Universitat Autònoma de Barcelona

Computational Model of Visual Perception: From Colour to Form

SeyedArash Akbarinia

Universitat Autònoma
de Barcelona

Computational Model of Visual Perception: From Colour to Form

A dissertation submitted by **SeyedArash Akbarinia** to the Universitat Autònoma de Barcelona in fulfilment of the degree of **Doctor of Philosophy** in the Departament de Ciències de la Computació.

Bellaterra, July 13, 2017

Director	**Dr. C. Alejandro Párraga** Centre de Visió per Computador Universitat Autònoma de Barcelona
Thesis committee	**Dr. Marcelo Bertalmío** Department of Information and Communication Technologies Universitat Pompeu Fabra **Dr. Nicolai Petkov** Intelligent Systems University of Groningen **Dr. Joost van de Weijer** Centre de Visió per Computador Universitat Autònoma de Barcelona
International evaluators	**Dr. Casimir Ludwig** School of Experimental Psychology University of Bristol **Dr. Thorsten Hansen** Department of General and Experimental Psychology Justus-Liebig-Universität

This document was typeset by the author using LaTeX 2ε.

ISBN: 978-84-945373-4-9

Printed by Ediciones Gráficas Rey, S.L.

Science is not only a disciple of reason but, also, one of romance and passion.
— Stephen Hawking

Dedicated to you from zero to infinity . . .

Acknowledgements

This doctorate project would not have been possible if it was not for the help, contribution, support and love of many distinct individuals. Unfortunately, I cannot mention names of every single one of them here, otherwise the acknowledgements would be even lengthier than the dissertation itself. However, I would like to genuinely express my deepest appreciation to all those who assisted me to complete my Ph.D.

I would like to express my profound gratitude to my advisor *Prof. C. Alejandro Párraga* for his continuous support and stimulation throughout my Ph.D study. His great critical mindset taught me immensely about science and I honestly am indebted to him for that. Besides that, his patience and guideline significantly helped me during challenging times. Por lo tanto, muchísimas gracias *Alejandro* por todo.

A very special gratitude is dedicated to *Prof. Xavier Otazu* and all the members of the NeuroComputation and Biological Vision Team (*NeuroBiT*) for their constructive suggestions and encouraging comments that considerably helped me to improve the quality of my research. I would like to thank every single one of them very much, moltíssimes gràcies *Xavier*, *Xim* i *David*; and muito obrigado *Lídia*.

My sincere thanks also goes to *Prof. Karl R. Gegenfurtner* and all the members of the *Allgemeine Psychologie* group at the Justus-Liebig-Universität for providing me with a unique and fascinating opportunity to stay in their laboratory as a research exchange. They enlightened me with a different perspective of the visual sciences, which certainly was very valuable for my investigations. Vielen Dank *Gießen*.

Many thanks to every single and each one of my colleagues and friends at the *Centre de Visió per Computador* for sharing with me their precious scientific opinions, offering me their generous help when needed, and probably even more importantly to make my experience during these three years extremely delightful and memorable. *Moltes gràcies, muchas gracias, thanks a lot, kheyli mamnun, tesekkur ederim, ganxie ni, molte grazie, dhanyavad, bahut shukriya, mut shakkran, hvala lepo, efharisto poli, ...*

Last but not the least, I would like to thank friends and family ...

Words fall short to express my wholehearted gratitude to my lovely parents and my brother for literally everything. I basically owe you all I have accomplished. *Baba*, *Maman* va *Ehsan* kheyli doostetoon daram. And to the rest of my family back home in Iran and all over the world, you will never guess how much I miss you.

And finally, to the love of my life, mi *Raquel*. I am very delighted that we are sharing this journey together. I simply would not have been able to finish my Ph.D if it was not for your inspiration, both scientifically and emotionally. Te quiero mucho. And muchas gracias a mi *familia nueva* in Spain, who helped me a lot in this period.

Abstract

The original idea of this project was to study the role of colour in the challenging task of *object recognition*. We started by extending previous research on *colour naming* showing that it is feasible to capture colour terms through parsimonious ellipsoids. Although, the results of our model exceeded state-of-the-art in two benchmark datasets, we realised that the two phenomena of *metameric lights* and *colour constancy* must be addressed prior to any further colour processing. Our investigation of metameric pairs reached the conclusion that they are infrequent in real world scenarios. Contrary to that, the illumination of a scene often changes dramatically. We addressed this issue by proposing a *colour constancy* model inspired by the dynamical centre-surround adaptation of neurons in the visual cortex. This was implemented through two overlapping asymmetric Gaussians whose variances and heights are adjusted according to the local contrast of pixels. We complemented this model with a generic contrast-variant pooling mechanism that inversely connect the percentage of pooled signal to the local contrast of a region. The results of our experiments on four benchmark datasets were indeed promising: the proposed model, although simple, outperformed even learning-based approaches in many cases. Encouraged by the success of our contrast-variant surround modulation, we extended this approach to detect boundaries of objects. We proposed an *edge detection* model based on the first derivative of the Gaussian kernel. We incorporated four types of surround: full, far, iso- and orthogonal-orientation. Furthermore, we accounted for the pooling mechanism at higher cortical areas and the shape feedback sent to lower areas. Our results in three benchmark datasets showed significant improvement over non-learning algorithms.

To summarise, we demonstrated that *biologically-inspired* models offer promising solutions to computer vision problems, such as, *colour naming, colour constancy* and *edge detection*. We believe that the greatest contribution of this Ph.D dissertation is modelling the concept of *dynamic surround modulation* that shows the significance of *contrast-variant* surround integration. The models proposed here are grounded on only a portion of what we know about the human visual system. Therefore, it is only natural to complement them accordingly in future works.

Key words: *visual perception, computer vision, visual neuroscience, colour, form, contrast, surround modulation*

Resumen

La idea original de este proyecto fue estudiar la importancia del color en el *reconocimiento de objetos*. Comenzamos extendiendo la investigación previa sobre *nombrar colores* y demostrando la viabilidad de capturar términos de color a través de elipsoides. Aunque nuestros resultados superaron el estado-del-arte en dos bases de datos, vimos que los fenómenos de *luces metaméricas* y *constancia de color* debían ser tratados antes de cualquier procesamiento de color. Nuestra investigación de pares metaméricas mostró que son infrecuentes en el mundo real. Contrariamente a eso, la iluminación de una escena a menudo cambia drásticamente. Abordamos este problema proponiendo un modelo de *constancia de color* inspirado en la adaptación dinámica del centro-envolvente de las neuronas en la corteza visual. Esto se implementa a través de dos gaussianos asimétricos superpuestos, cuyas varianzas y alturas se ajustan al contraste local. Complementamos este modelo con un mecanismo genérico de agrupación variante por contraste que inversamente conecta el porcentaje de señal agrupada al contraste de una región. Los resultados sobre cuatro bases de datos fueron prometedores: nuestro modelo superó incluso los enfoques basados en el aprendizaje en muchos casos. Alentados por el éxito obtenido, ampliamos este enfoque para detectar los bordes de los objetos. Proponemos un modelo de *detección de bordes* basado en la primera derivada del kernel gaussiano. Incorporamos cuatro tipos de envolvente: completa, distante, orientación isogonal y ortogonal. Además, contamos con el mecanismo de agrupación en las áreas corticales superiores y la retroalimentación de la forma enviada a las zonas más bajas. Nuestros resultados en tres bases de datos mejoraron el estado-del-arte en los algoritmos sin aprendizaje.

En resumen, hemos demostrado que los modelos *inspirados biológicamente* ofrecen soluciones para visión por computador, como *nombrar colores*, *constancia de color* y *detección de bordes*. Creemos que la mayor contribución de esta tesis doctoral es el modelado del concepto de *modulación envolvente dinámica* que muestra la importancia de la integración de envolvente *variante por contraste*. Los modelos propuestos se basan en sólo una parte de lo que sabemos sobre la visión humana. Por lo tanto, es natural complementarlos en trabajos futuros.

Palabras clave: *percepción visual, visión por computador, visual neurociencia, color, forma, contraste, modulación envolvente*

Resum

La idea original d'aquest projecte va ser estudiar la importància del color al *reconeixement d'objectes*. Comencem estenent la investigació prèvia sobre l'*anomenament de colors* i demostrant la viabilitat de capturar termes de color a través d'el·lipsoides. Tot i que els nostres resultats van superar l'estat de l'art utilitzant dues bases de dades, vam veure que els fenòmens de *llums metamèriques* i *constància de color* havien de ser tractats abans de qualsevol processament de color. Sobre la nostra investigació de parells metamèriques concloem que són infreqüents en el món real. Contràriament a això, la il·luminació d'una escena sovint canvia dràsticament. Abordem aquest problema proposant un model de *constància de color* inspirat en l'adaptació dinàmica del centre-envoltant de les neurones al còrtex visual. Això s'implementa a través de dues gaussianes asimètriques superposades, les variàncies i les alçades de les quals s'ajusten amb el contrast local dels píxels. Complementem aquest model amb un mecanisme genèric d'agrupació variant per contrast que connecta inversament el percentatge de senyal agrupada amb el contrast d'una regió. Els resultats sobre quatre bases de dades van ser prometedors: el model proposat superava, en molts casos, els models basats en aprenentatge. Encoratjats per l'èxit obtingut, ampliem aquesta proposta per detectar les vores dels objectes. Proposem un model de *detecció de vores* basat en la primera derivada del nucli gaussià. Incorporem quatre tipus de voltants: completa, distant, orientació isogonal i ortogonal. A més, comptem amb el mecanisme d'agrupació en les àrees corticals superiors i la retroalimentació de la forma, que és enviada a les zones més baixes. Els nostres resultats en tres bases de dades van millorar l'estat de l'art en els algoritmes sense aprenentatge.

En resum, hem demostrat que els models *biològicament inspirats* ofereixen solucions per a visió per computador, com *anomenament de colors, constància de color* i *detecció de vores*. Creiem que la major contribució d'aquesta tesi doctoral és el modelatge del concepte de *modulació envoltant dinàmica* que mostra la importància de la integració de *l'entorn que varia segons el contrast*. Els models proposats es basen en una part del que sabem sobre la visió humana. Per tant, és natural complementar-los en treballs futurs.

Paraules clau: *percepció visual, visió per computador, visual neurociència, color, forma, contrast, modulació envoltant*

Contents

List of Figures

List of Tables

1 Introduction

In our everyday experience, we see the world so naturally and perceive our surroundings so effortlessly that we have little appreciation for all the astonishing computations occurring in our brains. We are utterly unconscious of the complexity of our visual system to the extent that in the sixties, pioneering scientists of artificial intelligence believed that functional machine vision could be accomplished through a summer project [182]. Today, after more than half-century of exhaustive theoretical and empirical research, we know how unrealistic this view has been. Neurobiological studies suggest that more than sixty percent of our brain is involved in vision-related tasks [69, 237], manifesting the great effort needed to process visual information. This is also evident in convolutional neural networks (CNN) which require millions of tuned parameters to reach human-like performance in just one specific task. Despite remarkable technological progress made in the field of computer vision, the capabilities of human vision remain vastly superior to their artificial counterparts [30, 79, 212], suggesting that machines can learn a great deal from visual systems evolved over millions of years [106]. This has been the motivation behind the approach we followed in this research, namely, *biologically-inspired computer vision*.

