shows the lowest RMSE followed by PSNR, PSNR-HVS-M and then all the configurations of WLF-DEE-C. SSIM and VSNR have the highest RMSE. Confidence intervals (Figure 19) show that WLF-DEE-C (I, J, K, L, P) has no difference in performance with the other tested metrics but the overlap with S-CIELAB confidence interval is minimal.

Analysis with significance of the difference are presented with 5% significance level for Pearson correlation with logistic fitting only. Based on the definition in [48], two metrics can be significantly different if $-1.96 < t - value < 1.96$. This analysis confirms that WLF-DEE-C K is not significantly different in performance with respect to the other tested metrics for TID-masked noise and TID-image denoising databases. For TID-quantization noise instead, WLF-DEE-C K is

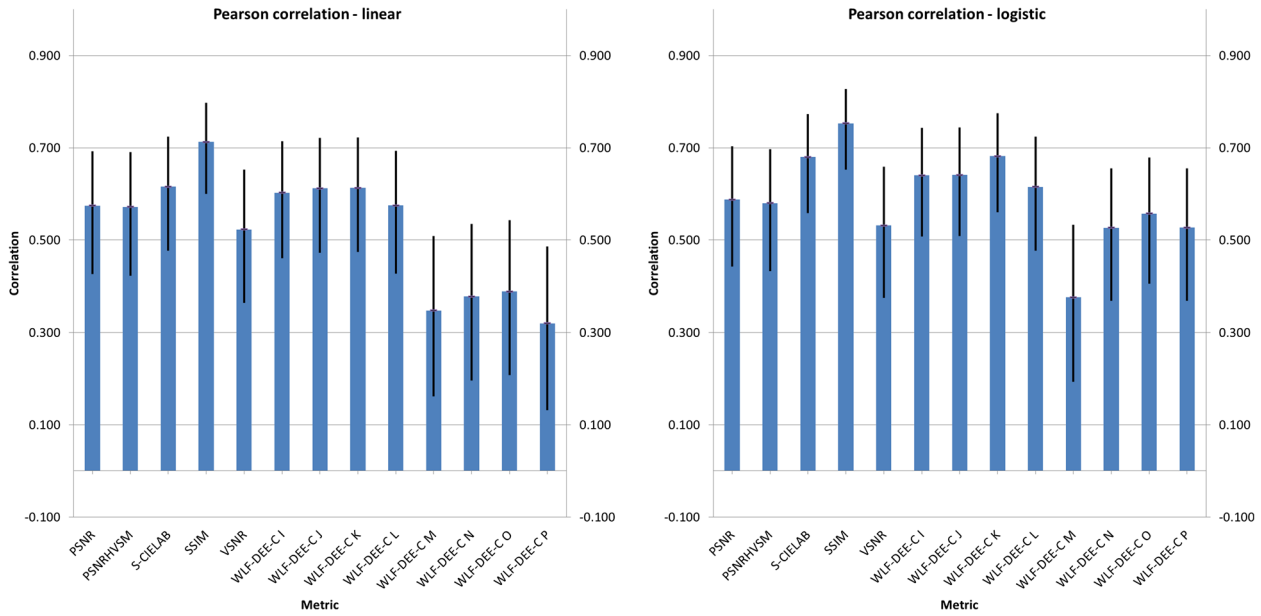


Figure 8 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-masked noise database. Results using linear fitting on the left and logistic fitting on the right.

not significantly difference in performance with respect to SSIM and S-CIELAB. For TID-contrast change WLF-DEE-C K is significantly difference in performance only with respect to VSNR. For Ajagamelle database, WLF-DEE-C K is not significantly different in performance only from SSIM while for Pedersen database is not significantly different in performance from the other tested metrics except S-CIELAB.

Overall, WLF-DEE-C K shows its particular strength on those databases where a change in contrast between the

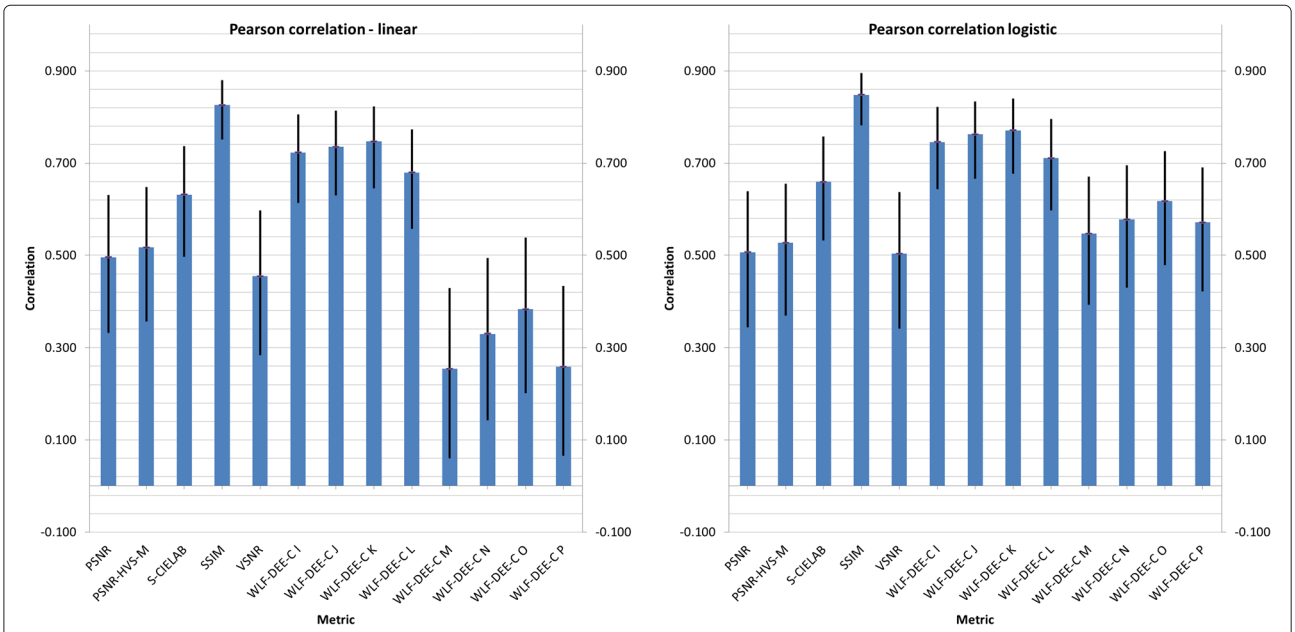


Figure 9 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-quantization noise database. Results using linear fitting on the left and logistic fitting on the right.

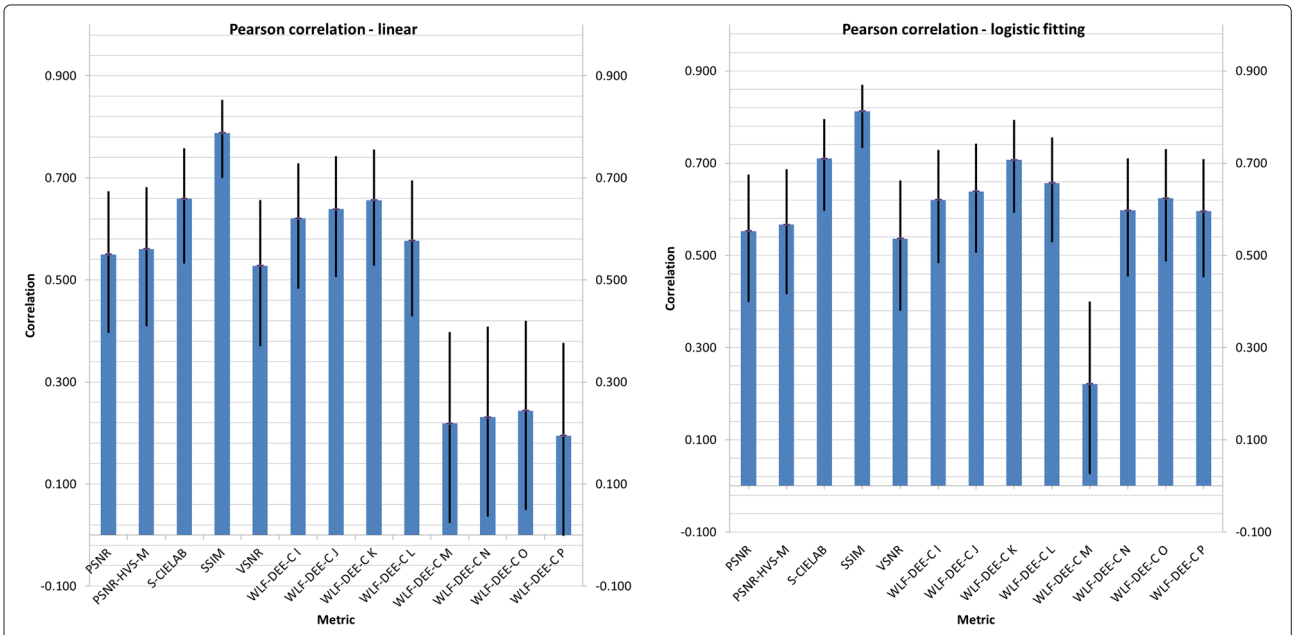


Figure 10 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-image denoising database. Results using linear fitting on the left and logistic fitting on the right.

original and its reproduction is triggered by a change of color attributes and not particular distortions. In conclusion, WLF-DEE-C K promotes itself as a new metric for predicting the perceived magnitude of contrast between an original and a reproduction, fulfilling the purpose for which it was developed.

6 Conclusions

Recent studies have shown that contrast is an important image attribute that falls under the umbrella of image quality [3]. In this paper, we have developed a new metric based on recent work on contrast filtering. This metric called WLF-DEE consists of two key features: a multi-

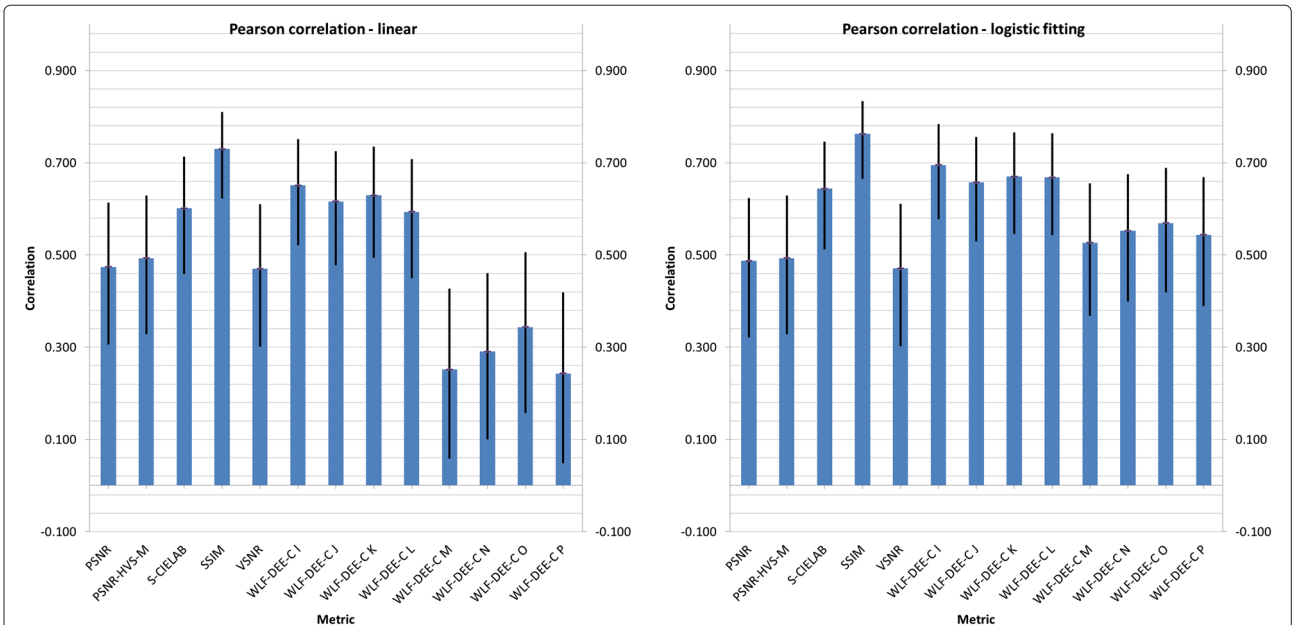


Figure 11 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on TID-contrast change database. Results using linear fitting on the left and logistic fitting on the right.

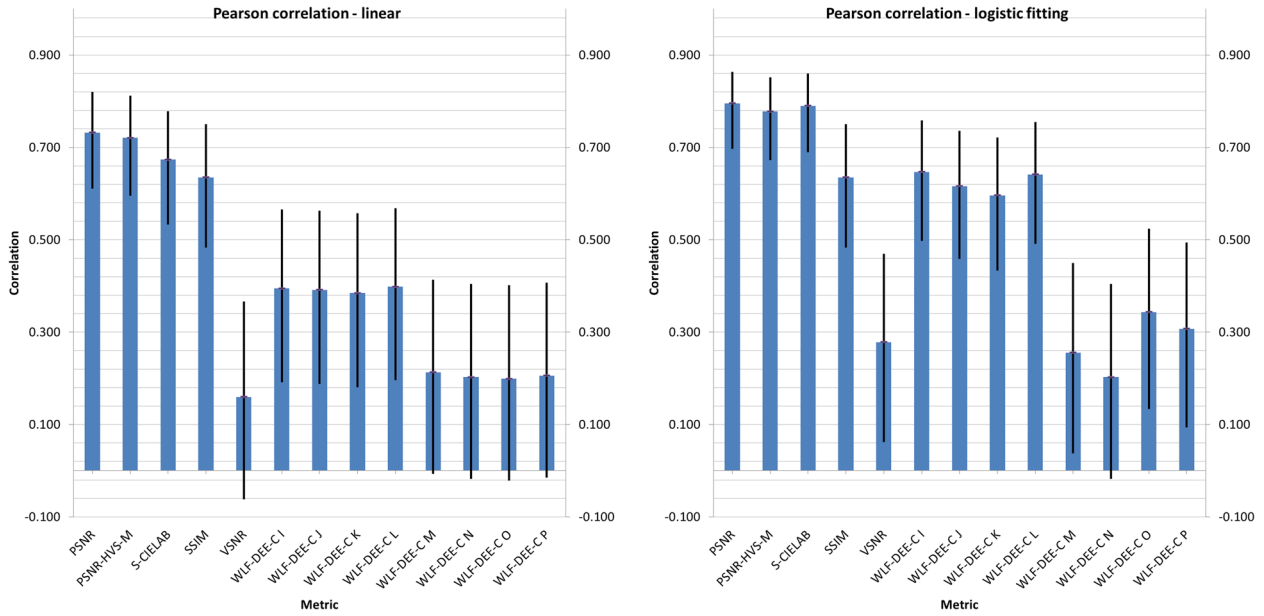


Figure 12 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on Ajagamelle database. Results using linear fitting on the left and logistic fitting on the right.

level filtering based on the work on contrast from Tadmor and Tolhurst [4] and the new Euclidean color difference formula in log-compressed OSA-UCS space proposed by Oleari et al. [5].

Extensive tests and analysis are carried out on four different categories of the well-known TID database and on two databases providing different distortions directly related to color and contrast. Comparisons with other state-of-the-art metrics are also presented. Results promote WLF-DEE as a new efficient metric for estimating the perceived magnitude of contrast between an original and a reproduction.