Biologically-inspired artificial intelligence and computer vision are among those disciplines that have received a great amount of attention during the past decade [114], thanks to the promising results achieved by CNNs that were originally inspired by our very own cortical organisation [146]. CNN models have obtained one of their greatest accomplishments in the field of computer vision and image processing, owing it to their architecture that resembles the neural circuits of the human brain. Consequently, many within the scientific community [53, 85] believe that in order to comprehensively decipher the basis of visual intelligence, today more than any other time, we require and should expect an amplification of the collaboration between the three distinct vision communities: neuroscientists (brain), cognitive-scientists (mind), and computer-scientists (machine). It is thought that only this collective work might allow us to understand the *bigger picture*.

This doctorate dissertation is an attempt towards this multidisciplinary approach. We intended to learn from the physiological and psychophysical literature about the underlying mechanisms of visual perception. We were curious about *how neuronal spikes lead to perception and eventually to awareness*. Subsequently,

we developed models that aim to simulate those mechanisms closely enough to be of interest to investigators who study visual perception, but also sufficiently practical to be executed in standard computers. To be concrete, during this Ph.D we implemented three *biologically-inspired* computer vision algorithms: (i) *colour constancy*, (ii) *colour naming*, and (iii) *edge detection*.

1.1 The rationale to get inspired by the visual cortex

There are various motives to pursue a biologically-inspired approach towards computer vision [114]. For us, the first and foremost reason is the large gap between the performance of the human visual system (HVS) and its artificial counterpart [30, 79, 212]. This is due to a number of fundamental differences that can be broadly divided into two groups according to their connection with the overall task:

1. *Generic*: those common operations that must be executed regardless of what the eventual objective is. This group can be interpreted as low-level preprocessing mechanisms that are task-irrelevant.

2. *Specific*: those certain actions that are only required for a particular goal, *e.g.* detecting the colour of an object or recognising a scene. This group can be interpreted as high-level visual information processing that are task-relevant.

1.1.1 Generic processing

Poor image quality has been reported to harm the performance of the best CNN models [58]. At the same time those networks are easily deceived by playing with the statistics of images [169, 174, 218] (*e.g.* labelling static white noise as an object with a high certainty). There are multiple other examples that one can refer to illustrate the fundamental differences between the abilities of human and machine vision with respect to low-level visual information processing [150]. We restrict those to two cases that are prerequisite of all other visual information processing: *scene illuminant* and *noise*.

Scene illuminant

In our daily life, dramatic changes occur in the spectral composition of the light reflected from a scene (*e.g.* the gamut of physical colours at sunset almost doubles in comparison to the "flat" midday illumination [122]; set aside artificial illuminants). There are also large intensity variations within a scene itself [81] (*e.g.* sunny versus shadowy region, mutual inter-reflections, *etc.*). Despite this, we usually experience a constant and robust perception of the surroundings and its composing objects.

Contrary to that, insufficient illumination causes great challenges for machine vision, an affect that most people have probably experienced, for example, while capturing a picture with cameras of mobile phones in a dim evening.

Although, in a number of aspects (*e.g.* resolution, angle of view, bit-depth, dynamic range) the hardware capabilities of the modern camera sensor is arguably comparable to human photoreceptors [211], our personal experience of a scene is still vastly superior to the captured images. This hints at underlying efficient processes that take place in our brain in order to "stabilise" the illuminant across and within scenes. Naturally, in the case of computers, a similar procedure must occur prior to any further higher-level visual processing. This is the subject of study for a few computer vision lines of research [97] (*e.g.* colour constancy, high-dynamic-range (HDR) imaging, colour stabilisation, *etc.*).

Noise

One of the puzzling features of our visual system is its extreme robustness to noise [105]. For instance, our performance is much higher than that of CNN models in distorted images [59]. People who are unlucky enough to require glasses experience many times a great surprise when cleaning their lenses. One starts asking oneself "How could I have possibly seen anything with this amount of dirt?". Despite this, our visual system does not get hindered and works rather smoothly. A similar scenario is experienced while driving in the rain or in fog. Although, a large portion of the windshield is covered with water drops, most people still can manage to continue driving. Similar conditions have caused accidents for autonomous driving cars that in addition to cameras are equipped with a number of other sensors to detect obstacles [66]. Currently, it is hard to imagine such autonomous vehicles to function relying mostly on visual information, like humans do.

A large portion of our tolerance to noise in comparison to machine vision does certainly not originate from the hardware (camera sensors versus retina photoreceptors). Overall, the superiority of the human visual system primarily stems from the fact that our visual cortex with its complex processing (software) is able to intelligently interpret the information from our eyes. It goes without saying that a "noise removal" mechanism is an essential preprocessing stage and is required prior to any other higher-level visual task processing [39].

1.1.2 Specific processing

Figure 1.1 illustrates two masterpieces by artist Coles Phillips. To our eyes the edges between foreground and background are well defined in his paintings, despite the fact that the physical colours (intensity in chromatic channels) of some objects

and their backgrounds are identical. This is refereed to as volume completion, a task in which the visual system determines the relationships between different components in order to form meaningful entities [223]. Such cases are extremely ambiguous for machine vision and they greatly disturb their routines to detect, recognise or segment objects.

Figure 1.1 – "Fadeaway" technique by Coles Phillips. Objects merge seamlessly with the background, yet in the perception of viewers the edges remain well defined.

Similar to what we discussed above with respect to generic processing, below we elaborate on some of the most important differences between human and machine vision with two examples regarding efficiency of each system while facing insufficient clues and information.

Occlusion

Although in everyday activities, our visual system continually operates among occluded objects, in most cases we are not conscious of this issue and behave as if occluded objects were present in their entirety. This capacity has also been reported in infants and even many other animal species [123]. When we encounter an occluded object, not only we are aware of its identity but also we estimate its location and spatial extent correctly [105]. In fact, this hardly inhibits our functional abilities regardless whether we consider a partial occlusion in an static image or complete occlusions in video sequences. In contrast, occlusion handling is a great challenge for machine vision and is the subject of study for a few computer vision lines of research [212] (*e.g.* face recognition, object tracking, scene segmentation, *etc.*).

Silhouette

Humans, even at an early age, have almost no difficulties in recognising objects merely by sketches of their outlines (*i.e.* a few high contrast lines) [213]. This is evident in various simple cave arts, cartoons and comic books (despite the fact that many times lines drawn do not even correspond accurately to luminance or colour contrast edges [232]). Similar conditions can seriously challenge machine vision. In a recent study [141], it has been reported that the accuracy of many modern CNN approaches significantly lowers when the input image is the silhouette of an object or its grey-scale version. This suggests that our visual system encodes objects descriptions with a higher level of abstraction by considering physical and perceptual attributes of objects.

1.2 Models inspired by biology

In the previous section, we discussed fundamental differences between human and machine vision. In this section, we start by commenting on some mechanisms of the human visual system. After that, in order to showcase the practical benefits to computer vision from this line of research, we review two influential biologically-inspired models (one of the retina and another of the cortex).

1.2.1 The human visual system

We will review the relevant parts of the visual system in the introductory section of each chapter of this dissertation. Here, it is sufficient to remind our readers of a few very important general properties (for a more comprehensive explanation, refer to [68, 121, 184]). Our visual system consists of two functional components, the eye, which is analogous to a camera, and the visual cortex in the brain, which does all of the complex image processing (see Figure 1.2).

The photons that enter the eye are absorbed by millions of the photoreceptor cells at the back of the retina [62]. The retina has two kinds of photoreceptors: (i) rods that function in less intense light and are responsible for night vision, and (ii) cones that belong to three types – *i.e.* long-, medium-, and short-wavelength (LMS) which can be interpreted as biological equivalent to RGB sensors – and are responsible for colour vision. Cone cells are densely located in the fovea of the retina while rods are in the peripheral regions. This in turn results into dramatic drop of visual resolution with distance from the centre of vision. The signals produced by cone cell are combined in an antagonistic manner in the retina to from the opponent channels that convey information to the visual cortex through the lateral geniculate nucleus (LGN).

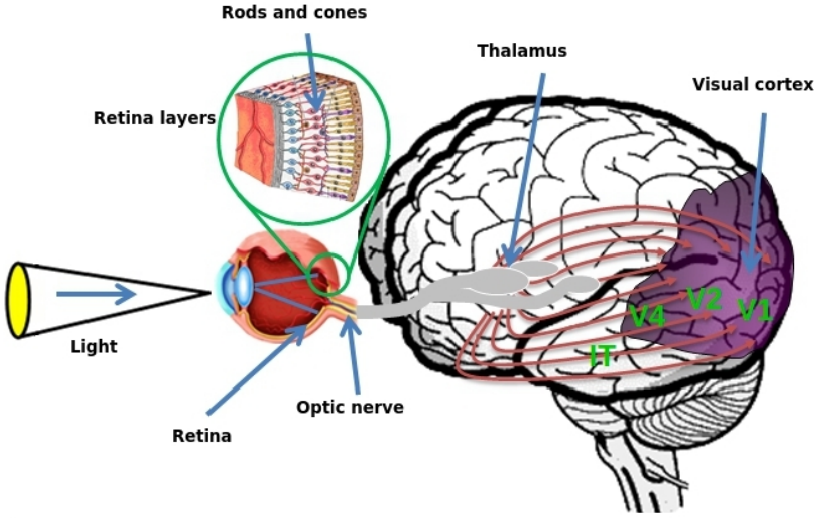


Figure 1.2 – A schematic view of the human visual system adapted from [160].

Neurons in the retina connect to further neurons in the visual pathway in small groups that are sensitive to local regions of the input image. These regions are called receptive fields (RFs). The visual information that arrives to the brain is processed by neurons of the primary visual cortex (V1) that have the smallest receptive field size of any visual area [42]. Neurons in area V1 are tuned to respond to generic low-level visual features (*e.g.* orientation, spatial frequency, phase, motion, *etc.*) within each of their receptive fields. As we advance deeper inside cortical areas, our knowledge of cerebral mechanisms involved becomes less clear. In higher visual areas, neurons have complex properties which are the product of pooling relevant information over several neighbouring spatial locations of the preceding areas [239]. This implies an increasing globality as we progress through the ventral stream (also known as the "what pathway" which is involved in object recognition).