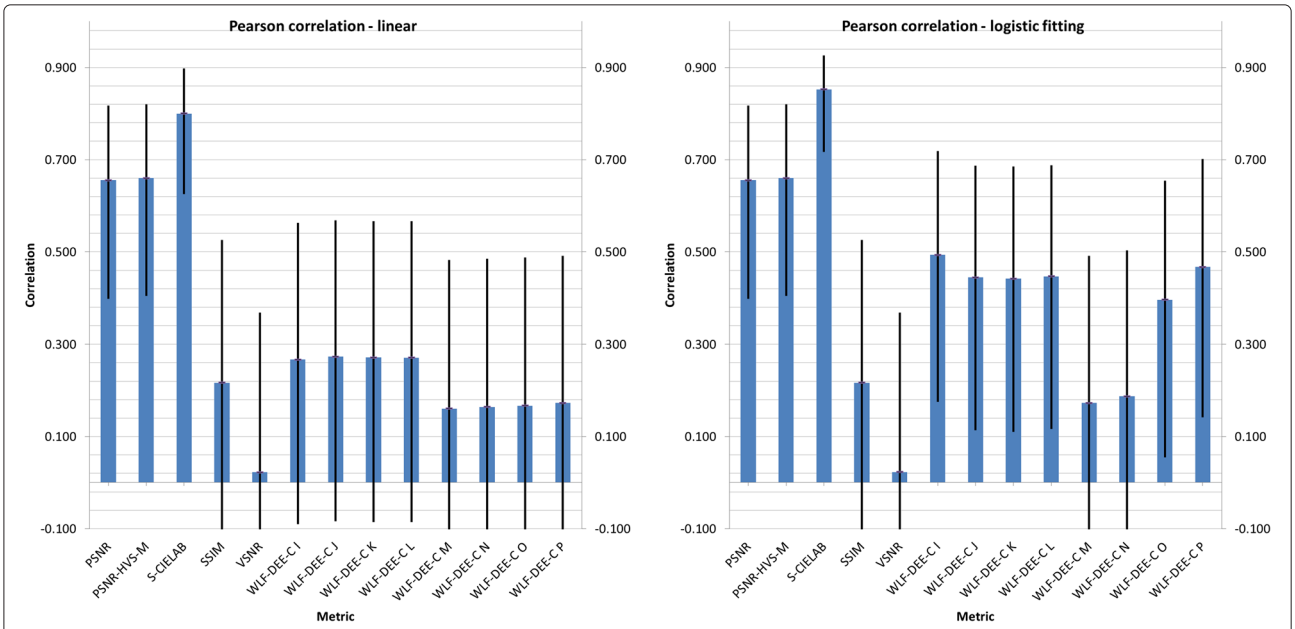


Figure 13 Pearson correlation for WLF-DEE and the selected state-of-the-art metrics on Pedersen database. Results using linear fitting on the left and logistic fitting on the right.

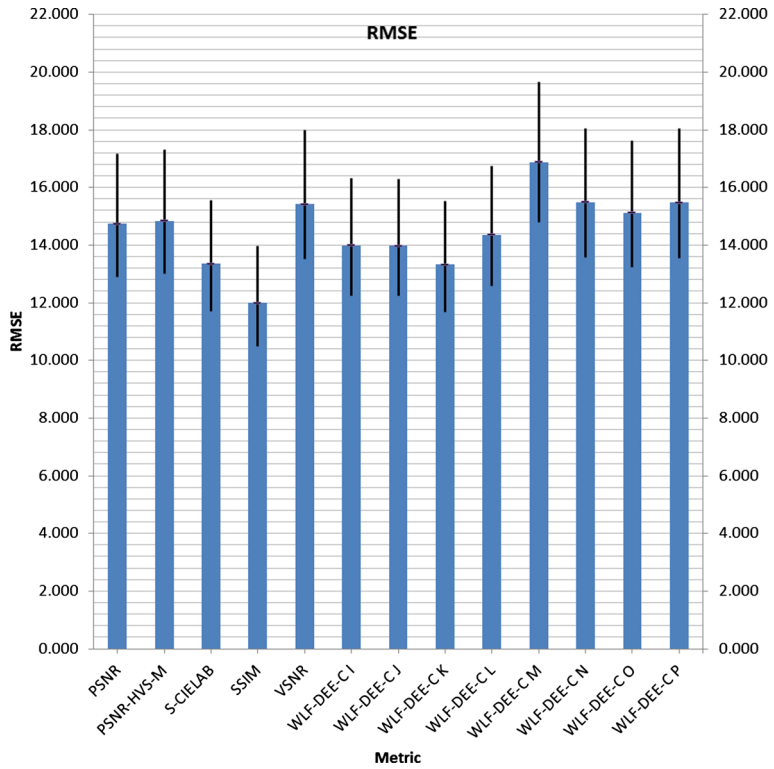


Figure 14 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-masked noise database.

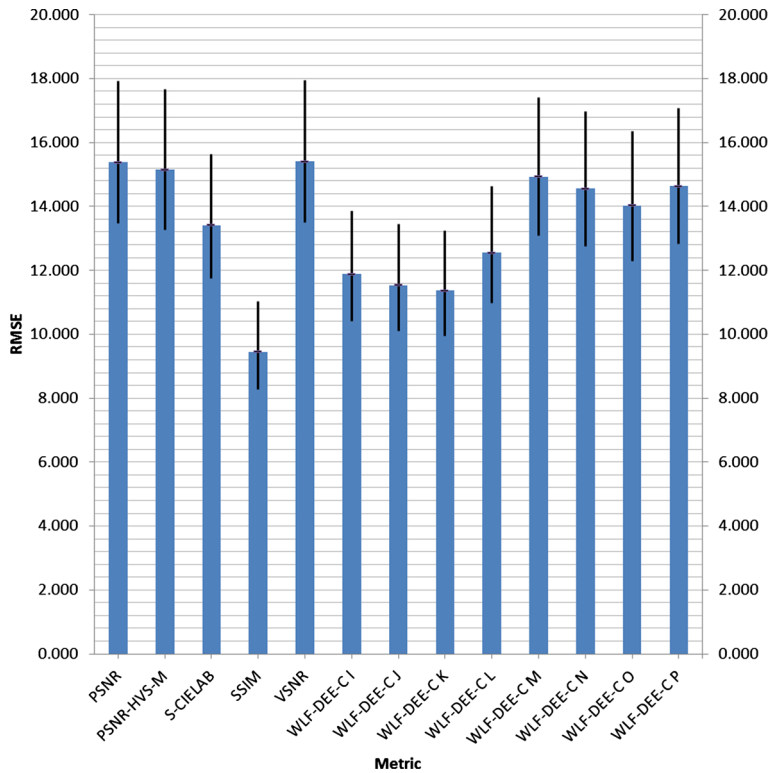


Figure 15 RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-quantization noise database.

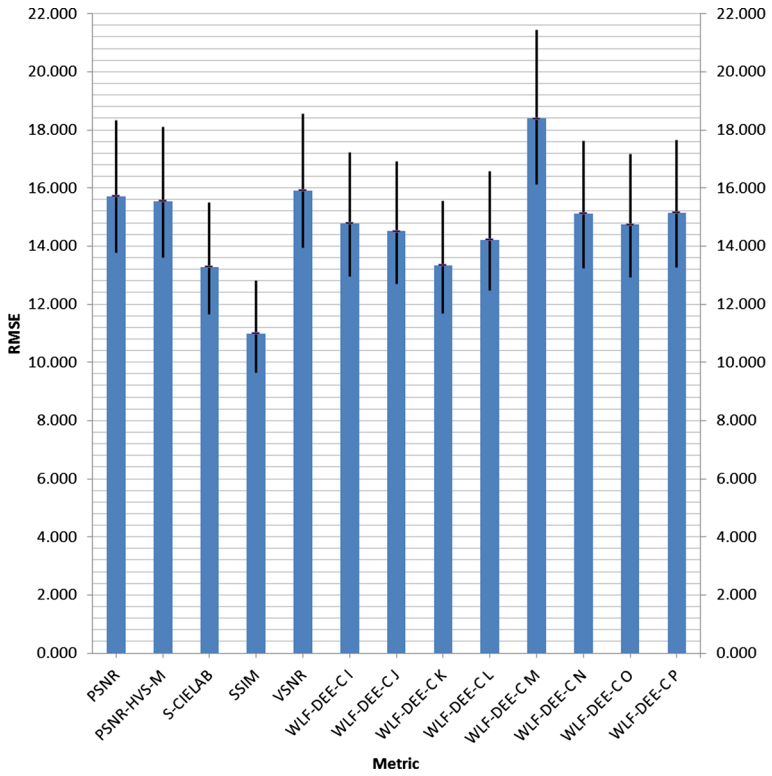


Figure 16 **RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-image denoising database.**

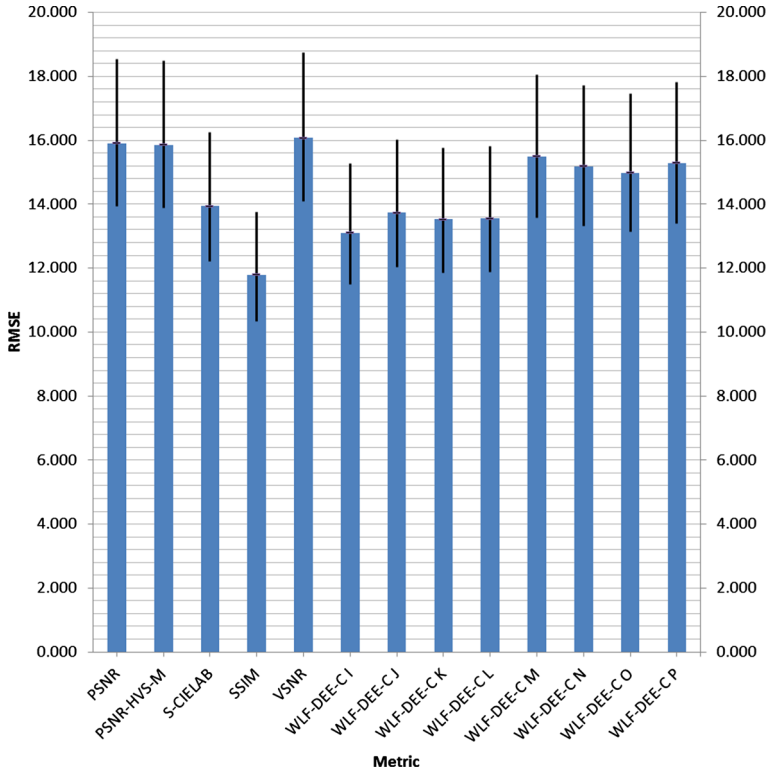


Figure 17 **RMSE for WLF-DEE and the selected state-of-the-art metrics on TID-contrast change database.**

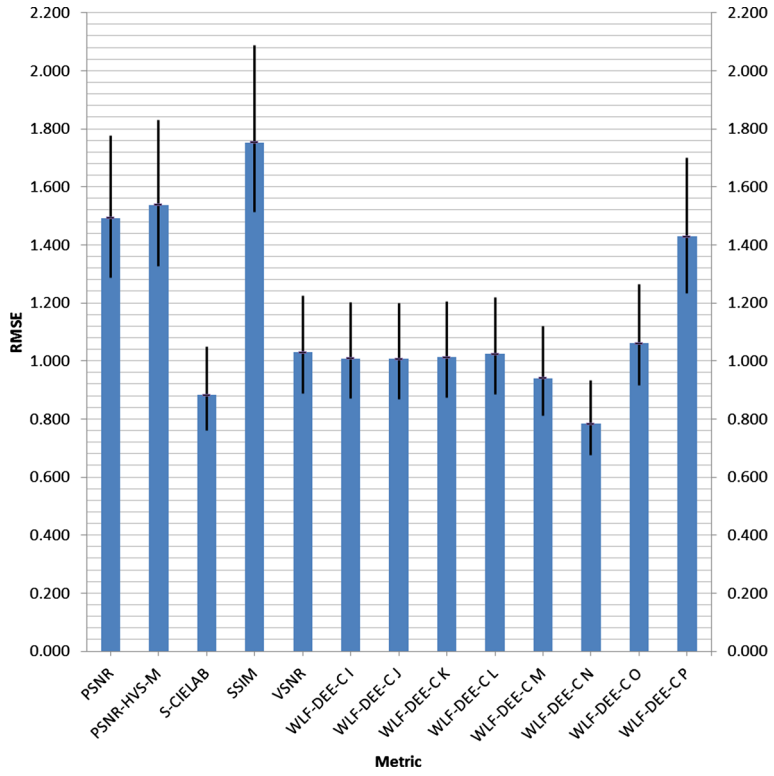


Figure 18 **RMSE for WLF-DEE and the selected state-of-the-art metrics on Ajagamelle database.**

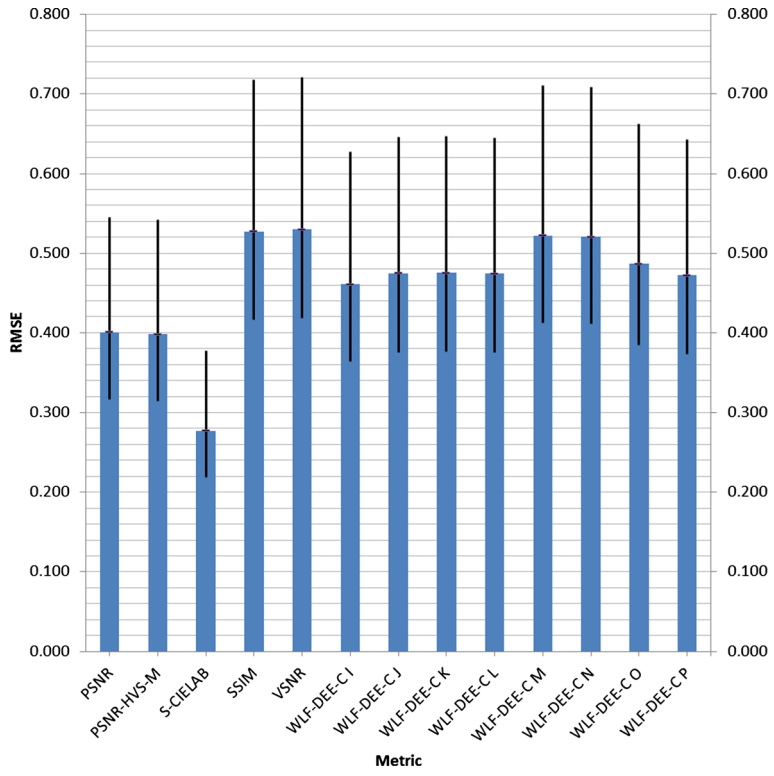


Figure 19 **RMSE for WLF-DEE and the selected state-of-the-art metrics on Pedersen database.**

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work has been supported by NFR over the SHP project. The authors would like to thank Fritz Albregtsen (University of Oslo), Sebastien Ajagamelle, Valentina Caracciolo and Sara Díaz De Cerio for their contribution in part of the project.

Author details
[1]The Norwegian Colour and Visual Computing Laboratory, Faculty of Computer Science and Media Technology, Gjøvik University College, P.O. Box 191, Gjøvik 2802, Norway. [2]Dipartimento di Fisica, Universita' degli Studi di Parma, Parco Area delle Scienze 7/A, Parma 43100, Italy.