Neurons in the secondary visual cortex (V2) have been reported to respond to extended lines and textural information in addition to sharing many properties of V1 cells. Area V4 was originally labelled as the colour centre of cortex, however many subsequent studies have shown that neurons in this region incorporate both shape and colour visual characteristics [203]. The inferior temporal (IT) part of cortex is considered to be the last stage in the ventral stream [108]. Large receptive fields of this area receive strong inputs from V2 and V4 and it has been proposed

that they play a crucial role in identifying visual objects (*e.g.* face recognition [112]).

1.2.2 Feature descriptors

The Fast Retina Keypoint (FREAK) is a local binary descriptor proposed by [12] following a topological sampling grid of the retina. From physiological studies we know that the size and density of retinal receptive fields increases as a function of their radial distance to the fovea [72]. Correspondingly, in FREAK regions around the centre of a pixel are modelled by narrow Gaussian kernels in order to replicate the retinal pattern. This leads to a higher density of sampling points in the central region, which decreases exponentially towards the periphery of the retina (*i.e.* Gaussian with larger standard deviations). Furthermore, it has been reported that receptive fields in the retina largely overlap [176]. A similar concept was incorporated in FREAK by overlapping Gaussian kernels resulting in higher discriminative power. The last step in FREAK is thresholding between pairs of receptive fields that are fed to a greedy learning mechanism in order to select a restricted number of pairs. Interestingly, the clusters obtained by this approach appear to be in great agreement with what we know about the human retina.

FREAK is one example of biologically-inspired feature descriptors (refer to [11] for a more comprehensive review) that primarily differs from other methods in its sampling strategy: using kernels of different size for each sampling point. This simple idea borrowed from our retina resulted in a robust local binary descriptor. The variability in receptive field sizes is also overwhelmingly present in the visual cortex. Correspondingly, we will show the benefits of modelling different kernel sizes in the phenomenon of colour constancy (discussed in the chapter 3).

1.2.3 Object recognition

In a series of articles [194, 200, 201, 202] a robust object recognition network (called HMAX) that models some basic parts of the feedforward mechanisms of the visual pathway was proposed. In the HMAX model, an image is first convolved with a set of kernels that are responsive to specific features and spatial frequencies (similar to simple neurons in the visual cortex) at each layer. Next, a non-linear max-operator pools signals over a group of these simple units resembling the activity of complex cells in the visual cortex. Higher layers (*e.g.* area V4) collect information from lower ones (*e.g.* area V1) through linear summation of complex cells responses. Finally, a concatenation of responses from higher level units results into feature descriptors similar to those obtained at the area IT of cortex [57]. At the end of this pipeline, these feature descriptors are fed into standard machine learning techniques to recognise objects.

In addition to its promising performance, this model possess a number of fundamental merits: (i) the features used were simple, (ii) a decent accuracy was obtained with a limited number of samples in the training procedure, and (iii) the learned descriptors were generic enough to be tuned easily for other types of objects. Architectures similar to the HMAX (for the benefit of space we do not discuss other models here; interested readers are referred to [219]) demonstrate in practice that the simple and complex cells discovered by Hubel & Wiesel [120] are indeed the backbones of higher processing mechanisms.

Although, current learning solutions driven by CNN models outperform hand-crafted models [140], we decided to use the HMAX model as an example due to its easy-to-grasp features. Alternatively, one can elaborate in details (*e.g.* [236]) about the biologically-inspired components of current popular deep-learning models, such as its convolutional operators (local features) [86] or its backpropagation architecture (feedback connections) [146].

1.3 Organisation of the dissertation

This doctorate dissertation is centred on three main subjects, which correspond to its three main parts:

- Scene illuminant
- Colour names
- Object edges

In the first part (*Scene Illuminant*), we discuss the implications of changing the source of light for human and artificial visual systems. We start by explaining the phenomenon of metamerism and whether it can pose a serious challenge for colour perception. After that we analyse how the colour of an object remains constant to us across different illuminants. At the end, we propose a *colour constancy* model grounded on the contrast-dependant surround modulation of neurons in the visual cortex.

In the second part (*Colour Names*), we continue the discussion about the importance of colours in scene understanding. We reflect on the handful of colour categories (from an infinite set of combinations) that have become universal colour names across different languages and cultures. After that, we examine low-level visual mechanisms that might be responsible for those universal colour terms. At the end, we propose an ellipsoidal *colour naming* model grounded on recent physiological studies of the primary visual cortex.

In the third part (*Object Edges*), we study some aspects of form and shape that have been reported to be closely linked to colour processing in the visual cortex. We start by describing the importance of surround modulation in the perception of forms and boundaries. After that, we explain the related physiological and psychophysical studies regarding the processing mechanisms involved. At the end, we propose an *edge detection* model grounded in the different types of surrounding regions.

Summary of published works

Parts of the materials presented in this doctorate dissertation have been published in the following journals and conferences:

- Section 2 (*Metamerism*)
  - Metameric mismatching in natural and artificial reflectances [2], *Vision Sciences Society (VSS)*, 2017.
  - Metamers in real world scenarios [3], *International Colour Vision Society (ICVS)*, 2017.
- Section 3 (*Colour Constancy*)
  - Colour constancy as a product of dynamic centre-surround adaptation [185], *Vision Sciences Society (VSS)*, 2016.
  - Colour Constancy: Biologically-inspired Contrast Variant Pooling Mechanism [10], *British Machine Vision Conference (BMVC)*, 2017.
  - Colour Constancy Beyond the Classical Receptive Field [Under review] [8], *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2017.
- Section 4 (*Colour Categorisation*)
  - Biologically plausible colour naming model [6], *European Conference on Visual Perception (ECVP)*, 2015.
  - NICE: A Computational Solution to Close the Gap from Colour Perception to Colour Categorization [186], *Journal of PLoS ONE*, 2016.
- Section 5 (*Boundary Detection*)
  - Biologically plausible boundary detection [7], *British Machine Vision Conference (BMVC)*, 2016.

  - Dynamically adjusted surround contrast enhances boundary [4], *European Conference on Visual Perception (ECVP)*, 2016.
  - Feedback and Surround Modulated Boundary Detection [9], *International Journal of Computer Vision (IJCV)*, 2017.

- Miscellaneous
  - Can biological solutions help computers to detect symmetry? [5], *European Conference on Visual Perception (ECVP)*, 2017.
  - New biologically-inspired solutions to old computer vision problems, *Barcelona Computational, Cognitive and Systems Neuroscience (BARCSYN)*, 2016.

Scene Illuminant Part I

How do we account for the source of light?

2 Metamerism

In this chapter, we analyse metameric surfaces in real world scenarios. Metamerism is when two distinct reflectance spectra result into identical colours under certain lighting conditions (see Figure 2.1). Because this phenomenon arises at sensory level (colour absorption), which is the lowest level of any visual system, it can have important implications for any high-level visual tasks. Here, we discuss whether metameric pairs can challenge our visual system with the phenomenon of colour constancy (the ability to preserve the perceived colour of objects under different illuminations), which is the topic of the next chapter in this dissertation.

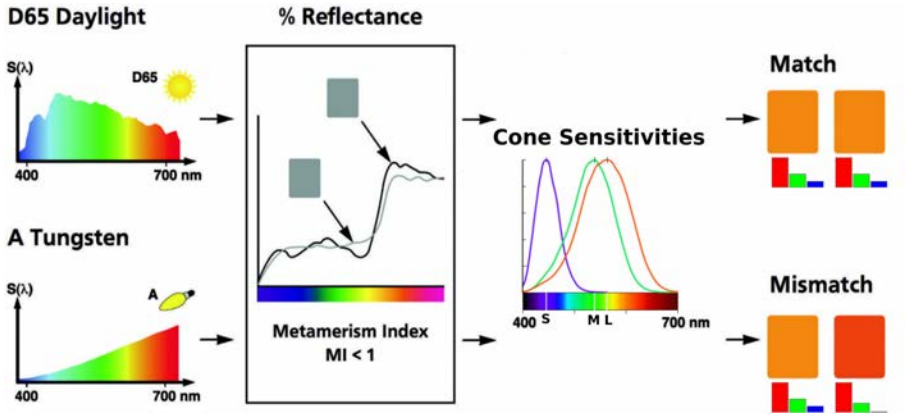


Figure 2.1 – What is metamerism? Two different reflectance spectra result into identical tristimulus values under the illumination of natural daylight and therefore they both appear as orange to us. However, the very same pair mismatch under the Tungsten light and consequently we perceive one of them as more reddish.

2.1 Introduction

The human visual system (HVS) and most digital cameras sample the continuous spectral power distribution of light through three classes of receptors – *i.e.* biologi-

cal long, medium and short wavelength-sensitive cone cells (LMS) or artificial red, green and blue imaging sensors (RGB). This implies that two distinct reflectance spectra can result in identical tristimulus values under one illuminant and differ under another (see Figure 2.1) – a phenomenon known as metamer mismatching [241]. This in turn can potentially become a serious challenge for our visual perception or as a matter of fact for artificial perception as well. For instance, the surfaces of two objects with an identical colour in daylight might appear very different under a typical florescent lamp. The frequency, magnitude, and overall consequences of this issue are still a matter of debate and subject to further research.

A large number of studies have addressed different aspects of the metamerism phenomenon. Lennie [149] associated the sparse distribution of signals in the primary visual cortex (V1) to seldom occurrences of metamers in natural scenes. Finlayson & Morovic [76] mathematically formulated infinite sets of metamers for a given CIE XYZ observation. Foster *et al.* [82] studied naturally occurring metamers in hyper-spectral images under three different daylights and reached the conclusion that their frequency is low. It still remains unclear how frequent metamers are in other real world scenarios, for instance under commonly used artificial sources of light (in particular florescent lamps which have a peaky and narrow-band spectral power distribution).

The extent of metamer mismatching in the visual environment has been investigated in a series of articles. Logvinenko *et al.* [153] theoretically demonstrated that the metamer mismatch volume between two illuminants is rather large and consequently the extent of colour mismatch is significant enough to question the colour constancy paradigm [154]. Evidence supporting an important role of metamer mismatches for perception came from Witzel *et al.* [240], who reported a strong correlation between the size of these mismatch volumes and the degree of colour constancy in an asymmetric matching task. However, in an empirical investigation into the size of these volumes, Zhang *et al.* [251] came to the conclusion that in practice their bodies are substantially smaller than the theoretical ones. There are still a number of ambiguities about the perceptual magnitude of these mismatching volumes in real world scenarios and whether the colour constancy phenomenon is indeed hindered by them. In this chapter, we have investigated the metamerism phenomenon from two viewpoints:

1. Frequency of metameric pairs among surfaces of natural and man-made objects under real world scenarios, using naturally occurring reflectance functions and common artificial lights.

2. Magnitude of perceptual colour shifts in metameric pairs under different sources of light, that is to say how different metameric pairs appear when they mismatch under one illuminant.

In order to address these questions, we gathered a large set of reflectance spectra and studied their colours under various types of illuminant. For each pair of reflectance spectra, we estimated the perceived colour difference using the CIE $\Delta E2000$ [156] metric. We conducted our analysis by computing various degrees of metamerism under different combinations of lower and higher thresholds representing perceptual discriminability. Correspondingly, if the colour difference of a reflectance spectra pair is smaller than the lower threshold under one illuminant (*i.e.* perceptually indiscriminable colours) and larger than the higher threshold under another illuminant (*i.e.* perceptually discriminable colours), this pair is considered to be *metamer*. This definition in some literature is referred to as *paramer* when colours do not match perfectly but the difference is visually unnoticeable [224]. If the colour difference of a reflectance spectra pair is smaller than the lower threshold under both illuminants, this pair is considered by us as *isomer*, meaning the original reflectance spectra of both surfaces are close to identical.

2.2 Method

In this section, we explain the configurational details of the conducted experiments in order to simulate real world scenarios.

2.2.1 Reflectance spectra dataset

We gathered 11,302 reflectance spectra of various natural and man-made surfaces from different sources. Details of the collected datasets are presented in Table 2.1.

The range of wavelengths sampled across all datasets was between 400 and 700 nm (corresponding to the sampling spectra of retinal cones) with distinct intervals (*i.e.* 1, 2, 4, 5 and 10 nm). We unified all the reflectance spectra of the different datasets to intervals of 1 nm through linear interpolation.

In Figure 2.2 we have plotted the angular hue histogram of each dataset considering D65 illumination. One can notice that in the majority of the datasets, *e.g.* Cambridge, FReD and Lumber, the dominant hue is around green, yellow and orange colours. This is due to the fact that the main components of these datasets are natural objects. In a few datasets where artificial colour samples have been measured, this distribution is more uniform across the hue circle, *e.g.* Munsell and Agfa.

Each of the 11,302 reflectance spectrum is to be compared to all other reflectance spectra in order to investigate whether a metameric pair is formed. This means the entire set of experimented spectra pairs contains 63,861,951 possible entries ($\frac{11302\times11301}{2}$).

Name		Samples	Main components
Cambridge	[190]	3276	Fruits, leaves, and pelage
FReD	[17]	2323	Flowers
Munsell	[177]	1600	Munsell chips
Lumber	[129]	1056	Tree logs and their leaves
Papers	[109]	803	Coloured papers
Barnard	[20]	702	Miscellaneous objects
Westland	[238]	404	Plants and soil
Matsumoto	[166]	339	Fruits and leaves
Agfa	[164]	289	Hues at regular intervals
Forest	[116]	272	Sawn timber and its branches
Natural	[183]	182	Colourful plants
Artist	[26]	56	Artist paintings

Table 2.1 – Description of the collected reflectance spectra datasets.

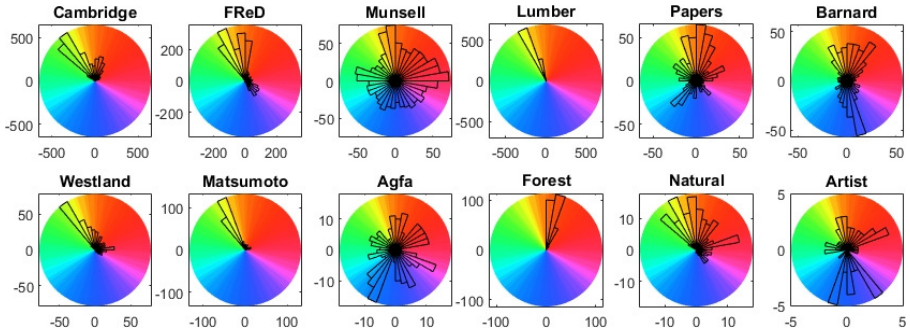


Figure 2.2 – Histogram of hue angles in CIE L*a*b* colour space for each of the collected datasets.

2.2.2 Tested illuminants

To represent a wide range of real world scenarios that our visual system experiences in daily basis, we studied the phenomenon of metamerism under fourteen different illuminants with distinct spectral shapes. Half of the analysed illuminants are naturally occurring daylights while the others are commonly used artificial lamps. In order to investigate the lower bound of our metameric sets, we studied a few illuminants with relatively similar spectral power distributions. We further tried to include illuminants with smooth curvatures as well as irregular narrow-band ones.

Details of the tested illuminants are reported in Table 2.2 and their spectral power distributions are illustrated in Figure 2.3.

	x	y	CCT	Description
D40	0.38	0.38	4001	Evening sunlight
D65	0.31	0.33	6504	Noon daylight
D75	0.30	0.31	7504	North sky daylight
D250	0.25	0.25	25235	Clear blue poleward sky
C	0.31	0.32	6774	Average north sky
Sky97	0.28	0.29	9666	Half cloudy sky
Sky213	0.25	0.26	21283	Bright snow sky
Prime	0.32	0.16	3676	High discriminatory [220]
Halo A19	0.46	0.42	2783	Domestic energy saver
Lorry Light	0.44	0.41	2958	Tungsten halogen light
FL36	0.41	0.41	3624	Domestic white florescent
CFL27	0.46	0.42	2726	White compact florescent
MH43	0.37	0.39	4260	Industrial metal halide
Street Lamp	0.51	0.44	2282	High pressure sodium

Table 2.2 – Description of the illuminants examined. The second and third columns show the CIE chromaticity coordinates for the 2° field of view (1931). Correlated colour temperatures (CCT) are given in the fourth column in Kelvin units.

2.2.3 Colour sensitivity function

We used the CIE 1931 standard colorimetric observer (2°) to compute tristimulus CIE XYZ values [48]:

$$\begin{aligned} X &= \int_{400}^{700} I(\lambda)\, R(\lambda)\, \bar{x}(\lambda)\, d\lambda \\ Y &= \int_{400}^{700} I(\lambda)\, R(\lambda)\, \bar{y}(\lambda)\, d\lambda \\ Z &= \int_{400}^{700} I(\lambda)\, R(\lambda)\, \bar{z}(\lambda)\, d\lambda \end{aligned} \tag{2.1}$$

where $I(\lambda)$ is the illuminant spectral power distribution at wavelength λ nm; $R(\lambda)$ is the reflectance spectra function; and $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ are the CIE 1931 standard colour matching functions. In our computations we normalised each illuminant such that the luminance value of an ideal reflector is equal to unity ($Y = 1$).

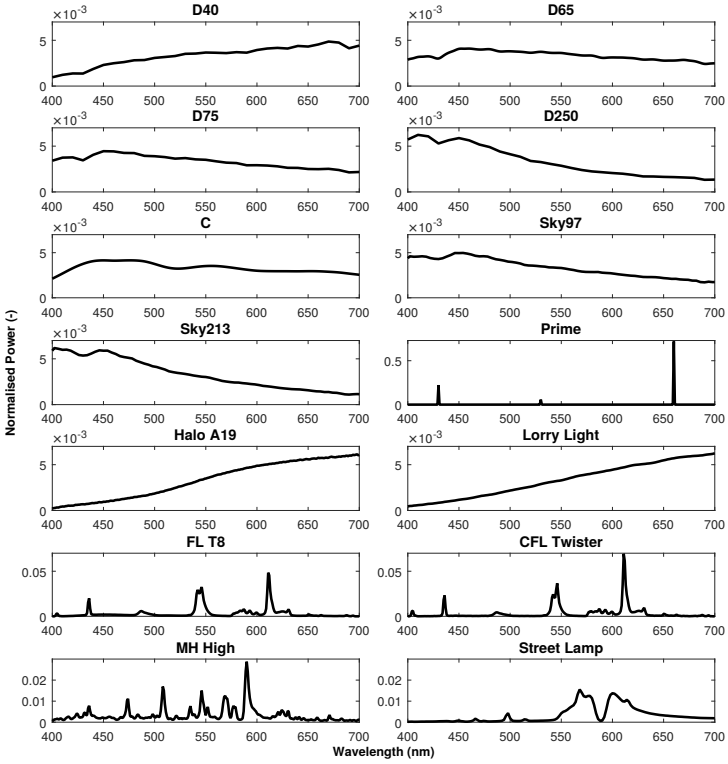


Figure 2.3 – Spectral power distributions of the examined illuminants. For visualisation proposes each distribution is normalised by sum of all its power.

2.2.4 Colour difference function

We estimated the perceived colour difference between two reflectance spectra using the CIE $\Delta E2000$ metric [156] in CIE L*a*b* colour space that is known to be perceptually uniform within a reasonable approximation [48]. Therefore, we converted the tristimulus CIE XYZ values into CIE L*a*b* by computing the white point for each illuminant under the assumption that the reflectance spectrum of a white object is unity.

We decided to measure the phenomenon of metamerism through CIE $\Delta E2000$ because: (i) in real world scenarios there are never two surfaces with reflectance spectra resulting in numerically identical tristimulus values; and (ii) in this study we are more interested in perceptual understating of colour vision and therefore

two surfaces that are perceived indistinguishable meet our criteria.

2.2.5 Procedure of metamer analysis

We computed the colour differences (CIE $\Delta E2000$) of all 63,861,951 reflectance spectra pairs under each of the tested illuminants. We compared these $\Delta E2000$s in two different ways:

1. Illuminant pairs – since we tested fourteen distinct illuminants the entire set of illuminant pairs contain 91 entries ($\frac{14\times13}{2}$).

2. Multi illuminants – comparing all fourteen illuminants simultaneously and therefore capturing all possible metamers across different illuminants.

We considered two surfaces as metameric if they produce perceptually discriminable colours under one illuminant while resulting into indiscriminable colours under another illuminant. Therefore, in order to decide whether a reflectance spectra pair is metameric or not, we defined a set of nominal threshold values ΔE^{Thr}_{low} and ΔE^{Thr}_{high}. If the colour difference of a reflectance spectra pair is smaller than ΔE^{Thr}_{low} under one illuminant (indiscriminable) and larger than ΔE^{Thr}_{high} under another illuminant (discriminable), this pair is determined to be a metamer. In case of multi illuminants comparison we considered a pair as metameric if at least under one of the fourteen illuminants its colour difference is lower than ΔE^{Thr}_{low} and in at least one illuminant it is larger than ΔE^{Thr}_{high}.