Received: 8 January 2013 Accepted: 29 May 2013
Published: 16 July 2013

References

1. A Poirson, B Wandell, Appearance of colored patterns: pattern-color separability. J. Opt. Soc. Am. A. **10**(12), 2458–2470 (1993)
2. M Pedersen, JY Hardeberg, Survey of full-reference image quality metrics: classification and evaluation. Foundations Trends. Comput. Graph. Vis. **7**, 1–80 (2012)
3. M Pedersen, N Bonnier, JY Hardeberg, F Albregtsen, Attributes of image quality for color prints. J. Electron. Imaging. **19**, 011016–1–011016–13 (2010)
4. Y Tadmor, D Tolhurst, Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. Vis. Res. **40**, 3145–3157 (2000)
5. C Oleari, M Melgosa, R Huertas, Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space. J. Opt. Soc. Am. A **26**, 121–134 (2009)
6. H Von Helmholtz, Versuch das psychophysische Gesetz auf die Farbenunterschiede trichromatischer Augen anzuwenden. Z Psychol Physiol Sinnesorgane **3**, 1–20 (1892)
7. CIE, Colorimetry. Tech. Rep. 15. (CIE, Vienna, 2004)
8. Colorimetry – Part 4: CIE 1976 L*a*b* Colour space, ISO 11664-4:2008 (CIE S 014-4/E:2007). CIE, Vienna, 2006
9. MR Luo, B Rigg, Chromaticity-discrimination ellipses for surface colours. Color Res. Appli. **11**, 25–42 (1986)
10. FJJ Clarke, R McDonald, B Rigg, Modification to the JPC79 color-difference formula. J. Soc. Dyers Colorists. **100**, 128–132 (1984)
11. M Luo, B Rigg, BFD(l:c) colour-difference formula: Part I - development of the formula. J. Soc. Dyers Colorists. **103**, 86–94 (1987)
12. MR Luo, B Rigg, BFD(l:c) color-difference formula. Part II – performance of the formula. J. Soc. Dyers Colorists. **103**, 126–132 (1987)
13. CIE, Industrial colour-difference evaluation. CIE Technical Report 116, (Central Bureau of the CIE , Vienna, Austria, 1995)
14. M Luo, G Cui, B Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000. Color Res. Appli. **26**(5), 340–350 (2001)
15. E Bando, JY Hardeberg, D Connah, in *Human Vision and Electronic Imaging X, ed. B. Rogowitz, T. Pappas, S. Daly, Proc. of SPIE - IST Electronic Imaging,SPIE,* vol. 5666. Can gamut mapping quality be predicted by color image difference formulae (SPIE Bellingham, 2005), pp. 180–191
16. JY Hardeberg, E Bando, M Pedersen, Evaluating image difference metrics for gamut mapped images. Coloration Tech. **124**(4), 243–253 (2008)
17. H Kivinen, M Nuutinen, P Oittinen, in *CGIV 2010 – Fifth European Conference on Colour in Graphics, Imaging*. Comparison of colour difference methods for natural images (Joensuu Finland, 2010), pp. 510–515
18. G Hong, M Luo, in *Proceedings of SPIE: 9th Congress of the International Colour Association* vol. 4421, ed. by Chung R, Rodrigues A. Perceptually based colour difference for complex images (SPIE Bellingham, 2002), pp. 618–621
19. C Oleari, Color opponencies in the system of the uniform color scales of the Optical Society of America. J. Opt. Soc. Am. A. **21**, 677–682 (2004)
20. C Oleari, Hypotheses for chromatic opponency functions and their performance on classical psychophysical data. Color Res Appli. **30**, 31–41 (2005)
21. R Huertas, M Melgosa, C Oleari, Performance of a color-difference formula based on OSA-UCS space using small-medium color differences. J. Opt. Soc. Am. A. **23**(9), 2077–2084 (2006)
22. G Simone, C Oleari, I Farup, in *11th Congress of the International Colour Association (AIC)*. Performance of the Euclidean color-difference formula in log-compressed OSA-UCS space applied to modified image-difference metrics (Sydney Australia, 2009), p. 81
23. DR Pant, I Farup, Riemannian formulation, and comparison of color difference formulas. Color Res. Appl. **37**, 429–440 (2012)
24. N Ponomarenko, V Lukin, K Egiazarian, J Astola, M Carli, F Battisti, in *10th International Workshop on Multimedia Signal Processing*. Color image database for evaluation of image quality metrics (Cairns, Queensland Australia, 2008), pp. 403–408 [http://www.ponomarenko.info/tid2008.htm]. Accessed 20 Feb 2008
25. X Zhang, BA Wandell, A spatial extension of CIELAB for digital color image reproduction. SID J (1997) [21/05/05:http://white.stanford.edu/~brian/scielab/]. Accessed 20 Feb 2008
26. Z Wang, AC Bovik, HR Sheikh, EP Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Proc. **13**(4), 600–612 (2004)
27. Z Wang, A Bovik, A universal image quality index. IEEE Signal Proc. Lett. **9**, 81–84 (2002)
28. K Egiazarian, J Astola, N Ponomarenko, V Lukin, F Battisti, M Carli, in *Proc. of the Second International Workshop on Video Proc. and Quality Metrics*. Two new full-reference quality metrics based on HVS. Scottsdale, Arizona, January 2006.
29. N Nill, A visual model weighted cosine transform for image compression and quality assessment. IEEE Trans. Commun. **33**, 551–557 (1985)
30. N Ponomarenko, F Silvestri, K Egiazarian, M Carli, J Astola, V Lukin, in *CD-ROM Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics*. On between-coefficient contrast masking of DCT basis functions. Scottsdale, Arizona, 25–26 January, 2007 4 p
31. D Chandler, S Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images. IEEE Trans. Image Proc. **16**(9), 2284–2298 (2007)
32. G Simone, C Oleari, I Farup, in *Proceedings from Gjøvik Color Imaging Symposium 2009,* An alternative color difference formula for computing image difference. no. 4 in Høgskolen i Gjøviks rapportserie, Gjøvik Norway, 2009, pp. 8–11
33. GM Johnson, MD Fairchild, in *The 9th Color Imaging Conference: Color Science and Engineering: Systems, Technologies, Applications*. Darwinism of color image difference models. Scottsdale, Arizona, 2001, pp.108–112
34. G Simone, M Pedersen, JY Hardeberg, Measuring perceptual contrast in digital images. J. Vis. Commun. Image Representation. **23**, 491–506 (2012)
35. J Frankle, JJ McCann, Method and apparatus for lightness imaging. United States Patent No. 4,384,336 (1983)
36. EH Adelson, CH Anderson, JR Bergen, PJ Burt, JM Ogden, Pyramid methods in image processing. RCA Eng. **29**(6) (1984)
37. C Enroth-Cugell, JG Robson, The contrast sensitivity of retinal ganglion cells of the cat. J. Phys. **187**, 517552 (1966)
38. G Sclar, JH Maunsell, P Lennie, Coding of image contrast in central visual pathways of the macaque monkey. Vis. Res. **30**, 1–10 (1990)
39. R Rodieck, Quantitative analysis of cat retinal ganglion cell response to visual stimuli. Vis. Res. **5**(12), 583–601 (1965)
40. A Michelson, *Studies in Optics*. (University of Chicago Press, Chicago, 1927)
41. M Pedersen, JY Hardeberg, in *CGIV 2008 Fourth European Conference on Color in Graphics, Imaging and Vision*. Rank order and image difference metrics (Terrassa Spain: IS&T, 2008), pp. 120–125
42. S Ajagamelle, Analysis of the difference of Gaussians model in perceptual image difference metrics. *Master's thesis*, Gjøvik University College and Grenoble Institute of Technology (2009)
43. J Dijk, In search of an objective measure for the perceptual quality of printed images. *PhD thesis*. Technische Universiteit Delft, 2004
44. 12640-3, I: ISO 12640–3:2007 graphic technology - prepress digital data exchange - Part 3: CIELAB standard colour image data CIELAB/SCID, International Organization for Standardization 2007. [http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=52497] Accessed 20 Feb 2008
45. MG Kendall, A Stuart, JK Ord, *Kendall's Advanced Theory of Statistics: Classical Inference and Relationship,* vol. 2, 5th edn. (Hodder Arnold Publisher, London, 1991)

46. QualityExperts Video, Objective perceptual assessment of video quality: full reference television. Tech. rep. (International Telecommunication Union,Geneva, 2004)
47. HR Sheikh, MF Sabir, AC Bovik, A statistical evaluation of recent full reference image quality assessment algorithms. IEEE Trans. Image Proc. **15**(11), 3440–3451 (2006)
48. Video Quality Experts Group, Final report from the video quality experts group: validation of reduced-reference and no-reference objective models for standard definition television. Phase I. Tech. rep. (International Telecommunication Union, Geneva, 2009)

doi:10.1186/1687-5281-2013-39
Cite this article as: Simone *et al.*: Multi-level contrast filtering in image difference metrics. *EURASIP Journal on Image and Video Processing* 2013 **2013**:39.

10.4 Paper D: Spatio–temporal Retinex–like envelope with total variation

Spatio-Temporal Retinex-like Envelope with Total Variation

Gabriele Simone and Ivar Farup
Gjøvik University College; Gjøvik, Norway.

Abstract

Many algorithms for spatial color correction of digital images have been proposed in the past. Some of the most recently developed algorithms use stochastic sampling of the image in order to obtain maximum and minimum envelope functions. The envelopes are in turn used to guide the color adjustment of the entire image. In this paper, we propose to use a variational method instead of the stochastic sampling to compute the envelopes. A numerical scheme for solving the variational equations is outlined, and we conclude that the variational approach is computationally more efficient than using stochastic sampling. A perceptual experiment with 20 observers and 13 images is carried out in order to evaluate the quality of the resulting images with the two approaches. There is no significant difference between the variational approach and the stochastic sampling when it comes to overall image quality as judged by the observers. However, the observed level of noise in the images is significantly reduced by the variational approach.

Introduction

A great amount of research has been done on Human Visual System (HVS), which is quite difficult to mimick as the HVS has complex and robust mechanisms to acquire useful informations from the physical environment. In particular the color of an area in a visual scene is heavily influenced by the chromatic content of the other areas of the scene. This psychophysiological phenomenon is known as locality of color perception.

One of the earliest models able to deal with locality of perception is Retinex, proposed by Land and McCann in 1971 [14], which is an image processing method that exhibits some behaviors similar to the HVS. The scientific community has continued to be interested in this model and its various applications, as reported in [17, 16]. In the basic Land and McCann implementation of Retinex, locality is achieved by long paths scanning across images. Different implementations and analysis followed after this first work. These can be divided into two major groups, and they differ in the way they achieve locality. The first group explore the image using paths or computing ratios with neighbors in a multilevel framework [7, 13, 15, 21, 8, 4]. and recent approaches work using in particular Brownian motions models [6, 18]. The second group computes values over the image with convolution mask or weighting distances [11, 1, 10, 5, 20].

A recent implementation, constructed to investigate the effects of different spatial samplings, replaces paths with random sprays, i.e. two-dimensional point distributions across the image, hence the name "Random Spray Retinex" (RSR) [19]. In a follow-up, Kolås et al. [12] developed the "Spatio-Temporal Retinex-like Envelope with Stochastic Sampling" (STRESS) framework, where the random sprays are used to calculate two envelope functions representing the local reference black and white points. Both algorithms need a high density of samples in order to lower the amount of noise, but they never sample the whole image in order to keep a local effect. Furthermore the number of sampling points needed increases drastically when increasing the image size and consequently also the computational time.

In this work, we propose and test an alternative method for calculating the two envelope functions of STRESS, replacing the the stochasting sampling with a constrained total variation method. We want to emphasize that although much of idea is the same, it is not just another implementation of STRESS as the two algorithms follow a different strategy for calculating the envelopes and they show different behaviors.

In order to give to the reader a complete and detailed overview, STRESS will be described in the next section, followed by our new proposal. Afterwards, a description of the method of evaluation of our proposal in addition to some implementation details are presented. Finally, experimental results are shown and conclusions are drawn.

The STRESS Algorithm

The STRESS (Spatio-Temporal Retinex-like Envelope with Stochastic Sampling) algorithm developed by Kolås et al. [12] aims to reproduce some of the adjustment mechanisms typical for the Human Visual System. The central part of the STRESS algorithm is to calculate, for each pixel, the local reference black and white points in each chromatic channel. This is done through calculating two envelope functions, the maximum and minimum envelopes, containing the image signal. The two envelopes, denoted as $E_{\max}$ and $E_{\min}$, are slowly varying functions, such that the image signal is always in between the envelopes or equal to one of them. In particular the two envelopes should have the following characteristics: 1) following the signal; 2) being smooth; 3) being edge preserving; 4) touching the global maximum of the image for $E_{\max}$, while the global minimum for $E_{\min}$.

For each pixel p_0, the two envelopes are estimated using a random spray modeled as follows:

$$E_{\min} = p_0 - \overline{v}\,\overline{r}, \tag{1a}$$

$$E_{\max} = p_0 + (1 - \overline{v}\,\overline{r}) = E_{\min} + \overline{r} \tag{1b}$$

where:

$$\overline{r} = \frac{1}{N}\sum_{i=1}^{N} r_i, \tag{2a}$$

$$\overline{v} = \frac{1}{N}\sum_{i=1}^{N} v_i \tag{2b}$$

N denotes the number of iterations, while r_i is the range of the samples and v_i the relative value of the center pixel given as:

$$r_i = s_i^{\max} - s_i^{\min}, \tag{3a}$$

$$v_i = \begin{cases} \frac{1}{2} & \text{if } r_i = 0 \\ \frac{p_0 - s_i^{\min}}{r_i} & \text{else} \end{cases} \tag{3b}$$

$s_i^{\max}$ and $s_i^{\min}$ are the maximum and minimum samples, found as:

$$s_i^{\max} = \max_{j \in \{1,2,...,M\}} p_j \quad , \tag{4a}$$

$$s_i^{\min} = \min_{j \in \{1,2,...,M\}} p_j \tag{4b}$$

where M is the number of samples and p_i is the pixel at iteration i.

Given the two envelopes, each pixel p_0 is adjusted as follows:

$$p_{stress} = \frac{p_0 - E_{\min}}{E_{\max} - E_{\min}} \tag{5}$$

In this way STRESS aims to enhance the contrast of the image, giving highlight/emphasis to details and balance the three channels of the image, thus performing a color correction.

Our Proposal: The STRETV Algorithm

The main feature of the STRESS algorithm is the calculation of the envelopes $E_{\max}$ and $E_{\min}$ for each channel. In our proposal the stochastic sample is replaced with the total variation method for calculating the two envelopes. The following equation describes our model, called *STRETV* (Spatio-Temporal Retinex-like Envelope with Total Variation):

$$\text{minimize } TV = \text{minimize} \int_\Omega |\nabla E| \, d\Omega + \frac{\lambda}{2} \int_\Omega |E - I|^2 \, d\Omega \tag{6}$$

where I is the original image channel, E is the maximum or minimum envelope, Ω is the domain of the image, $\int_\Omega |\nabla E| \, d\Omega$ is the total variational term, and $\frac{\lambda}{2} \int_\Omega |E - I|^2 \, d\Omega$ is the non-smooth fidelity term with λ weighting factor for the data attachment.

This minimization is subject to the three following constraints: following the signal, $E_{\max} \geq I$ and $E_{\min} \leq I$. The corresponding Euler-Lagrange equation is as follows:

$$\nabla \cdot \left(\frac{\nabla E}{\| \nabla E \|} \right) - \lambda (E - I) = 0 \tag{7}$$

where I is the original image channel, E is the maximum and minimum envelope, $\nabla \cdot \left(\frac{\nabla E}{\|\nabla E\|} \right)$ is the driving force to the smoothness, and $\lambda(E-I)$ is the driving force to the data attachment with $0 < \lambda \ll 1$.

The solution is computed using an Euler explicit time marching scheme for each color channel:

$$\nabla \cdot \left(\frac{\nabla E}{\| \nabla E \|} \right) - \lambda (E - I) = \frac{\partial E}{\partial t} \tag{8}$$

applying Neumann boundary conditions and the regularization proposed by Blomgren and Chan [2] at each iteration. At end of each iteration the constraints are enforced in order to avoid numerical errors.

Results

Preliminary tests indicate that STRETV works well with $\lambda = 0.1$ and $\lambda = 0.001$, using a unit time step ($\Delta t = 1$). The algorithm has been implemented in Matlab (using Parallel Computing Toolbox) and tested on a DELL Latitude Model E6520.

In order to evaluate the quality of STRETV, two perceptual experiments have been carried out. A set of 13 images chosen following the recommendations from [9, 3] were evaluated in a pairwise comparison on neutral grey background to a total of 20 observers. In the first perceptual experiment each STRETV image was compared to its original and the observers were asked to choose the image based on their preference. In the second perceptual experiment each STRETV image was compared to its relative processed with STRESS and in a first round the observers were asked to choose the image based on their preference while in a second round they were asked to pick the image with more noise. Whenever the STRETV image did not succeed in the first experiment, it was discarded from the second experiment.