We conducted our analysis under $\Delta E^{Thr}_{low} \in \{x|x = 0.5 + n \times 0.5; n = [0,\cdots,9]\}$ (from 0.5 to 5.0 with intervals of 0.5), and $\Delta E^{Thr}_{high} \in \{x|x = 0.5 + n \times 1.0; n = [0,\cdots,20]\}$ (from 0.5 to 20.5 with intervals of 1.0). We chose this large range of lower and higher thresholds due to two reasons:

1. There are disputes about which value of $\Delta E2000$ corresponds to one JND (just noticeable difference). Although often 1 unit of ΔE is mentioned as 1 JND, its uniformity throughout colour space is in question [34, 68]. We computed $\Delta E2000$ of points within the MacAdam ellipses [157], from centres to four vertices along axes, under illuminant C that those ellipses were originally studied. The average resulting $\Delta E2000$ within an ellipse is about 0.50 with a range from 0.23 to 0.89.

2. To study the perceptual magnitude of colour change for a metameric pair when they mismatch under one illuminant, *i.e.* whether they produce dissimilar colours within one category or they might result in colours that are utterly different.

2.2.6 Multidimensional scaling (MDS)

In order to better understand the relation between sources of light and degree of metamerism, we employed the non-metric multidimensional scaling (MDS) technique [197]. We can benefit from MDS to visualise the level of similarities across all tested illuminants. We applied MDS with the dissimilarity metric defined as the frequency of metamers for every combination of illuminant pairs. We chose the metric scaling distance (stress) as normalised with the sum of squares of the dissimilarities.

2.3 Results

2.3.1 Frequency of metameric pairs

In Table 2.3 we have reported absolute numbers of metameric pairs for different values of nominal thresholds (ΔE^{Thr}_{low} and ΔE^{Thr}_{high}); averaged over all 91 illuminant pairs we investigated. In general we can observe that although the absolute number of metameric pairs can appear to be a large quantity their corresponding frequency in the entire set of reflectance spectra pair is low. For instance for $\Delta E^{Thr}_{low} = \Delta E^{Thr}_{high} = 0.5$, there are 12,027 metameric pairs, this means their frequency in the entire set of 63,861,951 pairs is merely $1.9 \times 10^{-4} (\pm 1.2 \times 10^{-4})$; or in simple words only about two samples out of every ten thousands. As the higher threshold increases (*i.e.* a more restrictive discriminability measure), incidents of metameric pairs dramatically decrease. For instance for $\Delta E^{Thr}_{low} = 0.5$ and $\Delta E^{Thr}_{high} = 1.5$ there are just 502 metameric pairs making a frequency of $7.9 \times 10^{-6} (\pm 2.4 \times 10^{-5})$; or in simple words only about eight samples out of every million. The frequencies we observe in this study are in agreement with previously reported figures in the literature [82]. As the lower threshold increases (*i.e.* a less restrictive indiscriminability measure), occurrences of metameric pairs increase as well. For instance for $\Delta E^{Thr}_{low} = \Delta E^{Thr}_{high} = 1.5$ there are 53,072 metameric pairs making a frequency of $8.3 \times 10^{-4} (\pm 5.2 \times 10^{-4})$; or in simple words about eight samples out of every ten thousands.

Considering the multi illuminant analysis, we naturally observed a higher frequency of metamers since all the metameric pairs under any of the 91 illuminant pairs are accumulated in this analysis. For instance for $\Delta E^{Thr}_{low} = \Delta E^{Thr}_{high} = 0.5$ there are more than three times as many metameric pairs in the multi illuminant analysis than in the illuminant pairs analysis (*i.e.* 38,904 instances meaning a frequency of 6.1×10^{-4}; or in simple words six samples out of every ten thousands). For larger values of nominal thresholds the difference to the illuminant pair analysis is substantially larger; *e.g.* for $\Delta E^{Thr}_{low} = 0.5$ and $\Delta E^{Thr}_{high} = 1.5$ there are 12,543 metameric

High Threshold \ Low Threshold	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5
20,5	2	14	43	106	210	372	588	876	1254	1749
19,5	2	16	50	122	242	427	673	1004	1440	2013
18,5	3	19	58	142	278	488	770	1151	1655	2322
17,5	3	21	65	159	314	553	876	1320	1907	2690
16,5	3	23	73	178	354	625	996	1510	2195	3119
15,5	4	27	82	203	406	719	1150	1750	2567	3695
14,5	4	31	95	234	467	830	1337	2058	3066	4495
13,5	5	36	110	269	541	965	1571	2466	3773	5618
12,5	5	43	129	316	637	1144	1906	3098	4840	7238
11,5	7	51	153	377	762	1398	2432	4090	6425	9652
10,5	8	59	183	453	932	1805	3325	5610	8839	13225
9,5	9	71	222	554	1204	2553	4788	8022	12519	18499
8,5	11	84	276	713	1755	3904	7166	11793	18072	26717
7,5	13	107	361	1036	2886	6154	10987	17623	27033	40750
6,5	17	140	507	1811	4903	9917	17077	27511	43244	67286
5,5	23	196	888	3498	8557	16333	28278	46831	77210	128115
4,5	33	296	1998	6744	15064	29208	52793	94064	170756	
3,5	48	707	4641	13504	31035	64646	129432			
2,5	84	2394	11498	36029	91399					
1,5	502	9536	53072							
0,5	12027									

Table 2.3 – Average absolute numbers of metameric pairs under 91 illuminant pairs. Cells are colour coded with higher incidents being red and lower ones green. For cells where lower threshold is larger than higher threshold no metameric pairs are computed.

pairs making a frequency of 2.0×10^{-4}; or in simple words two samples out of every thousand. This is about 25 times larger than its respective figure in the illuminant pair analysis. We observe such a large increase due to the fact that one of the illuminants we have studied in this chapter is the Prime light [220] whose spectral power distribution is zero everywhere except in three wavelengths: 430, 530 and 660 nm. This light is designed to produce highly discriminable colours and therefore many reflectance spectra pairs that are of a similar colour under other illuminants appear completely different under the illumination of Prime (refer to the depicted examples in Figure 2.4).

Especially "problematic" are surface pairs that are perceptually within 1 JND ($\Delta E_{low}^{Thr} = 1.0$) under one illuminant and appear rather very different under another illuminant (*e.g.* $\Delta E_{high}^{Thr} = 4.5$, which is about the ΔE of one Munsell chip to its nearest neighbours). Such occurrences are rare in general, *i.e.* on average merely 296 incidents from the entire set of 63,861,951 pairs (see Table 2.3); or in simple

words about five samples out of every million. In our analysis such large differences are close to nonexistent under a combination of any natural daylights. Due to the same reasons explained above, we can observe that the frequency of these "problematic" cases for the multi-illuminant analysis is larger (more than 20 times), *i.e.* 5,946 pairs; or in simple words nine samples out of every hundred thousands.

2.3.2 Magnitude of changes under different illuminants

In Figure 2.4 we have illustrated a few examples of reflectance spectra pairs that are metameric at least under one illuminant and appear very different under some other illuminants. For instance in the first row we can observe that two surfaces with distinct reflectance spectra produce perceptually identical colours under the illumination of Street Lamp, *i.e.* $\Delta E = 0.23$. Illuminating the very same pair with most other lights yields to two discriminable colours and in case of Prime light they appear utterly different, *i.e.* $\Delta E = 80$. These surfaces appear as metameric under the illumination of Street Lamp since its spectral power distribution is concentrated only around the wavelength range of 550–650 nm where these two surfaces overlap.

Figure 2.4 – Metameric pairs under some illuminants that are utterly mismatching under other illuminants. The first column illustrates the original reflectance spectra of a pair. Columns two to the end are the colour of each surface under different sources of light. Under each illuminant the left colour belongs to the spectrum represented by the black line in column one and the right colour belongs to the red signal. ΔEs are reported under each pair of reflectance spectra.

In most rows of Figure 2.4 we can observe that two surfaces produce entirely different colours under the illumination of Prime light. This is due to the design of Prime light, which is to produce highly discriminable colours. Spectral power distribution of this illuminant is very peaky and narrow-banded (zero everywhere except in 430, 530 and 660 nm) and any small shifts in reflectance spectra of a surface can result into large colour differences. In few seldom cases we can observe an opposite phenomenon, for example in the third row the present surface pair is metameric under the illumination of Prime (*i.e.* $\Delta E = 0.19$), however they appear as two completely different colours under any other source of light.

In a number of cases we observe that two surfaces appear as metameric under florescent lamps (*i.e.* CFL27 and FL36) although the colour of same surfaces are clearly distinguishable under other illuminants, *e.g.* the second row of Figure 2.4. In many cases we can observe the opposite phenomenon where two surfaces are metameric under natural illuminants although under florescent illumination their colours appear as completely different, *e.g.* the seventh row. This is also due to the relatively peaky and narrow-banded spectral power distribution of many florescent lamps.

By comparing ΔEs under different 91 illuminant pairs we can analyse the magnitude of colour change when metameric pairs mismatch. In Table 2.4 we have compared incidents of metameric pairs when the source of light changes from the standard CIE illuminant D65 to six other illuminants: three natural ones (D40, D75 and D250) and three artificial ones (Street Lamp, FL36 and Prime). From this table we can observe that large colour changes never occur under the illuminant pair "D65 – D75", *i.e.* there are zero incidents when ΔE is lower than 0.5 under one illuminant and larger than 1.5 under another (refer to upper triangle of Table 2.4). Contrary to that, we can observe such large differences under the illuminant pair "D65 – Prime", for instance 623 pairs have a ΔE lower than 0.5 under one illuminant and larger than 1.5 in another. For similar lower and higher thresholds, we can observe 239 incidents under the illuminant pair "D65 – Street Lamp". Our experiments further demonstrate that enormous differences can occur under the illuminant pair "D65 – Prime", although very rarely, *e.g.* in ten incidents ΔE considerably increases from 0.5 to 20.5.

2.3.3 Illuminant pairs

In Table 2.5 we have reported the absolute number of metameric pairs under all illuminant pairs for our lowest nominal threshold, *i.e.* $\Delta E^{Thr}_{low} = \Delta E^{Thr}_{high} = 0.5$. We can observe that most metamers occur under combination of either Prime or Street Lamp with one other illuminant and the maximum incident of metameric pairs occur when these two illuminants are paired together, *i.e.* 36,691 occurrences

Table 2.4 – Absolute number of metameric pairs (from a set of 63,861,951 samples) under illuminant change from D65 to six other sources of light. Cells are colour coded with higher frequencies being red and lower ones green. For cells where the lower threshold is larger than the higher threshold no metameric pairs are computed.

with a frequency of 5.7×10^{-4}; or in simple words about six samples out of every ten thousands. This is not surprising since spectral power distributions of both lights are narrowly banded around certain colours: Prime light due to its design to produce high colour discrimination and Street Lamp as a consequence of its building material which is high pressure sodium.

In general it appears that metameric pairs are less common under combination of any natural daylights (see the square on the bottom right corner of Table 2.5). Lowest number of metameric pairs occur under the illuminant pairs “Sky213 – D250” ans “C – D75” with 561 and 570 incidents, respectively. Highest metameric pairs occur under the illuminant pair “D40 – D250” with 14,283 incidents. This is expected as these two illuminants have a large difference in their correlated colour temperature.