Figure 1-2 shows the 13 original images in the middle, the corresponding processed by STRETV on the right and the corresponding processed by STRESS on the left. Due to page limitations, it is not possible to analyze all the images but particular interesting results for STRETV can be found in Figures 1(c), 2(l), 2(r) where the reflections present in the water, windows and eyeglasses respectively are more visible and in Figure 2(o) where the red component is corrected and identification of objects in the map are perceptually easier.

Figure 3 shows the preference of the 20 observers on the tested images for the first psychophysical experiment and we can clearly see that STRETV succeeds on 11 of 13 images with six of them with a preference equal or greater than 85% of the rates. The reason of the defeat for Figure 1(i) and of the draw for Figure 2(f) with respect to their original can be found in the fact that observers perceived a loss of naturalness.

Figure 4 shows the preference of the 20 observers for the second psychophysical experiment, where STRETV was compared to STRESS. Figure 5 shows the perceived noise of the 20 observers for STRETV and STRESS. STRETV is preferred for 7 of the 11 images and it results to be less noisy for 9 of the 11 images.

Furthermore, STRETV has lower computational complexity than STRESS, $O(N \cdot n)$ against $O(N \cdot M \cdot n)$, where N is the number of iterations, n is the number of pixels in the image and M is the number of samples. As consequence STRETV is faster than STRESS implemented in MATLAB. On the other hand a new implementation of STRESS in CUDA provided by the original authors [12] is more efficient than STRETV.

A sign-test at 95% confident interval shows that STRETV is significantly better than the original and having the same performance of STRESS but producing images less noisy.

Conclusions

We have developed a new Retinex algorithm called STRETV (Spatio-Temporal Retinex-like Envelope with Total Variation), following the approach of Kolås et al. [12] for the STRESS algorithm (Spatio-Temporal Retinex-like Envelope with Stochastic Sampling), which adjusts each pixel calculating the local reference of lighter and darker points in each channel. This is done estimating two envelope functions, the maximum and minimum envelopes, containing the image signal, through the total variation method.

STRETV shows promising results in contrast enhancement and automatic color correction. A first psychophysical experiment on 13 images with 20 observers confirm the efficiency of the method with a noticeable success of STRETV on 11 images in comparison to the original. A second psychophysical experiment shows that the overall performance of STRESS is not significantly different at 95% confidence level. At the same time a higher preference of the observers of STRETV with respect to STRESS is noticed due to a lower perception of noise.

Future work can consist of extending STRETV to high dynamic range imaging rendering and color-to-grey conversion to

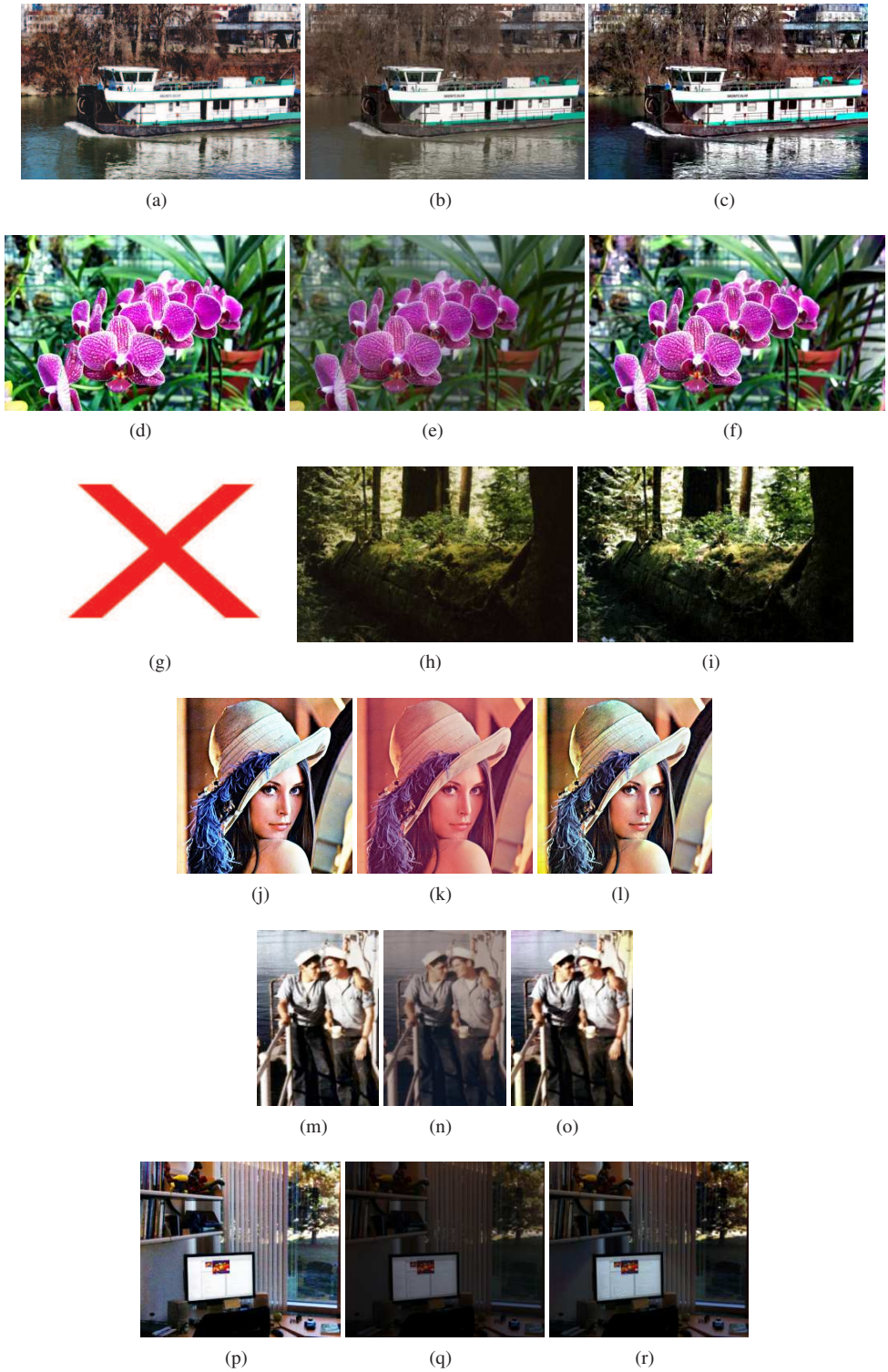


Figure 1. *STRESS on the left, original in the middle, STRETV on the right.*

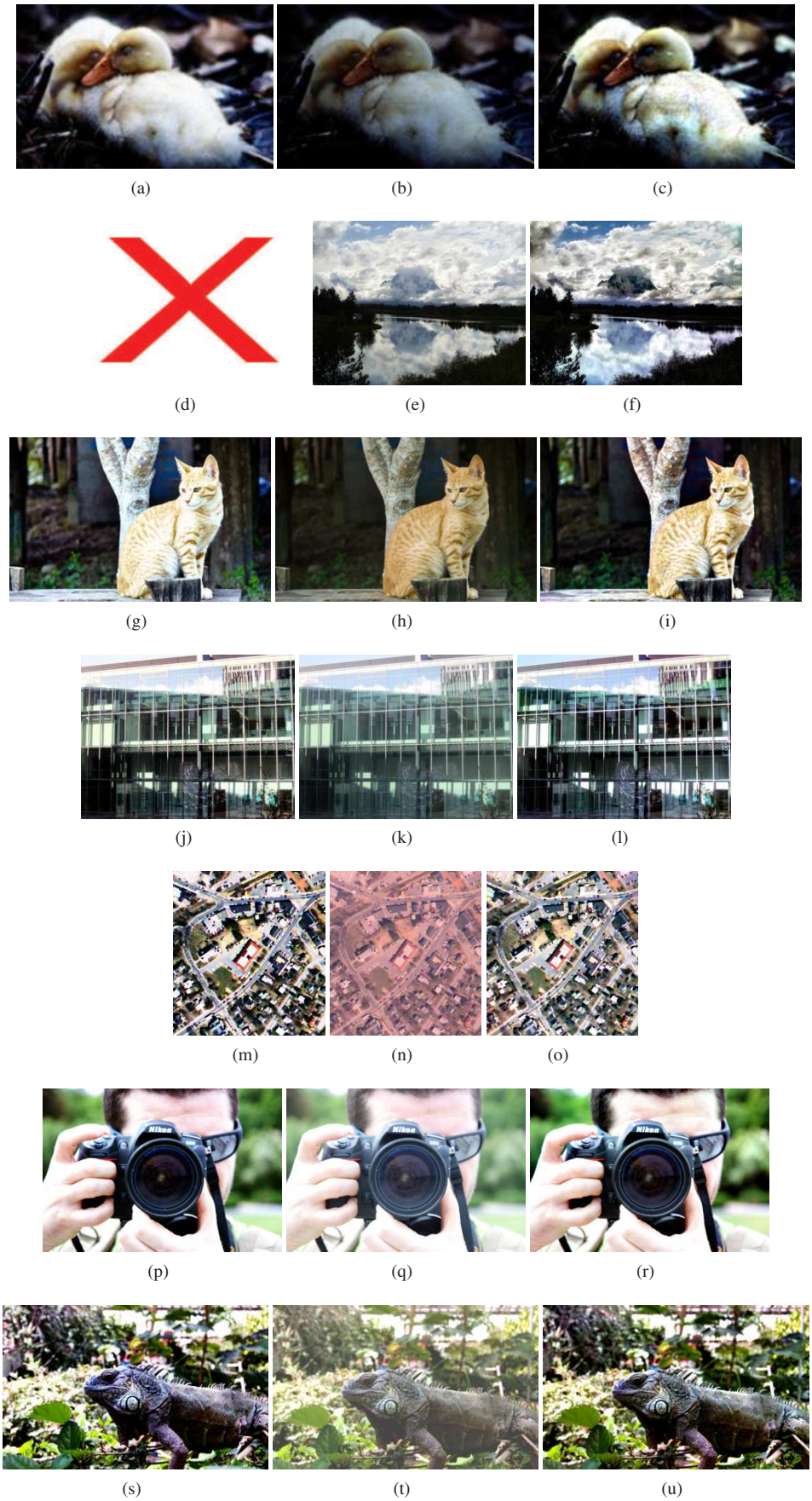


Figure 2. *STRESS on the left, original in the middle, STRETV on the right.*

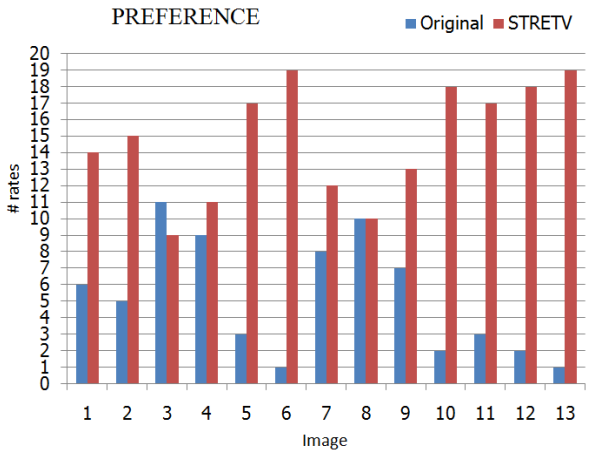


Figure 3. *Observer's preference of STRETV with respect to its original on the 13 tested images.*

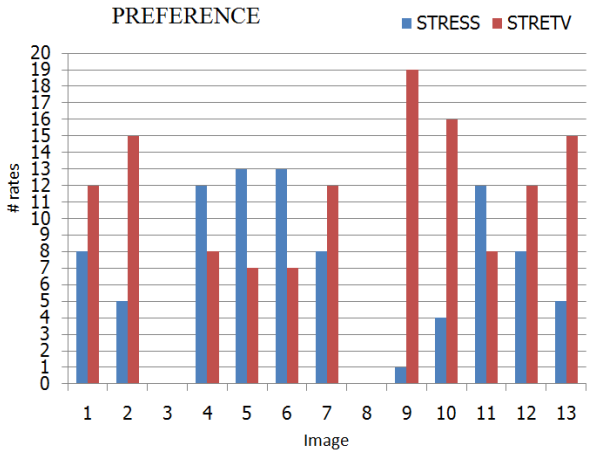


Figure 4. *Observer's preference of STRETV in respect of STRESS on the 11 tested images.*

be fully comparable with the STRESS and other spatial color algorithms.

Acknowledgment

This work has been supported by NFR over the SHP project. The authors would like to thank Fritz Albregtsen (University of Oslo) for his useful feedbacks and suggestions.

References

[1] K. Barnard and B. Funt. Investigations into multi-scale Retinex. In *Color Imaging in Multimedia*, pages 9–17. Technology, Wiley, 1998.

[2] P. Blomgren and T. F. Chan. Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, (7):304–309, 1998.

[3] CIE. Guidelines for the evaluation of gamut mapping algorithms. Technical Report ISBN: 3-901-906-26-6, CIE TC8-08, 156:2004.

[4] T. J. Cooper and F. A. Baqai. Analysis and extensions of the frankle-mccann Retinex algorithm. *Journal of Electronic Imaging*, 13(1):85–92, 2004.

[5] J. D. Cowan and P. C. Bressloff. *Visual cortex and the Retinex algorithm*, volume SPIE-4662, pages 278–285. The International Society for Optical Engineering, 2002.

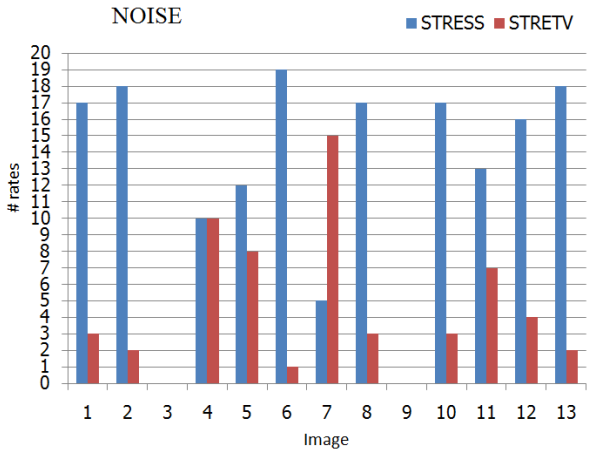


Figure 5. *Observer's perceived noise of STRETV and STRESS on the 11 tested images.*

[6] G. Finlayson, S. Hordley, and M. Drew. Removing shadows from images using Retinex. In *The Tenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, pages 73–79, Scottsdale, AZ, USA, November 2002. IS&T/SID.

[7] J. Frankle and J. McCann. Method and apparatus for lightness imaging. United States Patent No. 4,384,336, 1983.

[8] B. Funt, F. Ciurea, and J. J. McCann. Retinex in MATLAB. *Journal of Electronic Imaging*, 13(1):48–57, January 2004.

[9] J. Holm, I. Tastl, and T. Johnson. Definition & use of the iso 12640-3 reference color gamut. In *Fourteenth Color Imaging Conference: Color Science and Engineering Systems, Technologies, Applications*, pages 62–68, Scottsdale, AZ, 2006. IS&T/SID.

[10] F. O. Huck, C. L. Fales, R. E. Davis, and R. Alter-Gartenberg. Visual communication with Retinex coding. *Applied Optics*, 39(11):1711–1730, 2000.

[11] D. J. Jobson, Z. Rahman, and G. A. Woodell. Properties and performance of a center/surround Retinex. *IEEE Transactions on Image Processing*, 6(3):451–462, 1997.

[12] Ø. Kolås, I. Farup, and A. Rizzi. STRESS: A framework for spatial color algorithms. *Journal of Imaging Science and Technology*, 55(4):040503, 2011.

[13] E. H. Land. The Retinex theory of color vision. *Scientific American*, 237:108–128, 1977.

[14] E. H. Land and J. J. McCann. Lightness and Retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, jan 1971.

[15] D. Marini and A. Rizzi. A computational approach to color adaptation effects. *Image and Vision Computing*, 18:1005–1014, 2000.

[16] J. McCann and A. Rizzi. *The Art and Science of HDR Imaging*. John Wiley, 2011. ISBN: 978-0-470-66622-7.

[17] J. J. McCann. Guest editorial: Special section on Retinex at 40. *Journal of Electronic Imaging*, 13(1):6–7, 2004.

[18] R. Montagna and G. D. Finlayson. Constrained pseudo-brownian motion and its application to image enhancement. *Journal of the Optical Society of America A*, 28(8):1677–1688, August 2011.

[19] E. Provenzi, M. Fierro, A. Rizzi, L. D. 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(1):162–171, January 2007.

[20] Z. Rahman and G. A. Woodell. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13:100–110, 2004.

[21] A. Rizzi, D. Marini, and L. D. Carli. LUT and multilevel brownian Retinex colour correction. *Machine Graphics & Vision*, 11(1):153–169, 2002.

Author Biography

Gabriele Simone *received his BiT in 2005, and his MSIT in 2007 both at University of Milan - Department of Information Technology, Italy. He is currently pursuing a PhD in Color Imaging. He is a member of the Norwegian Color Research Laboratory at Gjøvik University College and his main research topic is contrast measure, image difference metrics, and tone mapping algorithms in HDR images.*

Ivar Farup *received the M.Sc. degree in physics from the Norwegian University of Science and Technology, Trondheim, Norway, in 1994, and the Ph.D. degree in applied mathematics from the University of Oslo, Oslo, Norway, in 2000. Since 2000, he has been an Associate Professor at Gjøvik University College, Gjøvik, Norway, mainly focusing on color imaging.*

10.5 Paper E: Termite Retinex: A new implementation based on a colony of intelligent agents

Journal of Electronic Imaging

SPIEDigitalLibrary.org/jei

Termite Retinex: a new implementation based on a colony of intelligent agents

Gabriele Simone
Giuseppe Audino
Ivar Farup
Fritz Albregtsen
Alessandro Rizzi

SPIE

IS&T
imaging.org

Journal of Electronic Imaging 23(1), 013006 (January 2014)

Termite Retinex: a new implementation based on a colony of intelligent agents

Gabriele Simone,[a,*] Giuseppe Audino,[b] Ivar Farup,[a] Fritz Albregtsen,[c] and Alessandro Rizzi[b]
[a]Gjøvik University College, Faculty of Computer Science and Media Technology, The Norwegian Colour and Visual Computing Laboratory, P.O. Box 191, N-2802 Gjøvik, Norway
[b]Universitá degli Studi di Milano, Dipartimento di Informatica e Comunicazione, I-20135 Milano, Italy
[c]University of Oslo, Department of Informatics, P.O. Box 1080 Blindern, N-0316 Oslo, Norway

Abstract. The original presentation of Retinex, a spatial color correction and image enhancement algorithm modeling the human vision system, as proposed by Land and McCann in 1964, uses paths to explore the image in search of a local reference white point. The interesting results of this algorithm have led to the development of many versions of Retinex. They follow the same principle but differ in the way they explore the image, with, for example, random paths, random samples, convolution masks, and variational formulations. We propose an alternative way to explore local properties of Retinex, replacing random paths by traces of a specialized swarm of termites. In presenting the spatial characteristics of the proposed method, we discuss differences in path exploration with other Retinex implementations. Experiments, results, and comparisons are presented to test the efficacy of the proposed Retinex implementation. [DOI: 10.1117/1.JEI.23.1.013006]

Keywords: Retinex; ant colony system; spatial color algorithm; image enhancement.

Paper 12537P received Dec. 22, 2012; revised manuscript received Nov. 28, 2013; accepted for publication Dec. 11, 2013; published online Jan. 17, 2014.

1 Introduction

During the past decades a significant amount of research has been undertaken to understand human visual perception, which is not a trivial task as the human visual system (HVS) has complex and robust mechanisms to acquire useful information from the environment. In particular, the color appearance of an area is influenced by the chromatic content of the other areas of the scene. This psychophysiological phenomenon is referred to as locality of color perception.

In order to deal with this locality in image appearance, different image processing methods and frameworks have been developed with the intent to exhibit behaviors similar to the HVS, such as Automatic Color Equalization (ACE),[1] Spatio-Temporal Retinex-inspired (STRESS),[2] image Color Appearance Model (iCAM), and its evolutions,[3,4] and the various Retinex implementations, which are the interest of this work.

The original Retinex model of lightness and color perception was proposed by Land and McCann about 50 years ago.[5,6] In the light of current developments, there has been a renewed interest in color appearance.[7,8] The principal assumption of the Retinex theory is that the sensation of color comes from a comparison among the various areas in the scene. In this comparison performed separately on each of the three color channels, a series of ratios and products among the various parts of the image are computed to calculate each pixel appearance. In this way, the appearance of each pixel depends on what surrounds the pixel; therefore, pixels of the same value can trigger different sensations. This locality of perception is achieved, in the original proposal from Land and McCann, by long paths scanning across the image, accounting for pixel ratio computation in each chromatic channel. The details of the ratio-product-reset Retinex algorithm will be presented later in the paper.

Different implementations and analysis have been developed since this first work, and these can be divided into three major groups, which differ in the way they achieve locality. The first group explores the image using paths or extracting random pixels around the pixel of interest or computing ratios with neighbors in a multilevel framework.[9–11] The second group instead computes values over the image with convolution masks or weighting distances.[12,13] The third group uses differential mathematical techniques based on Poisson-type equations and variational approaches (e.g., Refs. 14, 15, 16 and 17).

Recent implementations constructed to investigate the effects of different spatial samplings replace paths with random sprays, for example, two-dimensional point distributions across the image, hence the name Random Spray Retinex (RSR).[18] All these algorithms require a high density of samples in order to lower the amount of noise, but they never sample the whole image in order to keep the local effect. Furthermore, the number of sampling points that are needed increases drastically and consequently also the computational time, when increasing the size of the image.

In this work we start from the random path approach of the first group. In particular, we focus on the Brownian motion models.[10,19] Here, the idea of the paths is implemented using a swarm intelligence model. Swarm intelligence can be considered a branch of artificial intelligence techniques dealing with the modeling of the collective behavior of simple agents interacting with the environment and among themselves. These models, in general physically or biologically inspired, provide metaheuristics for a wide set of

*Address all correspondence to: Gabriele Simone, E-mail: gabriele.simone@hig.no

combinatorial optimization problems. There are several families of swarm intelligence, and among them is the ant colony system (ACS),[20] upon which we base our work.

The first studies on the development of artificial ants date back to the 1990s and is based on the work of Dorigo et al.[20] The idea is derived from the behavior of real ants foraging for food. In this context, the objective is to understand how almost blind animals, such as ants, can manage to establish the shortest path routes from their colony to feeding sources and back. Biological studies have shown that social insects coordinate their activities via stigmergy, a form of indirect communication mediated by modifications of the environment. In the case of ants, the communication among individuals, or between individuals and the environment, is based on the production of a chemical substance called pheromone. Inspired by these behavioral properties of the ants, Dorigo et al. developed the so-called ACS for solving computational problems. Today their pioneering work has been diversified to solve a wider class of numerical problems.[21]

In this work, we propose a new implementation of Retinex, following the approach of the first group, in particular substituting the Brownian paths with a so-called termite colony exploration of the image, a diversified and tuned model for image processing derived from the ACS for traveling salesman problem (TSP). In daily life, termites are also known as white ants, and as this model attempts an eager exploration in search of the local reference white, we have called our algorithm Termite Retinex (TR).

The rest of this paper is organized as follows. Section 2 briefly describes the ACS system, followed by our proposal TR in Sec. 3. Section 4 presents the method of evaluation, and next, the results are presented and discussed in Sec. 6. Finally, in Sec. 7, conclusions are drawn.

2 Ant Colony System Model

The basic motivation of the presented work is to implement an alternative way to explore the image, driven by mechanisms different from simple randomness. To this aim, we have chosen a swarm intelligence method. Many of them can be suitable for this purpose; however, in this paper, we are more interested in testing if this approach can be successful rather than choosing the best swarm intelligence method. For this reason, we have chosen the ACS since it is one of the first and for its simplicity.

The following is a brief introduction to the first ACS model proposed by Dorigo et al.[20] and the TSP since this is relevant for our proposal. The TSP is an NP-hard problem in combinatorial optimization and theoretical computer science, where given a list of cities and their pairwise distances, the task is to find the shortest possible tour that visits each city exactly once. ACS has been shown to be able to converge to the optimal solution of instances of this problem with short computational time, when cities are on a plane, and a path (edge) exists between each pair of cities (i.e., the TSP graph is completely connected).[22]

Three ideas from natural ant behavior are transferred to the artificial ant colony.

1. The preference for paths with a high pheromone level,
2. The higher rate of growth of the amount of pheromone on shorter paths,
3. The trail mediated communication among ants.

An artificial ant k in city r chooses the city s to move to among those that do not belong to its working memory M_k by applying the following probabilistic formula:[23]

$$p_k(r,s) = \begin{cases} \frac{(\tau_{r,s})^{\alpha}(\eta_{r,s})^{\beta}}{\sum_{u \notin M_k}(\tau_{r,u})^{\alpha}(\eta_{r,u})^{\beta}} & \text{if } s \notin M_k \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $\tau_{r,u}$ is the amount of pheromone trail on edge (r,u), $\eta_{r,u}$ is a heuristic function called visibility, which is the inverse of the distance between cities r and u, and α and β are parameters that allow a user to control the importance of the trail versus the visibility. The memory M_k is the tabu list of the k'th ant, which contains the cities that it has already visited. City s is inserted in the list when the ant transits from city r to city s. The choice criteria of the parameters α and β can differ widely according to the problem for which the ACS is used. A guideline on how to choose the values of the different parameters for the TSP problem can be found in Ref. 22.

3 Termite Retinex

The original Retinex model developed by Land and McCann[5,6] is based on computing the relative channel lightness (L) at a point i as the mean value of the relative channel lightnesses (l) computed along N random paths from point j to the point i (Fig. 1).

$$L^i = \frac{\sum_{h=1}^{N} l_h^{i,j}}{N}, \tag{2}$$

where

$$l_h^{i,j} = \sum_{x=j}^{i} \delta \log\left(\frac{I_{x+1 \in \text{path}}}{I_{x \in \text{path}}}\right), \tag{3}$$

where I_x is the lightness intensity of the pixel x, I_{x+1} is the lightness intensity of the pixel $x+1$, h is indicating the path, and δ represents a threshold mechanism.

$$\delta = \begin{cases} 1 & \text{if} \left| \log\left(\frac{I_{x+1 \in \text{path}}}{I_{x \in \text{path}}}\right) \right| > \text{Threshold} \\ 0 & \text{if} \left| \log\left(\frac{I_{x+1 \in \text{path}}}{I_{x \in \text{path}}}\right) \right| \leq \text{Threshold} \end{cases}. \tag{4}$$

The threshold mechanism helps to discount slow varying gradients; however, it has been proven that the algorithm maintains its fundamental properties even without it.[24] During the computation of l along the path h, the value

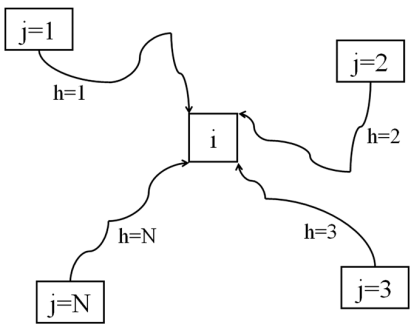


Fig. 1 N random paths from point j to the point i.

of l can become positive. In terms of image, this means that a lighter area is found. In this case, a reset mechanism triggers, forcing the relative lightness to the unitary value, making the average computation restart from this area.

The reset mechanism forces to restart the chain of ratios (sum of logs) from the point of reset by considering the lightness value found at the reset point as the new local reference white. The point i can be either along the path, like in the original Retinex, or always at the end of the path, like in the Milano (MI)-Retinex family. This difference is discussed in Ref. 8. We have chosen to implement the Termites using the MI-Retinex approach. This results in higher computational costs but is an easier way to test and control the Termite swarm and the overall behavior.

A Brownian motion method is used to generate these paths,[10,19] which has been extensively studied from a theoretical point of view. Brainard and Wandell[25] described them formally using stochastic methods: for a large number of very long paths, the method converges; for infinitely long paths, the image tends to the original image globally normalized by its channel maximum values. This limit behavior has no interest since it loses the local behavior. Brownian paths work efficiently for a path-based Retinex (Fig. 2), but the approach has some drawbacks. Its computational complexity is low, $O(N \log N)$, where N is the number of pixels, but processing a whole image requires a large amount of memory. Thus, performing as many visits as desirable for each pixel becomes unfeasible when the image is too large.

When attempting to model the human vision system, an important characteristic to consider is the locality of the visual sensation.[8,26] The way the image is explored affects the influence that the surrounding pixels have on the final estimate of the visual sensation. The main objective of the use of a colony of termites is to model a novel method by which the image content is explored, and thus the locality of perception, with an alternative mechanism. As consequence, a new set of parameters is created to describe and investigate the local behavior of Retinex.

In order to create TR, ACS requires some modifications, which comprises the following assumptions and constraints:

1. Pixels are considered as cities: a termite can choose to move only to one of the eight neighboring pixels (no jumps). Exploring larger neighborhoods and allowing the termite to jump might lead to not discovering the proper reference white.
2. The choice of a pixel is based on its distance and intensity value: the visibility η is substituted with the bilateral distance c as defined below, which we will refer to as closeness.
3. There is a higher preference for paths with a low poison, as we want divergence, in order to explore different areas of the image: the poison is the inverse to the amount of pheromone: $\theta = 1/\tau$, which acts as a repulsion force, inverse of the attraction force, pheromone in ACS.

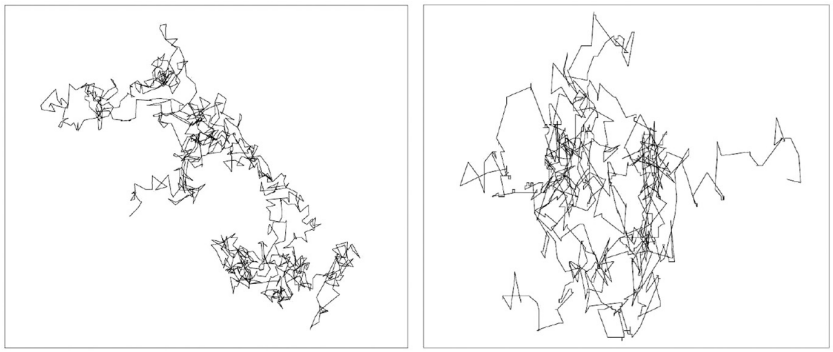
So in our modified model, an artificial termite k in pixel r chooses to move to the pixel s among those that belong to the eight-neighborhood N_8 and that do not currently belong to its working memory M_k by applying the following probabilistic formula:

$$p_k(r,s) = \begin{cases} \dfrac{(\theta_s)^\alpha (c_{r,s})^\beta}{\sum_{u \notin M_k \text{ and } u \in N_8} (\theta_u)^\alpha (c_{r,u})^\beta} & \text{if } s \notin M_k \text{ and } s \in N_8 \\ 0 & \text{otherwise} \end{cases}, \tag{5}$$

where θ_u is the amount of poison on pixel u, $c_{r,u}$ is the bilateral distance between pixels r and u, and α and β are parameters weighting the importance of the poison versus the closeness, which is directly related to the brightness of the pixel. In that case, all the surrounding pixels have the same probability, one pixel is drawn randomly with uniform probability. In our model, the memory M_k is the tabu list of the k'th termite, which contains the coordinates of the pixels that the termite has already visited. This list is updated by inserting the coordinates of pixel s when the termite transits from pixel r to pixel s. Reminding that the poison is the inverse of the amount of pheromone, once a termite has transited on pixel u, the quantity of poison on pixel u is updated as follows:

$$\tau_u = \tau_u + Q, \tag{6}$$

$$\theta_u = \frac{1}{\tau_u}, \tag{7}$$



(a) (b)

Fig. 2 Traditional Brownian motion investigation on the left, pseudo-Brownian motion investigation proposed by Montagna and Finlayson[19] on the right. Images provided with courtesy by Montagna and Finlayson.[19]

where Q is a chosen amount of poison, according to how strong a divergence is desired. As in Retinex model, each pixel is treated independently; after processing a pixel, the amount of poison at all pixels must be reset to the initial value of 1 before starting to process a new pixel.