	Prime	Street Lamp	CFL27	FL36	MH43	Halo A19	Lorry Light	D40	D65	C	D75	Sky97	Sky213	D250
Prime		36691	21126	19459	27989	29316	27951	23802	18256	16946	16828	14642	11588	11095
Street Lamp	36691		17053	18410	9560	8727	9844	13625	19125	20435	20579	22821	26123	26576
CFL27	21126	17053		2585	9575	10318	9415	6526	7734	8494	8730	10440	13410	13637
FL36	19459	18410	2585		10602	11443	10424	7141	7003	7573	7747	9171	11947	12146
MH43	27989	9560	9575	10602		4843	4492	5211	10181	11465	11617	13871	17263	17718
Halo A19	29316	8727	10318	11443	4843		1467	5890	11720	13048	13248	15634	19184	19585
Lorry Light	27951	9844	9415	10424	4492	1467		4515	10403	11719	11939	14329	17903	18290
D40	23802	13625	6526	7141	5211	5890	4515		6358	7638	7914	10312	13938	14283
D65	18256	19125	7734	7003	10181	11720	10403	6358		1444	1624	4096	7810	8183
C	16946	20435	8494	7573	11465	13048	11719	7638	1444		570	2826	6522	6879
D75	16828	20579	8730	7747	11617	13248	11939	7914	1624	570		2492	6246	6617
Sky97	14642	22821	10440	9171	13871	15634	14329	10312	4096	2826	2492		3814	4183
Sky213	11588	26123	13410	11947	17263	19184	17903	13938	7810	6522	6246	3814		561
D250	11095	26576	13637	12146	17718	19585	18290	14283	8183	6879	6617	4183	561	

Table 2.5 – Comparison of number of metameric pairs under all illuminant pairs for $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 0.5$. Cells are colour coded with higher frequencies being red and lower ones green. Our entire set of samples contains 63,861,951 reflectance spectra pairs.

This is also evident in Table 2.4 that in general frequency of metamerism is smaller under natural illuminants. The lowest values are obtained under the illuminant pair "D65 – D75". This is not surprising as both illuminants have a similar spectral power distribution and correlated colour temperature. Occurrences of metameric pairs are higher when illumination shifts from D65 to one artificial light. The highest frequencies are observed under the illuminant pairs "D65 – Street Lamp" and "D65 – Prime". In general we can observe that frequency of metameric pairs is increased by one order of magnitude when "D65 – D75" is compared to "D65 – Street Lamp" or "D65 – Prime" for any definition of metamerism where lower and higher thresholds are equal to each other, *i.e.* $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = \{0.5, \cdots, 4.5\}$ (anti-diagonal cells in Table 2.4).

It is worth mentioning that it is not the case that natural pairs always yield a lower degree of metamerism. For instance, frequency of metamerism is higher under illuminant pairs "D65 – D250" in comparison to "D65 – FL36". For $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 0.5$ frequency of metamerism is 1.3×10^{-4} in case of "D65 – D250" versus 1.1×10^{-4} of "D65 – FL36"; or for $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 1.5$ there are almost two times more metameric pairs under the illuminant pair "D65 – D250".

2.3.4 What determines metamers

Prior to conducting the MDS analysis and in order to cover a larger set of conditions, we extended the collection of illuminants with thirteen more sources of light that are presented in Table 2.6. We conducted the MDS analysis by giving as input the frequency of metameric pairs ($\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 0.5$, similar data shown in Table 2.5 but with all newly added illuminants). We have illustrated the output of MDS as configuration points in an arbitrary two dimensional space in Figure 2.5. The Euclidean distances between points in this figure approximate a monotonic transformation of the corresponding dissimilarities in frequency of metameric pairs across different illuminant pairs. Prime light and Street Lamp that produced the most number of metamers in our experiment, are located on both extrema of this figure.

	x	y	CCT	Description
D50	0.35	0.36	5001	Horizon daylight
FL29	0.44	0.40	2937	Standard FL4
LED29	0.47	0.46	2909	Streetlight LED
MH34	0.41	0.40	3452	Industrial metal halide
CFL38	0.39	0.39	3767	Cool white CFL
LED38	0.39	0.38	3811	Street view LED
LED43	0.36	0.34	4258	Outdoor LED
CFL44	0.36	0.37	4405	Cool white CFL
MH45	0.37	0.39	4499	Outdoor metal halide
FL45	0.35	0.36	4997	Standard FL8
LED50	0.34	0.36	5011	Cool white LED
CFL54	0.33	0.36	5417	Daylight CFL
FL64	0.31	0.34	6427	Standard FL1

Table 2.6 – Description of the thirteen extra illuminants we investigated. The second and third columns show the CIE chromaticity coordinates for the 2° field of view (1931). Correlated colour temperatures (CCT) are given in the fourth column in Kelvin units.

In Figure 2.5 we can observe a clear pattern among natural illuminants (open circles in Figure 2.5), *i.e.* they are orderly distributed from left to right as their correlated colour temperature increases. This order is also true for other smooth illuminants (*i.e.* Halo A19 and Lorry Light, which are similar to the illuminant A). We have connected these illuminants with dashed lines in Figure 2.5. One can observe a similar pattern for light-emitting diode (LED) illuminants. We believe this

is a novel and useful way to represent illuminants with respect to their properties related to metamerism.

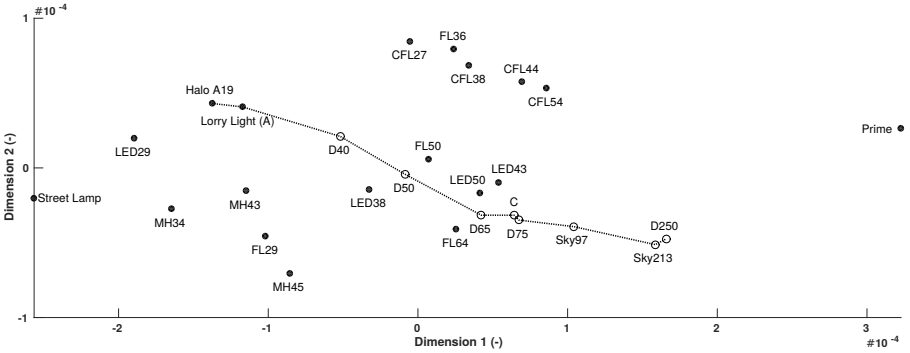


Figure 2.5 – Multidimensional scaling on frequency of metamers (*i.e.* $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 0.5$, similar data presented in Table 2.5 but with all newly added illuminants in Table 2.6). Open circles represent natural illuminants. Filled circles are artificial sources of light. Dashed line connects illuminants with a smooth spectral power distribution that are ordered according to their correlated colour temperature from left to right.

2.3.5 Variation of colour differences across illuminants

We computed variation of colour differences across all illuminants, *i.e.* how much ΔE of a given reflectance spectra pair (both metameric and non-metameric) changes under different illuminants. A variation of value 0 means ΔE stays constant for the colour of two surfaces under different illuminants. On average the absolute variation in the measure of ΔE for the colours of two surfaces across all of the 91 tested illuminant pairs is 3.19. Evidently, illuminants that shrink colour space like monochromatic lights cause minimal colour difference between two surfaces. Consequently ΔE under these illuminants is smaller in comparison to the other sources of light. The contrary is also true. High discriminatory illuminants expand colour space and produce larger quantities of ΔE. Therefore, the variation of colour differences across illuminants can be used as a measure to decide whether an illuminant generates more colours in comparison to other illuminants.

In Figure 2.6 we have illustrated the distribution of these variations for each of the fourteen examined illuminants in the form of a box plot. We can observe that the largest positive variations occur for the Prime light (its first quartile, median and

third quartile are all above zero). This is due to the previously explained narrow-band shape of this light and its specific design to enhance discriminability of colours. On the other side of the range, we can observe that the largest negative variations occur under the illumination of Street Lamp (its first quartile, median and third quartile are all below zero). This light is of high pressure sodium material and only has spectral power between wavelengths 550-650 nm. Our observations suggest it is most difficult to distinguish two colours under this Street Lamp and it is easiest under the Prime illumination.

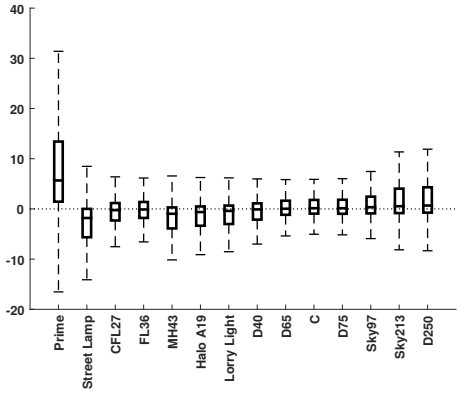


Figure 2.6 – Distribution of variation of colour differences under the fourteen tested illuminants. In other words changes of ΔE of a reflectance spectra pair from one illuminant to another; value 0 means ΔE stays constant for two surfaces under different illuminants. The bottom and top of the box are the first and third quartiles, and the band inside the box is the second quartile (the median). Dashed lines represent the minimum and maximum of all variations.

Among natural lights, D250 and Sky213 (sky at bright snow reflected by mirror) encounter the largest variations of colour differences. However, their median value is right on the zero line and their interquartile range is less polarised. This is in line with the description of daylight as the ideal illuminant [31] and the fact that colour rendering index (CRI) of natural lights is assumed to be 100%.

2.4 Discussion

Naturally, the interpretation of the results obtained in this study (and others) is conditioned by a number of factors, such as the value used to estimate the just

discriminable colour difference, the extent of similarity in the set of illuminants, and the variation of the reflectance spectra in the dataset. In spite of this, the results of our study are fairly consistent across these factors, suggesting that metamerism is infrequent in real world scenarios, under different types of natural and artificial illuminants. On average only two surface pairs out of every ten thousand samples will be metameric (perceptually indiscriminable, *i.e.* $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 0.5$). If we consider a higher threshold for perceptual discrimination (*i.e.* $\Delta E_{low}^{Thr} = \Delta E_{high}^{Thr} = 1.5$) we still do not observe many metameric pairs: only about eight samples out of every ten thousands. Refer to Table 2.3. This is in agreement with a previous study by Foster *et al.* [82].

In our study we find that metameric pairs are less common among natural illuminants. Frequency of metamers is the highest for two narrow-band illuminants: Prime light and Street Lamp (refer to Table 2.5). Two surfaces that are metameric under one illuminant can potentially yield to completely different colours under other sources of light, however this is very unusual. In about five samples out of every million pairs, two surfaces have a ΔE that is within 1 JND under one illuminant and a ΔE that is larger than 4.5 (about the colour difference of one Munsell chip to its neighbours) under another illuminant (refer to Table 2.3).