We underline that Eq. (5) uses the same convex combination as Eq. (1) but with a different leading principle: a trade-off between the brightness of a pixel to choose and the quantity of poison found on the pixel. In the case of high values of β, a pixel is chosen, even if already visited by another termite. On the contrary, high values of α force a termite to choose another direction.

The bilateral distance $c_{r,u}$ is defined as follows:

$$c_{r,u} = \frac{d_e + d_v}{\sqrt{2}}, \tag{8}$$

$$d_e = \sqrt{(x_r - x_u)^2 + (y_r - y_u)^2}, \tag{9}$$

$$d_v = |I(x_r, y_r) - I(x_u, y_u)|, \tag{10}$$

where d_e and d_v are the distance in coordinates and in intensity values, respectively, I is the image channel, and (x, y) are the coordinates of the pixels.

The use of the bilateral distance is known to be a suitable tool for edge-preserving.[27] In our model, it is possible to notice from Eq. (8) that when a termite has to choose among eight neighboring pixels with same intensity values, the preference goes to one of the four diagonal pixels with respect to the vertical and horizontal ones. This scheme is motivated and justified by the fact that the presence of halos are reduced in the final image with respect to the simple distance based only in intensity values [Eq. (10)]. An example is shown Fig. 3. This phenomenon of preference for the diagonal pixel gradually disappears with a high number of termites and with the participation of the poison.

4 Algorithm Characteristics

4.1 *Parameters Tuning*

In the TSP problem, all the metaheuristics attempt to find the optimal solution. In the field of spatial color algorithms (SCAs),[26] the optimal solution depends on the task of the algorithm. In the work that we are presenting, the goal of the filtering is a qualitative emulation of the HVS for an unsupervised image enhancement.

TR is a novel method to explore the image content alternative to Brownian randomness. Tuning the TR parameters properly can make the termites' paths similar to Brownian method; however, regarding this similarity, we would like to highlight two important points.

1. TR is an MI-Retinex implementation that differs from the original Retinex and also from the version by Montagna and Finlayson[19] and does not need to travel across the whole set of pixels in the image.
2. The poison is a feature that prevents the paths from a complete randomness since it affects locality and can be used as a tuning parameter in order to obtain different spatial effects.

In the case of TR, choosing the number of termites and the length of their path is a trade-off between the computational cost of the algorithm and the locality of the filtering. Introducing too many termites and paths that are too long

(a) Original

(b) Only intensity values

(c) With bilateral distance

Fig. 3 Original on the top, halos appearing using only the simple distance on intensity values in the bottom-left, using bilateral distance in the bottom-right.

may lead to a higher computational cost. Moreover, as we are interested in finding a local reference white and not the global white of the image, a termite should never touch all the pixels of the image. On the other side, paths that are too short can be a constraint on the locality, and too few termites due to insufficient image sampling can lead to a high chromatic noise. Examples of results using different spatial investigations are shown in Figs. 4 and 5. In general, a large number of termites traveling for a long path (number of pixels N_s) is needed in order to investigate a wide neighborhood for searching the local reference white. In the examples of Figs. 4 and 5, it can be noticed that overincreasing $k \cdot Ns$ results in a loss of the local effect, as evidenced by the light background in the details bottom right that is darker than the one above and the one on the left, both with a lower sampling.

The poison θ and the weight α, which control the divergence of the swarm, are also essential for investigation: low values of θ strongly enforce termites to choose different directions from each other, while high values of θ make termites converge. An example of the behavior of the swarm with different values of α and β with constant quantity of poison $Q = 1$ is shown in Fig. 6, while an example of constant values of α and β and increasing quantity of poison Q in steps of a factor of 10 from 10^{-4} to 1 is shown in Fig. 7. A preliminary test among few people was run just too quickly preset the Q value. The presence versus the absence of halos was the chosen criterion that led to the use of value 1. In fact, in Fig. 7, it is easy to notice that low values of Q introduce halos in the filtered image.

A rule of thumb derived from a small sampling from a previous investigation suggests that $N_s \approx 70\%$ of the length of the diagonal of the image results in a good score in observers' image quality preferences.[28,29] Regarding α, β, and Q, the suggested starting values are 0.1, 0.9, and 1, respectively. We underline to the reader that the parameters α, β, and Q do not change during pixel recomputation; they are kept constant for the whole processing of the image.

As a result of more comprehensive set of experiments, we can confirm this rule of thumb as a starting value for the parameter setting. Since the choice of parameters is affected by the image content, a finer tuning could be necessary. We remind that these settings mean that the poison θ has very low importance, while the closeness c has very high importance. This causes a termite to simply choose a pixel based on its closeness even if it has been previously visited by another termite, resulting in a milder change in contrast.

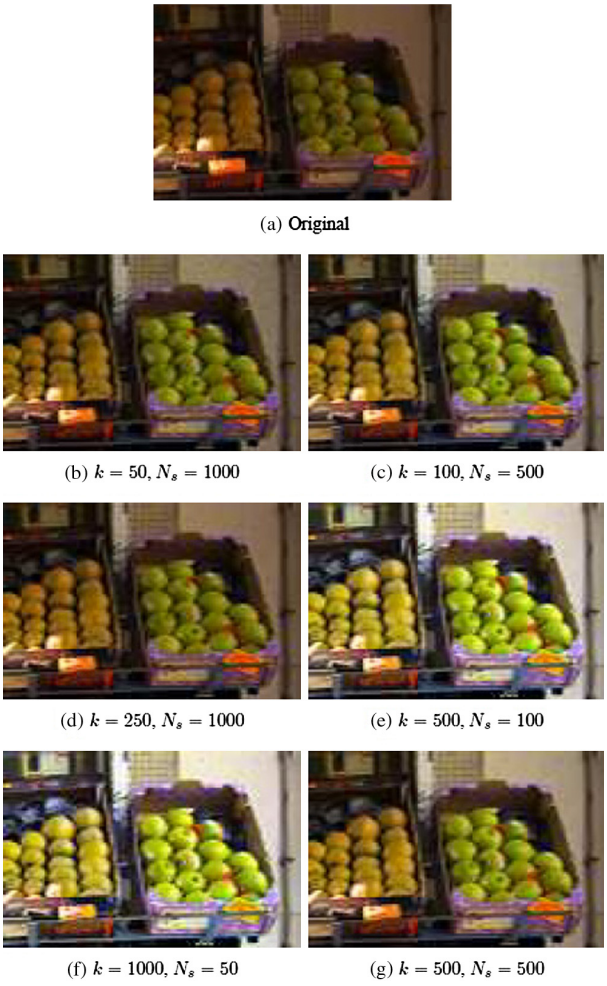
(a) Original

(b) $k = 50, N_s = 1000$ (c) $k = 100, N_s = 500$

(d) $k = 250, N_s = 1000$ (e) $k = 500, N_s = 100$

(f) $k = 1000, N_s = 50$ (g) $k = 500, N_s = 500$

Fig. 4 Results obtained on varying the number of termites k and path length N_s. The different configurations lead to a loss of details in some cases and in an increase of noise in others.



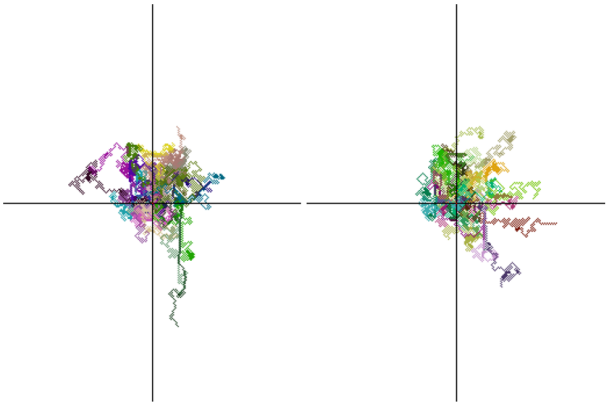
(a) Original

(b) $k = 50, N_s = 1000$ (c) $k = 100, N_s = 500$

(d) $k = 250, N_s = 1000$ (e) $k = 500, N_s = 100$

(f) $k = 1000, N_s = 50$ (g) $k = 500, N_s = 500$

Fig. 5 Details of Fig. 4.



(a) $\alpha = 0.9, \beta = 0.1$ (b) $\alpha = 0.1, \beta = 0.9$

Fig. 6 Termite investigation with two different configurations of α and β and quantity of poison $Q = 1$. The path of each termite is distinguished by a different color. A high value of α and a low value of β make the termite swarm to explore a wide area of the image as shown in the bottom-left, while a low value of α and a high value of β make the termite swarm to explore a smaller area of the image as shown in the bottom-right.

These tests, presented in the following section, will allow us to rewrite Eqs. (5) and (6) as follows:

$$p_k(r,s) = \begin{cases} \dfrac{(\theta_s)^{0.1}(c_{r,s})^{0.9}}{\sum_{u \in N_8} (\theta_u)^{\alpha}(c_{r,u})^{\beta}} & \text{if } s \in N_8 \\ 0 & \text{otherwise} \end{cases}, \tag{11}$$

$$\tau_u = \tau_u + 1, \tag{12}$$

$$\theta_u = \frac{1}{\tau_u}. \tag{13}$$

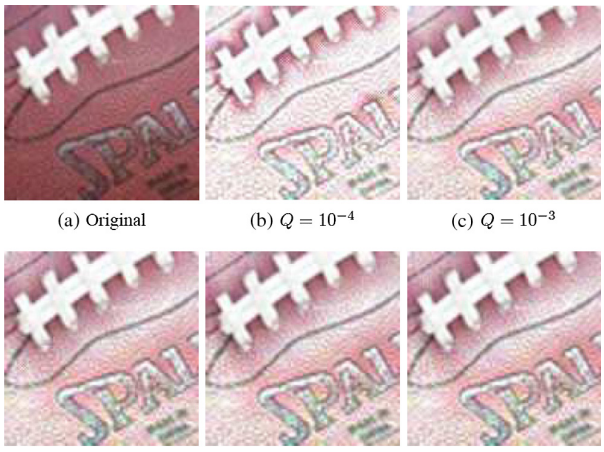
We highlight that Eq. (5) becomes simplified as the memory M_k is no longer applied in Eq. (11), allowing the termite to revisit a pixel.

4.2 *Computational Complexity*

The computational complexity of the original ACS proposed by Dorigo et al.[20] is $O(NC \cdot n^3)$, where NC is the number of ant cycles and n is the number of cities in an instance of the TSP problem. Despite its higher computational complexity, the ACS reaches the optimal solution of the TSP problem in a shorter computational time than other heuristics.[30] In our case, the ant cycle is not necessary because we do not need to converge to an optimal solution, and furthermore, at each pixel recomputation, each termite does not have to touch all the pixels. As a consequence, the computational complexity of the Termite Retinex is given by

$$O(k \cdot N_s \cdot n), \tag{14}$$

where k is the number of termites, N_s is the number of pixels (length of the path) visited by a termite, and n in this case is the number of pixels in the image. The TR follows the same computational complexity of other SCAs, such as RSR[18] or STRESS,[2] which have a computational complexity of $O(N \cdot M \cdot n)$, where N is the number of iterations, M is the number of samples, and n is the number of pixels in the image.



(a) Original (b) $Q = 10^{-4}$ (c) $Q = 10^{-3}$

(d) $Q = 10^{-2}$ (e) $Q = 10^{-1}$ (f) $Q = 1$

Fig. 7 Original on the top followed by termites with $\alpha = 0.1$, $\beta = 0.9$ and increasing quantity of poison Q.

As each pixel recomputation is independent from each other, and termites can be seen as autonomous coworkers, TR has been implemented in compute unified device architecture (CUDA) to exploit parallel computing on GPU. This leads to a gain in speed of approximately a factor of 60 with respect to a CPU-based implementation. The performances have been tested on Intel i3-2310M with 4.0 GB of RAM, Windows 7 64 bit and a GeForce GT 540M as GPU. With this configuration, which is just an entry-level configuration for home use, whose performances are very different from a typical high-level laboratory configuration,[31] TR takes ~7 min to process a 512×512 image with the rule of thumb mentioned above.

5 TR and Color Constancy

The scientific discussion about color constancy (CC) started in the late 19th century with the study of the appearances of objects in different illuminations. In 1872, Hering wrote: "The approximate constancy of the colors of seen objects, in spite of large quantitative or qualitative changes of the general illumination of the visual field, is one of the most noteworthy and most important facts in the field of physiological optics. Without this approximate constancy, a piece of chalk on a cloudy day would manifest the same color as a piece of coal does on a sunny day, and in the course of a single day it would have to assume all possible colors that lie between black and white."[32] As it can be noticed from this quote, the relation between the physics of light and the visual appearance it produces has a long scientific history.

In digital imaging literature, two main different approaches to CC can be found according to their goal in modeling this phenomenon: computer vision CC (in some scholar work referred as computational CC, but we prefer to refer it as CV CC since both CCs have computational aspects) and human vision CC (HV CC). They have distinct goals, different kind of outcomes are expected, and different measures of performance are required.

CV CC has the goal of separating illuminant from reflectance or alternatively estimating the physical reflectance of objects in different illuminants, or alternatively estimating the

illuminant spectral or colorimetric component. Separating reflectance from the illuminant in the color signal (reflectance illumination) is a well-known ill-posed problem;[33,34] thus these algorithms need constraints or assumptions on the scene content, illumination, or geometry. In the case of reflectance estimation, the measurement of the error could be the difference at each wavelength between the actual measured surface reflectance and the estimated one, but this is not practical and dependent on the devices involved. Thus, the scaled integrated reflectance or its chromatic coordinates is often used. As well, in the case of illuminant estimation, the correlated color temperature or its chromatic coordinates is used. In any case, CV CC aims to totally cancel the interaction between reflectance and illumination. For a detailed overview on CV CC, the reader can refer to Ref. 35.

In HV CC, illuminant component is not totally canceled; it generates appearances that are close to reflectance, but show significant departures from reflectance. These departures serve as important signatures of the underlying visual mechanisms.[8] Algorithms belonging to the Retinex family, and thus TR, share this approach and attempt to mimic the response of human vision.

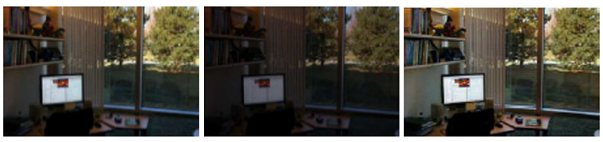
These two approaches to CC start from the same scientific history and have common points and differences. Both

(a) Image 1

(b) Image 2

(c) Image 3

(d) Image 4



(e) Image 5

Fig. 8 Original in the middle, that processed by Termite Retinex (TR) on the right, and that processed by Random Spray Retinex (RSR) on the left.