Although metameric pairs are not frequent in real world scenarios, their magnitude of perceptual colour difference can be enormous in certain cases, for example under the illuminant pair "Prime – Street Lamp" in thirty incidents ΔE significantly increases from a value smaller than 0.5 to one larger than 20. Such big changes never occurred under natural daylights. For instance, under the illuminant pair "D65 – D75" there are no incidents of two surfaces with a ΔE of smaller than 0.5 under one illuminant and a ΔE of larger than 1.5 under another.

Although at first our findings appear to be in contrast to the large size of those theoretical metamer mismatching volumes [153], a more thorough consideration suggests that the frequency reported by Foster *et al.* and us is not that very different from the ratio of those metamer mismatching volumes to the entire colour space. For instance, the average theoretical mismatching volume under illumination condition of "D65 – A" is computed as 140 [251]. If we approximate the entire CIE XYZ colour space as a cube of length 100 (in the study of Zhang *et al.* [251] CIE Y was normalised to a value of 100 for the ideal reflector), and assume that all colours are equally probable, this means that the theoretical mismatching volume covers an area of about 1.4×10^{-4} of the entire colour space, which is very close to the frequency of metameric pairs found by Foster *et al.* and us.

Alternatively, another way to interpret these mismatching volumes would be in comparison to the MacAdam ellipses. Let us assume these mismatching volumes are in shape of a cube in CIE XYZ colour space. This means a volume of 140

is emerged from a cube of approximately length 5.2. In this cube a cylinder of radius 2.6 will be fit. This is larger than the average axis of MacAdam ellipses in CIE XYZ which is about 2.0. However, average empirical mismatching volume under illumination condition of "D65 – A" is computed as 18 [251]. With similar computations we reach to a cylinder of radius 1.3 which is even smaller than the average axis of MacAdam ellipses. This suggests that size of these mismatching volumes are roughly comparable to the MacAdam ellipses. Therefore they might not be substantial enough to interfere with colour constancy.

2.5 Conclusion

In order to answer whether metameric pairs could pose a major problem for our visual perception, at least two factors must be taken into account: (i) how often they occur, and (ii) how different they appear when they mismatch. We discussed the former by arguing that results of our experiments suggest that metameric pairs are rare in real world scenarios. We believe the later is neither a big issue due to the following reasons:

1. The frequency of surface pairs that are metameric under one illuminant and appear very different under another illuminant significantly drops as the colour difference increases. For instance, across all the illuminant pairs we studied there are merely 502 surface pairs out of a set of 63,861,951 elements that have a ΔE that is smaller than 0.5 under one illuminant and a ΔE that is larger than 1.5 (which is outside of the range of computed values for the MacAdam ellipses) under another. This falls further to 84 pairs for those with a ΔE larger than 2.5; or in simple words merely one pair out of every million samples. Furthermore, such large differences are close to nonexistent under natural daylights.

2. The variation of ΔEs for each reflectance spectra pair under different illuminant conditions (inter-variation) is smaller than the variation of colour differences for one surface under different illuminations (intra-variation). Our visual system overcomes the issue of intra-variation through local and global adaptation among other mechanisms. Consequently the colour of objects appear as constant to us under different sources of light [94]. Furthermore, we perceive the colour of objects within their context and it is well established that shape and form influence perception of colours, too [203]. We believe similar mechanisms can help our visual perception to mitigate issues caused by rare metameric pairs. For instance, it is very unlikely that a metameric pair is placed exactly in the same local surrounding, therefore

this difference of surround influence might help our visual perception to discriminate between a metameric pair.

Therefore, based on the results of this study, we can conclude that metameric pairs do not pose a serious issue to our colour perception, specifically to the phenomenon of colour constancy which is discussed in the next chapter.

3 Colour Constancy

In the previous chapter, we discussed the problem of metamerism and whether it can cause complications for colour perception. Another crucial factor in recognising objects is to preserve their perceived colour under different illuminations which occur in real world scenarios. This phenomenon is known as colour constancy (see Figure 3.1). In this chapter, we start by introducing the importance of colour constancy for visual information processing. After this, we explain the challenges involved in estimating the scene's source of light. Finally, we demonstrate computationally that contrast variation in the surrounding regions plays a crucial role for the mechanisms involved in colour constancy.

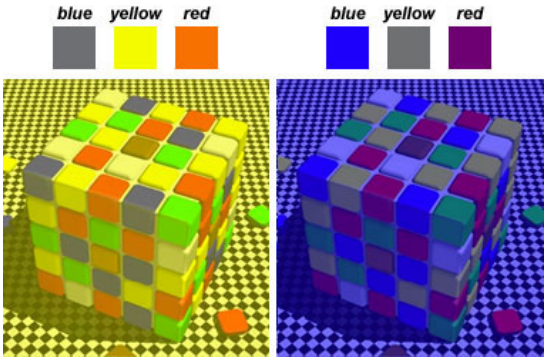


Figure 3.1 – What is colour constancy? A *Rubik's Cube* under the illumination of yellow and violet lights. Although the physical values of tiles are different, we perceive them largely within the same colour category (a few examples are illustrated on top of the figure). Picture is adapted from "HandPrint" website (https://www.handprint.com/).

3.1 Introduction

Colour is an essential property of our visual world. Apart from its intrinsic aesthetic and emotional value, it provides valuable information about the environment by

breaking the luminance pattern of cast shadows, facilitating the segmentation of objects from each other and the background [36]. Despite dramatic changes in the spectral composition of the light reflected from a scene (*e.g.* the gamut of physical colours at sunset almost doubles in comparison to the "flat" midday illumination [122]), to our visual perception the colour of an object appears to be largely the same across illuminants and throughout the day. This ability (termed *colour constancy*), is more impressive if we consider the fact that mathematically, the problem of separating illumination form reflectance is *ill-posed*. If we ignore angular dependencies, the light captured by a camera sensor over the visible spectrum ω is represented as

$$\int_{\omega} R(\lambda)L(\lambda)S(\lambda)d\lambda, \tag{3.1}$$

where $R(\lambda)$ is the spectral reflectance of the objects present in the scene, $L(\lambda)$ is the spectral power distribution of the illumination, and $S(\lambda)$ is the camera's sensor spectral sensitivity function. Even if we assume that $S(\lambda)$ is known – an assumption easily violated in images acquired by commercial cameras – it is impossible to infer $R(\lambda)$ and $L(\lambda)$ with only one equation. Thus, rectifying biased images has infinite possible solutions. It is worth mentioning that in general computational models of colour constancy do not consider the phenomenon of metamerism since they are extremely rare in real world scenarios [2] as we discussed it in the previous chapter.

Although there is no agreement on the precise mechanisms and brain areas responsible for colour constancy, most researchers group them according to the neural level where they likely operate [125]:

1. *Sensory level*: modelled by simple linear transformations of the photoreceptor responses, *e.g.* scaling responses by their mean activities over the image [158, 231].

2. *Perceptual level*: modelled by considering various perceptual "cues", such as, specular highlights [148], mutual reflections [90], and achromaticity of edges [225].

3. *Cognitive level*: modelled by considering colour memory and/or the identification of objects to be able to compensate for the effects introduced by familiar objects [110].

The relative contributions of each of these processing levels is still a matter for debate. However, most researchers acknowledge that cognitive contributions are likely to be small due to the fact that the colour constancy phenomenon can be

largely explained by low level mechanisms present in the retina and areas V1 and V4 of the visual cortex [81].

The significance of colour constancy to both human vision and computer vision communities is demonstrated by the many studies in object detection, tracking, feature extraction, *etc.* [19, 98, 99, 193] from visual perception [16, 81, 144, 162] and computer vision [63, 80, 100, 117] perspectives, which have historically had different objectives. Most visual perception and neuroscience work aims at understanding the phenomenon while most computer vision work aims at predicting the effects of colour constancy. However, one can assume there might be computational advantages in incorporating the knowledge acquired by the brain's neural machinery after millions of years of evolution. To this end, the finely-tuned combination of low-level (mostly hard-wired) and high-level (mostly cognitive) mechanisms that the primate brain has achieved after millions of years of evolution might be understood in terms of the *bias/variance* trade-off common in machine learning [96]. The choice of the best bias will depend on the nature of the training data (*e.g.* how much is known in advance about the problem) and the system's noise.

Biological systems face similar choices. A simple organism living in a fix environment does not need a strong bias and all individuals can safely share the same neural configuration. More complex organisms such as primates face variable environments and need to dedicate part of their brains to learning during their lifetime while leaving large scale neural structures like the sensory cortex genetically specified. This particular combination of bias/variance in complex organisms allows them to adapt to different environments while still keeping crucial survival skills. In the case of colour constancy, most of the brain computations are arguably done at the sensory level [81] indicating that "bias" may perhaps plays a larger role than "variance" (*i.e.* more of a *normalisation* problem than a *learning* problem). This is perhaps the reason why current learning-based solutions have considerable trouble to replicate their results in new (non-learned) datasets [81, 87], using dataset-dependent parameters. Additionally, the majority of methods are constrained to consider only one source of illumination, which in effect hinders their applicability on real scenes [100].

3.1.1 Computational Solutions

From a mathematical point of view, retrieving the colour of a surface illuminated by light of unknown spectral distribution is underdetermined, and to computationally rectify biased images (in the same way colour constancy does) it is common to impose several assumptions regarding the scene illuminant, the statistical distribution of colours or edges, *etc.* [100]. In general, these algorithms can be divided into two categories: (i) learning-based approaches and (ii) low-level features-driven

methods.

Learning-based approaches, *e.g.* [1, 44, 89, 206], train machine learning techniques on some relevant image features. One group of learning-based algorithms is "gamut mapping", which originated from the influential work of Forsyth [80], and was extended by others [18, 73, 74, 102, 170], following the assumption that only a finite set of colours is observable in real world images. Another large group of algorithms considers reflectance as the random variable of a normal distribution under a Bayesian framework [35, 95, 196]. Although learning-based approaches can obtain accurate results, they rely heavily on training data, which is likely to be cumbersome (*i.e.* their overall performance depends on the quality of their training data) and slow [100].

The majority of low-level features-driven methods can be summarised by the following Minkowski framework [77, 225]

$$L_c(p) = \left(\int f_c^p(x) dx \right)^{\frac{1}{p}} = k e_c, \tag{3.2}$$

where $f(x)$ is the image intensity value at the spatial coordinate x; c represents one of the three chromatic channels $\{R, G, B\}$; p corresponds to the Minkowski norm; and k is a multiplicative constant chosen such that the illuminant colour, e, is a unit vector.