(a) Image 6

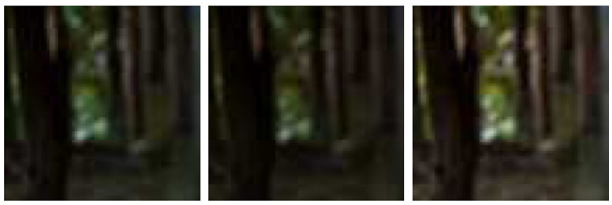
(b) Image 7

(c) Image 8

(d) Image 9

Fig. 9 Original in the middle, that processed by TR on the right, and that processed by RSR on the left.

(a)



(b)

(c)

Fig. 10 Different details of Fig. 6. Original in the middle, that processed by TR on the right, and that processed by RSR on the left.

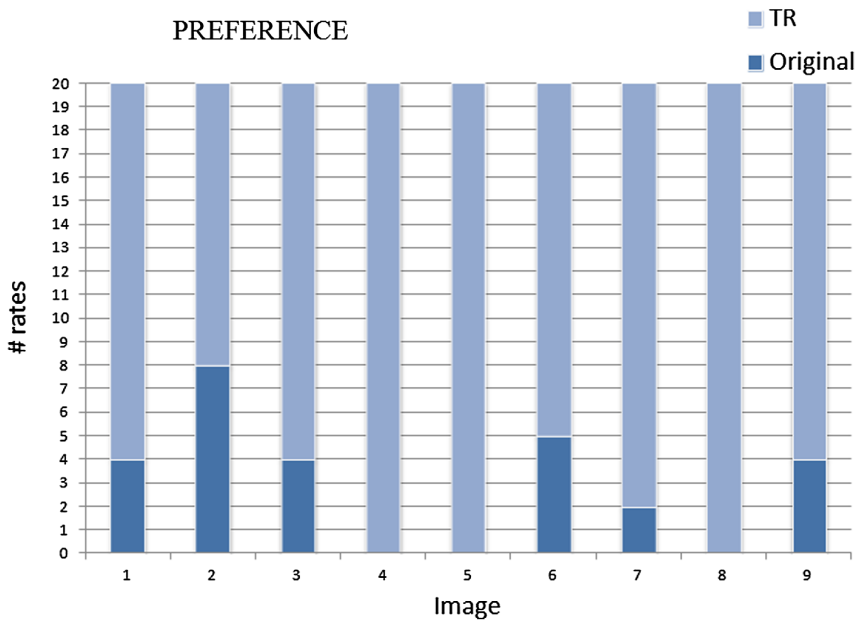


Fig. 11 First experiment results: observers overall preference of TR with respect to its original on the nine tested images.

approaches introduce a sort of normalization with regard to the illuminant but can differ in the estimated pixel output according to the use of spatial arrangement of the scene. CV CC aims at computing reflectance and HV CC aims at computing appearance. These two different outputs are sometimes considered identical. This comes from early experiments that proved that appearance correlates with scaled integrated reflectance,[36] but this has been found to hold true only for flat surfaces in uniform illumination.[37,38] There are many examples in which human appearance does not correlate with reflectance.[39] The simplest example is simultaneous contrast: HV CC reports different appearance of grays on white and black surrounds, while CV CC has to report these grays as identical.[8] For many years, these two variables, reflectance and appearance, have been treated as a unique correlated feature. This has been the incorrect assumption in many cases of weak differentiation between CV CC and HV CC. Moreover, considering these two approaches as similar has led to simplified assumption about vision, in some cases useful for CV CC, but unrealistic for HV CC.

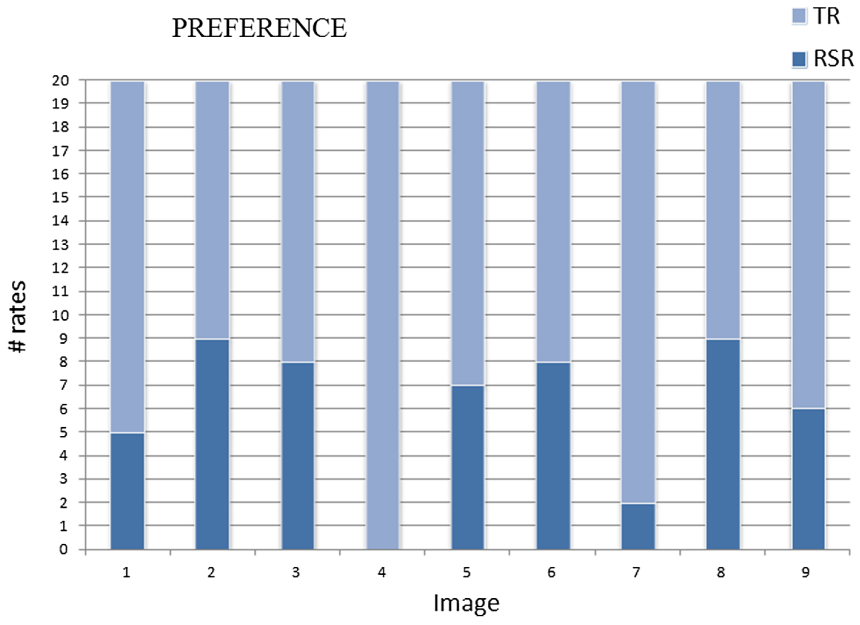


Fig. 12 Second experiment results of the first round: observers overall preference of TR with respect to RSR on the nine tested images.

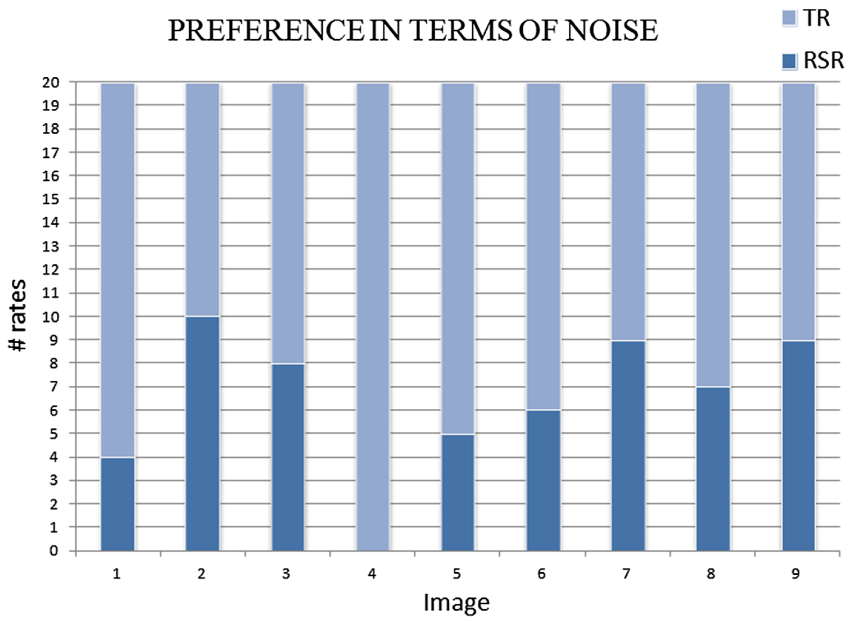


Fig. 13 Second experiment results of the second round: observers preference in terms of perceived noise of TR and RSR on the nine tested images.

The first HV CC algorithm has been Retinex by Land and McCann.[5] In the original Retinex algorithm and in all its derived versions, the spatial locality of the filtering is dependent on the image content and on the way the image is explored. As presented in the above sections, TR can be regarded as a tool to investigate locality.

In 1983, Land retired as Polaroid Chairman of the Board and presented the first departure from the original Retinex model. He simplified the algorithm and lost some of its most powerful properties. In these latter papers, Land describes two versions: the first was the same as Land and McCann[6] but without the reset[40] and the second calculates a designator[8] driven by the average pixel values.[41] The first used a series of paths that began in the surround and traveled to a central pixel of interest. Since each path did not reset, it used the average of pixel values as a way of normalizing the output to the maxima in each waveband. Thus, the calculation became a Gray World comparison, instead of the White Patch behavior of the original Retinex. The paper discussed calculating color appearances found in a flat color Mondrian uniformly illuminated. The designator of the 1986 paper calculates the ratio of a central area to an extended fixed surround. This meant that the threshold operation was replaced by averaging over a large area, and the Gray World average in 1983 was replaced with local-average dependence in 1986. Land emphasized that the designator was still the three-channel Retinex model of color.

These two variants were easier to formalize and consequently simpler to implement; this gave rise to a renewed interest about Retinex, starting from the 1997 work from NASA.[42] Following Land 1986 and the derived NASA implementation, this local averaging approach has had a good success, obtaining interesting color normalization results, mainly without dealing with the problem of visual appearance.

HV CC can be of inspiration for CV CC algorithms, such as Ebner's work.[43] This relevant work in the field of CV CC starts with a clear differentiation between the two kinds of CC, while at the same time presenting some similar features. It is based on local space averaging as Land designator, and it addresses the problem of high computational cost with a parallel implementation. Parallel and multilevel implementations are useful ways to get closer to the amazing efficiency of the human vision system. It also differs from the classic CV CC evaluation methods presented above since it uses an object recognition task, a method also used in similar applications of HV CC algorithms.[44]

Qualitative and quantitative measurements of how accurately HV CC algorithms are able to predict the HVS visual appearance should be the metric for HV CC algorithms.[26] For a correct quantitative evaluation, any model of vision has strict requirements concerning calibration of input information; however, calibrating the input image is a delicate process that cannot be done with generic images acquired in unknown conditions. Thus, in many cases, these algorithms are used as image enhancers. Among the many image enhancers available in the literature, the algorithms like TR, inspired by the original Retinex and MI-Retinex models, have some interesting properties. Basically, they behave as edge enhancer and gradient suppressors, whose intensities

(a) Original (b) TR

Fig. 14 An example of unsupervised color correction.

(a) Original (b) TR

Fig. 15 An example of unsupervised dynamic range stretching.

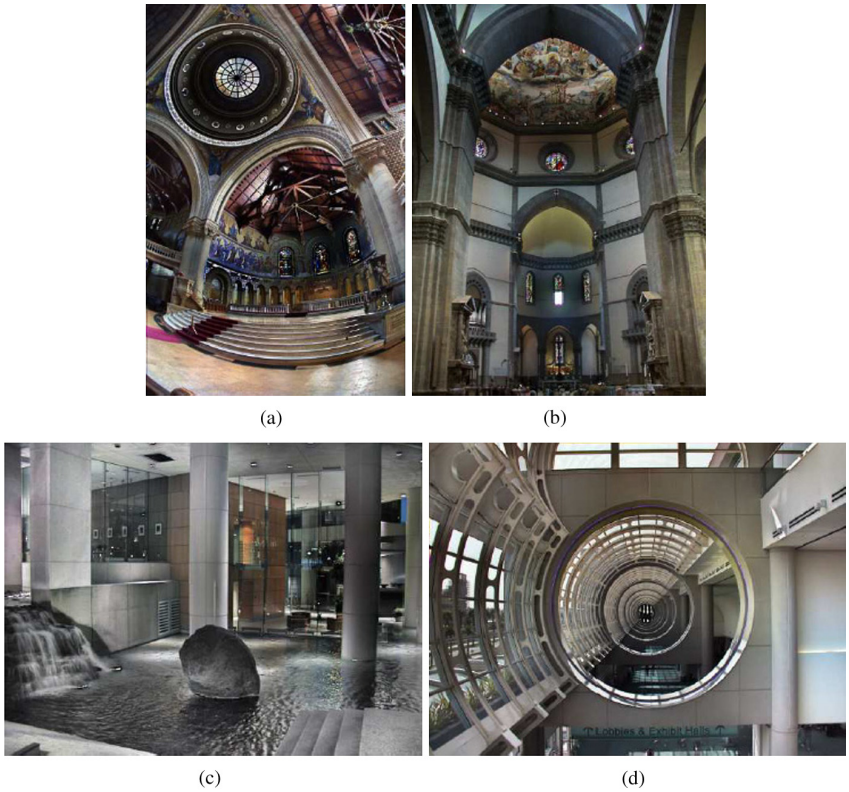
and localities depend on the chosen parameters. Regarding these algorithms being used as image enhancers, tests have proven that a filtering that goes qualitatively in the direction of the human visual appearance in the major part of the cases increases the user preferences.[45–47] For this reason and since we present TR as an alternative formalization of RSR and other MI-Retinex implementations, we have tested TR performances as described in the following section.

6 Test Results and Discussion

6.1 *Experiments*

Since the perceived quality of two images and the perception of single attributes, such as contrast or noise, are not influenced by the environment in which they are observed,[48,49] two experiments in uncontrolled environments with human observers were performed in order to evaluate the quality of TR. A set of nine images displayed in an sRGB color space, chosen following the recommendations from Refs. 50 and 51, were evaluated in a pairwise comparison for both experiments, on a neutral gray background by a total of 20 observers. All the participants were recruited from the computer science field with most of them having knowledge of image processing.

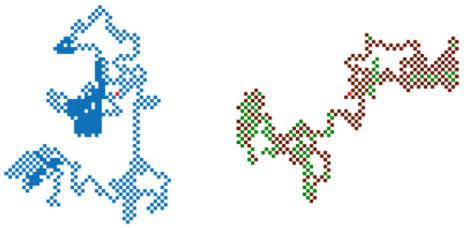
In the first experiment, each image was processed using TR and was compared to its original. Observers were asked to choose one of the images based on their overall preference. No indications of any image quality attribute were given to the participants.[52] Most of the participants of this first experiment also performed the second experiment.



(a) (b)

(c) (d)

Fig. 16 Examples of unsupervised tone rendering of famous high-dynamic-range images.[56] (a) Memorial, (b) Duomo, (c) Vancouver Building, (d) Convention center.



(a) M_k Active (b) M_k Disabled

Fig. 17 An example of path of one termite with memory active on the left and memory disabled on the right. In green the pixels that the termite have revisited. A continuous loop visit between only two pixels is prevented by the poison update.

In the second experiment, each image processed using TR was compared to the same image processed using RSR. In the first task, the observers were asked to choose the image according to their overall preference, while in the second task, they were asked to choose the image according to their preference only in terms of perceived noise. All the participants for this second experiment performed both the first and second tasks. This second experiment was designed with the purpose of evaluating the path-based sampling mechanism of TR against a spray-based one such as RSR.

Figures 8 and 9 show the nine original images in the middle, the corresponding processed by TR on the right, and the corresponding processed by RSR on the left. Figure 10 shows different details of Fig. 6. Figure 11 shows the preference of the 20 observers on the tested images for the first experiment. It can be noticed that TR succeeds on all the images, with three of them having a preference equal to 100%. A sign-test at 95% confidence interval shows that TR is significantly better than the original. The sign-test is a nonparametric statistical test that is a useful alternative to the familiar two-sample t-test in the case where the data do not follow the normal distribution.

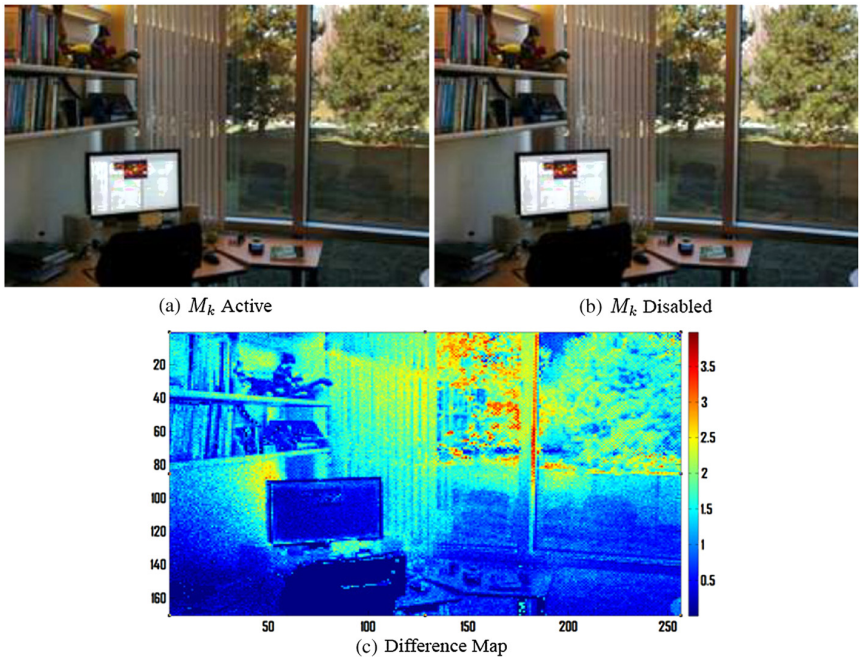
Figure 12 shows the overall preference of the 20 observers in the second experiment, where TR was compared to RSR. TR is preferred for all the nine tested images. Figure 13 shows the preference in terms of perceived noise of the 20 observers for TR and RSR. TR is able to render less noisy scenes for all the images with respect to RSR except for Fig. 2, where the preferences are tied. A sign-test at 95% confidence interval shows that TR is significantly better than RSR. Furthermore, the same test shows that TR is able to process images with higher quality and less noise with respect to RSR. In conclusion, from these tests, we can suggest TR as new effective version of path-based Retinex with the particular novelty of swarm intelligence behavior.

6.2 *Properties of TR*

Retinex is a white patch algorithm[8] and TR follows the same behavior. The brighter areas in the image are mapped toward white, and this is performed locally in a way that is edge-preserving. Furthermore, like other SCAs,[26] TR tends to produce images whose histograms are flatter than those of the originals. It also performs automatic color correction (Fig. 14) and dynamic range maximization if starting from a low dynamic range image (Fig. 15). In case of high dynamic range images, like other SCAs, TR can be used directly as a local tone-rendering operator. Examples are shown in Fig. 16.

6.3 *Implementation Issues*

As mentioned in Sec. 4.2, TR has been implemented in CUDA, like it has been done for other SCAs. In this



(a) M_k Active (b) M_k Disabled

(c) Difference Map

Fig. 18 On the top, results of TR on Fig. 4 with memory active and disabled, respectively. On the bottom, the difference map calculated using the Euclidean color difference formula for small-medium color differences in log-compressed OSA-UCS space (ΔE_E).[53,54] ΔE_E reports highest color differences around 4.0, which in terms of image differences are perceptual, but acceptable differences.[55]

case, another interesting issue to consider is the memory M_k, which affects if a termite can revisit a pixel and then choose another direction. As the memory M_k is a tabu list associated to each termite k, memory overhead may occur for a large number of termites and a long path.

A test on the same nine images shows that by disabling the memory M_k, which simply means not recording in the tabu list which pixels the k'th termite has visited and allowing it to walk through them again (Fig. 17), TR renderings are similar to the ones with memory M_k active. An example is shown in Fig. 18. The differences shown in Fig. 18(c) are calculated using the Euclidean color difference formula for small-medium color differences in log-compressed Optical Society of America Uniform Color Space (OSA-UCS) space (ΔE_E),[53,54] which reports highest color differences around 4.0, which in terms of image differences are perceptual, but acceptable differences[55] to present to the observers. No particular difference was reported from the participants in terms of perceived quality and noise. As benefit, the computational load decreases and a further gain in speed is achieved. However, the memory M_k can be used as a tuning parameter in order to obtain slightly different spatial effects.

7 Conclusions

We have developed a novel implementation of Retinex, reconsidering the idea of the paths and using an existing swarm intelligence model inspired from a biological process. This new algorithm named TR has been originated from the modification of the ACS model proposed by Dorigo et al.[20] In this case, the purpose of TR is not the optimization of some constraints but an exploration of the image content tuned, in particular, by two parameters, α and β, which weight the relative importance of the so-called poison and of the so-called closeness. Results report low importance of the poison and high importance of the closeness, which causes the termite swarm to concentrate to a particular region of an image.

We have carried out two perceptual experiments in order to evaluate the quality of TR. A first experiment on nine images with 20 observers confirms the efficacy of the method on all the tested images in comparison to the original. A sign-test at 95% confidence interval shows that images processed with TR are significantly better than the original images. A second experiment shows that the overall performance of TR is higher than RSR, and a sign-test at 95% confidence interval confirms this statement.

In conclusion, we have presented a novel effective version of path-based Retinex with the particular novelty of swarm intelligence behavior.

Acknowledgments

This work has been funded by the Research Council of Norway over the SHP project. The authors would like to thank Marius Pedersen and Jon Yngve Hardeberg (Gjøvik University College) for their useful feedback and suggestions.

References

1. A. Rizzi, C. Gatta, and D. Marini, "From Retinex to automatic color equalization: issues in developing a new algorithm for unsupervised color equalisation," *J. Electron. Imaging* **13**(1), 75–84 (2004).
2. Ø. Kolås, I. Farup, and A. Rizzi, "Spatio-temporal retinex-inspired envelope with stochastic sampling: a framework for spatial color algorithms," *J. Imaging Sci. Technol.* **55**(4), 1–10 (2011).
3. M. D. Fairchild and G. M. Johnson, "The iCAM framework for image appearance, image differences, and image quality," *J. Electron. Imaging* **13**(1), 126–138 (2004).
4. J. Kuang, G. M. Johnson, and M. D. Fairchild, "iCAM06: a refined image appearance model for HDR image rendering," *J. Vis. Commun. Image Represent.* **18**(6), 406–414 (2007).
5. E. H. Land, "The Retinex," *Am. Sci.* **52**(2), 247–64 (1964).
6. E. H. Land and J. J. McCann, "Lightness and Retinex theory," *J. Opt. Soc. Am.* **61**(1), 1–11 (1971).
7. J. J. McCann, "Retinex at 40," *J. Electron. Imaging* **13**(1), 6–7 (2004).
8. J. J. McCann and A. Rizzi, *The Art and Science of HDR Imaging*, John Wiley, Hoboken, New Jersey (2011).
9. J. Frankle and J. J. McCann, "Method and apparatus for lightness imaging," U.S. Patent No. 4,384,336 (1983).
10. D. Marini and A. Rizzi, "A computational approach to color adaptation effects," *Image Vis. Comput.* **18**(13), 1005–1014 (2000).
11. B. Funt, F. Ciurea, and J. J. McCann, "Retinex in MATLAB," *J. Electron. Imaging* **13**(1), 48–57 (2004).
12. D. J. Jobson, Z. Rahman, and G. A. Woodell, "Properties and performance of a center/surround Retinex," *IEEE Trans. Image Process.* **6**(3), 451–462 (1997).
13. K. Barnard and B. Funt, "Investigations into multi-scale Retinex," in *Color Imaging in Multimedia*, pp. 9–17, John Wiley and Sons, Hoboken, New Jersey (1998).
14. R. Kimmel et al., "A variational framework for Retinex," *Int. J. Comput. Vis.* **52**(1), 7–23 (2003).
15. M. Bertalmío et al., "Perceptual color correction through variational techniques," *IEEE Trans. Image Process.* **16**(4), 1058–1072 (2007).
16. J. M. Morel, A. B. Petro, and C. Sbert, "A pde formalization of Retinex theory," *IEEE Trans. Image Process.* **19**(11), 2825–2837 (2010).
17. G. Simone and I. Farup, "Spatio-temporal Retinex-like envelope with total variation," in *Sixth European Conf. on Color in Graphics, Imaging and Vision*, IS&T, Amsterdam, The Netherlands (2012).
18. E. Provenzi et al., "Random spray Retinex: a new Retinex implementation to investigate the local properties of the model," *IEEE Trans. Image Process.* **16**(1), 162–171 (2007).
19. R. Montagna and G. D. Finlayson, "Constrained pseudo-Brownian motion and its application to image enhancement," *J. Opt. Soc. Am. A* **28**(8), 1677–1688 (2011).
20. M. Dorigo, V. Maniezzo, and A. Colorni, "The ant system: optimization by a colony of cooperating agents," *IEEE Trans. Syst. Man Cybern. B* **26**(1), 29–41 (1996).
21. M. Dorigo and T. Stützle, *Ant Colony Optimization*, MIT Press, Cambridge, Massachusetts (2004).
22. M. Dorigo and L. M. Gambardella, "Ant colonies for the traveling salesman problem," *Biosystems* **43**(2), 73–81 (1997).
23. M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: an autocatalytic optimizing process," Technical Report 91-016, Dipartimento di Elettronica e Informazione-Politecnico di Milano, Italy (1991).
24. E. Provenzi, L. D. Carli, and A. Rizzi, "Mathematical definition and analysis of the Retinex algorithm," *J. Opt. Soc. Am. A* **22**(12), 2613–2621 (2005).
25. D. H. Brainard and B. A. Wandell, "Analysis of the Retinex theory of color vision," *J. Opt. Soc. Am. A* **3**(10), 1651–1661 (1986).
26. A. Rizzi and J. J. McCann, "On the behavior of spatial models of color," *Proc. SPIE* **6493**, 649303 (2007).
27. C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. of the Sixth Int. Conf. on Computer Vision*, pp. 839–846, IEEE Computer Society, Bombay, India (1998).
28. G. Audino, "Termites: a Retinex implementation based on a colony of intelligent cooperating agents," Master's thesis, University of Milano-DICO, Italy (2012).
29. G. Simone et al., "Termites: a Retinex implementation based on a colony of agents," *Proc. SPIE* **8292**, 82920N (2012).
30. M. Dorigo and L. Gambardella, "Ant colony system: a cooperative learning approach to the travelling salesman problem," *IEEE Trans. Evol. Comput.* **1**(1), 53–66 (1997).
31. Y.-K. Wang and W.-B. Huang, "Acceleration of the retinex algorithm for image restoration by GPGPU/CUDA," *Proc. SPIE* **7872**, 78720E (2011).
32. E. Hering, *Outline of a Theory of Light Sense*, L. M. Hurvich and D. Jameson, Trans., Harvard University Press, Cambridge, p. 1964 (1905).
33. K. Barnard, V. Cardei, and B. Funt, "A comparison of computational color constancy algorithms—Part i: Methodology and experiments with synthesized data," *IEEE Trans. Image Process.* **11**(9), 985–996 (2002).
34. K. Barnard et al., "A comparison of computational color constancy algorithms. ii. Experiments with image data," *IEEE Trans. Image Process.* **11**(9), 985–996 (2002).
35. M. Ebner, *Color Constancy*, Wiley, Hoboken, New Jersey (2007).
36. J. J. McCann, S. P. McKee, and T. H. Taylor, "Quantitative studies in Retinex theory a comparison between theoretical predictions and

observer responses to the colormondrian experiments," *Vis. Res.* **16**(5), 445–458 (1976).
37. C. Parraman, A. Rizzi, and J. J. McCann, "Colour appearance and colour rendering of HDR scenes: an experiment," *Proc. SPIE* **7241**, 72410R (2009).
38. C. Parraman, J. J. McCann, and A. Rizzi, "Artist's colour rendering of HDR scenes in 3-d Mondrian colour-constancy experiments," *Proc. SPIE* **7528**, 752802 (2010).
39. J. Albers, *Interaction of Color*, Yale University Press, New Haven (1975).
40. E. H. Land, "Recent advances in Retinex theory and some implications for cortical computations: color vision and the natural image," *Proc. Natl. Acad. Sci. U. S. A.* **80**(16), 5163–5169 (1983).
41. E. H. Land, "An alternative technique for the computation of the designator in the Retinex theory of color vision," *PNAS* **83**(10), 3078–3080 (1986).
42. D. J. Jobson, Z. U. Rahman, and G. A. Woodell, "A multiscale Retinex for bridging the gap between color images and the human observation of scenes," *IEEE Trans. Image Process.* **6**(7), 965–976 (1997).
43. M. Ebner, "Color constancy based on local space average color," *Mach. Vis. Appl.* **20**(5), 283–301 (2009).
44. G. Ciocca et al., "Retinex preprocessing of uncalibrated images for color based image retrieval," *J. Electron. Imaging* **12**(1), 161–172 (2003).
45. A. Rizzi et al., "Automatic lightness and color adjustment of visual interfaces," in *Human Computer Interaction*, Torino, Italy (2003).
46. C. Parraman and A. Rizzi, "Searching user preferences in printing: a proposal for an automatic solution," in *Printing Technology SpB06* (2006).
47. C. Parraman and A. Rizzi, "User preferences in color enhancement for unsupervised printing methods," *Proc. SPIE* **6493**, 64930U (2007).
48. S. Zuffi et al., "Controlled and uncontrolled viewing conditions in the evaluation of prints," *Proc. SPIE* **6807**, 680714 (2008).
49. G. Simone, M. Pedersen, and J. Hardeberg, "Measuring perceptual contrast in uncontrolled environments," in *European Workshop on Visual Information Processing*, Paris, France, pp. 102–107 (2010).
50. J. Holm, I. Tastl, and T. Johnson, "Definition & use of the ISO 12640-3 reference color gamut," in *Fourteenth Color Imaging Conf.: Color Science and Engineering Systems, Technologies, Applications*, pp. 62–68, IS&T/SID, Scottsdale, AZ (2006).
51. CIE, "CIE 156:2004, Guidelines for the evaluation of gamut mapping algorithms," CIT Central Bureau, Vienna (2004).
52. M. Pedersen et al., "Attributes of image quality for color prints," *J. Electron. Imaging* **19**(1), 011016 (2010).
53. C. Oleari, M. Melgosa, and R. Huertas, "Euclidean color-difference formula for small-medium color differences in log-compressed OSA-UCS space," *J. Opt. Soc. Am. A* **26**(1), 121–134 (2009).
54. G. Simone, C. Oleari, and I. Farup, "Performance of the Euclidean color-difference formula in log-compressed OSA-UCS space applied to modified image-difference metrics," in *11th Congress of the Int. Colour Association*, Sydney, Australia, pp. 81 (2009).
55. J. Y. Hardeberg, *Acquisition and Reproduction of Color Images: Colorimetric and Multispectral Approaches*, Dissertation. com, Parkland, Florida (2001).
56. E. Reinhard et al., *High Dynamic Range Imaging—Acquisition, Display and Image-Based Lighting*, Morgan Kaufmann Publisher, San Francisco, California (2005).

Gabriele Simone received his bachelor of information technology in 2005 and master of information science and technology degree in 2007, both at University of Milan, Department of Information Technology, Italy. He is currently pursuing his PhD in color imaging at University of Oslo, and he was a member of The Norwegian Colour and Visual Computing Laboratory at Gjøvik University College and with main research topics contrast measure, image difference metrics, and tone mapping algorithms in HDR images.

Giuseppe Audino received his bachelor of digital communication degree in 2009 and master of information technology for communication degree in 2012, both at University of Milan, Department of Computer Science, Italy. He was a member of the The Norwegian Colour and Visual Computing Laboratory at Gjøvik University College in 2011, and his main research topic is swarm intelligent methods for image processing and code optimization using GPU programming.

Ivar Farup received his MSc degree in physics from the Norwegian University of Science and Technology, Trondheim, Norway, in 1994 and his PhD in applied mathematics from the University of Oslo, Oslo, Norway, in 2000. He is currently a full professor at Gjøvik University College, Norway, mainly focusing on color science and image processing.

Fritz Albregtsen is Cand. Real. from Department of Theoretical Astrophysics, University of Oslo, 1978. He has been with the Department of Informatics, University of Oslo, since 1983, as professor of digital image processing since 1997. He has coauthored more than 150 peer-reviewed research papers. His main research activities are at present analysis medical microscopy images and automatic segmentation of medical images. In addition, he is working on general quantitative metrics for image quality.

Alessandro Rizzi is an associate professor at the Department of Computer Science, University of Milano, teaching fundamentals of digital imaging, multimedia video, and human–computer interaction. He is one of the founders of the Italian Color Group and a member of several program committees of conferences related to color and digital imaging. Since 1990, he is researching in the field of digital imaging and vision. He is particularly interested in the perceptual issues related with digital imaging and lighting.