Substituting $p = 1$ in Eq. 3.2 reproduces the well known Grey-World assumption, in which the illuminant is estimated by presuming that all colours in the scene average to grey [40]. Setting $p = \infty$ replicates the White-Patch algorithm, which assumes that the brightest patch in the image corresponds to a specular reflection containing all necessary information about the illuminant [144]. In general, it is challenging to automatically tune p for every image and at the same time inaccurate p values may corrupt the results noticeably [100].

The incorporation of high-order image statistics into the Minkowski framework was proposed by van de Weijer *et al.* [225], under the assumption that the edges carry important information about the source of light, thus their algorithm is called "Grey-Edge". The Minkowski framework can be generalised further by replacing the $f(x)$ in Eq. 3.2 with its derivative

$$\left| \frac{\partial^n f_\sigma(x)}{\partial x^n} \right|, \tag{3.3}$$

where $|.|$ is the Frobenius norm; n is the order of the derivative; and σ is the scale of the Gaussian derivative filters convolved with the original image [84].

It has been noted [51, 143, 203, 214] that high-order derivatives have correspon-

dences with the centre-surround mechanism as modelled in colour perception research. This mechanism is activated when localised sensory regions of the retina are stimulated by light. These sensory regions (also called "receptive fields") are characterised in terms of their contribution to cortical neurons' stimulation as "centre" and "surround" [184].

The interplay between centre and surround in receptive fields (RF) is typically modelled by a Difference-of-Gaussians (DoG) [45, 91, 185, 215, 250]. Since, the second order image derivative can be approximated by DoG, they can be a good tool for modelling the sub-cortical mechanisms involved in colour constancy. This simple model of the low-level properties of the mammalian visual system has a long history starting with Enroth-Cugell and Robson in 1966 [65], continuing with Marr in 1980 [163] and more recently applied to colour constancy by Gao *et al.* [91]. However, the efficiency of DoG in estimating the illuminant strongly depends on finding an adequate width for the Gaussian kernel, σ, and the optimal weight of the broader Gaussian function, which are difficult to tune automatically. A solution to this problem has already been found by the human visual system (HVS) in the form of dynamic, contrast-based, centre-surround cortical interactions [14, 207] (see below), which are not present in the classical formulations. Although the ultimate purpose of these non-linear interactions is not known, we speculate here that they might play a role in colour constancy and accordingly, we propose a *fully automatic*, contrast-dependent colour constancy model that overcomes the need for hand-crafted parameters. In our colour constancy model we incorporate three well known properties of cortical (area V1) neurons:

1. The size of the minimum RF (also referred to as centre) varies according to the local contrast of the present stimuli, *i.e.* enlarged when exposed to low-contrast [207];

2. The influence of the surround on the centre varies depending on the local contrast of both centre and surround, with greater inhibition for higher contrast stimuli [14];

3. Cortical RFs increase their diameters systematically by approximately a factor of three from lower to higher areas [239], as they pool signals over a large neighbourhood from the levels below.

The above formulation presents major differences with current DoG-based models like that of Gao *et al.* [91], where the centre size is always constant and the contributions of both centre and surround to the receptive field responses are fixed. Also, the final estimation of luminance was previously based on a simple operation (*e.g.* selecting the peak by max-pooling), whereas we model hypothetical neurons

from a higher area (area V4 neurons) whose receptive fields are substantially larger than those of V1 neurons, pooling signals from area V1 according to the contrast of the corresponding stimulus. We show that this contrast-variant-pooling mechanism can even enhance performance of other models driven by high-order derivatives. To summarise, previous models adopt the *classical* receptive field approach while we go beyond, including the latest physiological findings.

Figure 3.2 shows a flowchart of our colour constancy model. Although a step forward in terms of plausibility, our functional approach still entails an oversimplification of the much more complex (and less well known) interactions between the different neural layers and cortical feedback from higher regions. Following the Occam's razor principle we aimed for the most parsimonious solution that can produce competitive results. It is wroth highlighting that we are not strictly interested in out-competing learning-based solutions in each of the testing datasets. Instead we want to produce an algorithm that works like the HVS does, *i.e.* produces the best possible results in all of the datasets at the same time and *with the same set of parameters*. Equally, we want our solution to be computationally efficient, that is, to incorporate the evolutionary knowledge accumulated by the primate brain in an algorithm potentially implemented in small portable devices. A more multidisciplinary objective of this work is to further understand the role of dynamically-sensitive visual cortical neurons. Throughout this chapter we will refer to our colour constancy model as *Adaptive Surround Modulation (ASM)*.

In summary, the main contributions of this chapter are: (i) the modelling of colour constancy based on more recent physiological findings, *i.e.* two overlapping asymmetric Gaussian functions whose kernels and weights adapt according to centre-surround contrast, (ii) the estimation of the chromaticity of the illuminant by modelling higher visual cortical areas (*i.e.* neurons with large RFs pooling signals from lower areas) according to their local contrast, and (iii) the dynamic generalisation of the colour constancy by using the same parameters to predict results in different datasets with no need to "recalibrate", mimicking what the HVS does.

3.2 Beyond the classical receptive field

In this section we review important physiological findings regarding surround modulation in the visual cortex and describe how we modelled these properties.

3.2.1 Surround modulation in area V1

The concept of non-classical receptive field (RF) became established by the work of Allman *et al.* [13] and today numerous studies show that most V1 cells in cat and

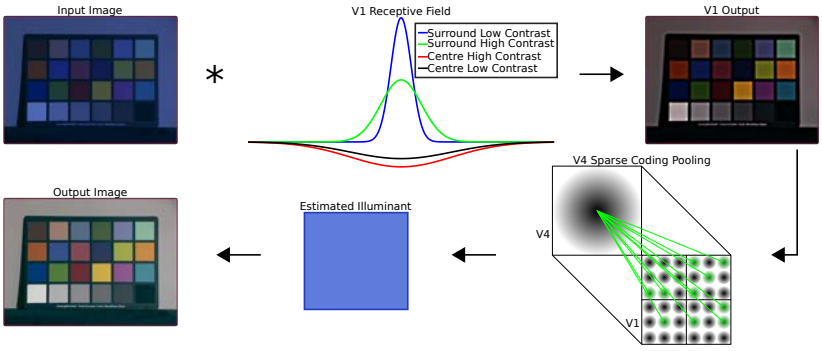


Figure 3.2 – The flowchart of our model. The input image is convolved with a centre-surround contrast-dependent asymmetric difference-of-Gaussian envelope (inspired by V1 neurons that have larger receptive fields at low contrast and are suppressed further by high contrast surround). The output of V1 is pooled by V4 neurons according to the sparse-coding principle considering global contrast of image.

macaque are suppressed by stimuli extending beyond a critical distance (for a full review refer to [14]).

Quantitative results suggest that RFs in cortical area V1 of macaque change their responses when measured at low contrast stimuli [207]. Figure 3.3 shows the responses of a typical macaque neuron when its RFs are exposed to a vertically-oriented sinusoidal grating of constant spatial frequency and varying size [14]. The dashed line at the bottom illustrates the average spontaneous firing rate of the neuron (no stimulation). The black curve represents the neuron's excitation when stimulated by a high (70%) contrast grating of increasing size (increasing grating radius). As the grating's size increases, more of the neuron's receptive field becomes stimulated producing an increase in the neuron's output, a process known as "facilitation". Its maximum output happens when the grating reaches a radius equal to sRF_{high}. After that, increasing the size of the grating only decreases the neuron's output, *i.e.* neighbouring neurons start to "suppress" the neuron's activity until it becomes close to zero. Correspondingly, the grey curve in Figure 3.3 represents the same neuron's activity as a function of grating size when stimulated by a low (12%) contrast grating. The peak of the grey curve (maximum stimulation radius or sRF_{low}) has now shifted to the right of the plot. The area between the two peaks (shaded in the plot) defines a "dual-role" region, *i.e.* gratings of radii

between these two values can either suppress or stimulate the neuron according to its contrast. The existence of this region implies a fundamental change in the way these visual cortex neurons operate, and we have incorporated it at the core of our colour constancy model. Now the receptive field of the neuron can be separated in three regions, a "centre" with radius up to sRF_{high}, a "surround" with radius larger than sRF_{low} and a dual-role area in between which operates like the surround (*i.e.* suppression) when contrast is high and operates like the centre (*i.e.* facilitation) when the contrast is low (see right insert in Figure 3.3).

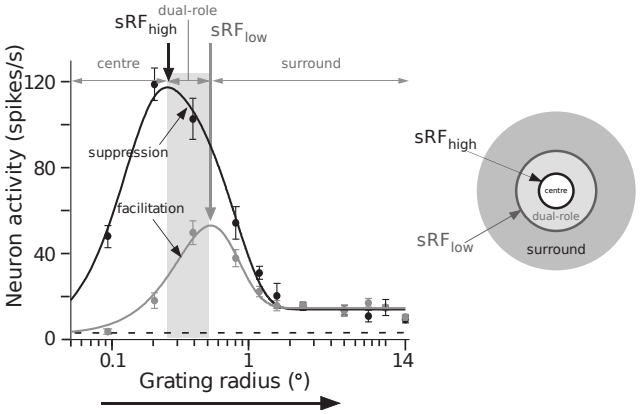


Figure 3.3 – Size tuning curve of an example cell in macaque V1, adapted from [14]. Black and grey curves show responses to a grating of high and low contrast, respectively. The dual-role area is suppressive for high contrast stimuli, whereas it acts as a facilitator in the case of low contrast. The scheme on the right represents the RFs of a V1 neuron. Arrow heads point to radii that determine sRF_{high}(0.26°) and sRF_{low} (0.54°).

Physiological recordings [207] have shown that the radius of the surround in V1 can be about five to six times larger than the value of sRF_{high} and its effects on the centre significantly more complex than those described above. Figure 3.4 illustrates changes in a typical V1 neuron's activity when the stimulation of the centre is fixed and the surround is stimulated by an annuli that becomes increasingly thinner. The plot shows results for three different cases (a) high-contrast is applied to both the centre and the surround; (b) low-contrast is applied to the centre and high-contrast to the surround and (c) low contrast is applied to both the centre and the surround. In all cases, centre-only stimulation (right side of the plot) produces higher neural activity than when both centre and surround are stimulated (left

side of the plot). However, suppression is larger for high contrast stimuli (black curve reaches zero when the whole of the surround is stimulated) and is minimal when both centre and surround are stimulated by low contrast gratings (solid grey curve) [134]. In all cases, suppression is strongest when the orientation of centre stimuli is parallel to that of the surround, an effect known as iso-orientation suppression. This effect can also turn into facilitation as the orientations of the stimuli applied to centre and surround move towards perpendicular directions and the contrast is low. In general, facilitation happens when centre and surround have different characteristics (*e.g.* different spatial frequency, phase or orientation) and it increases when these differences increase.

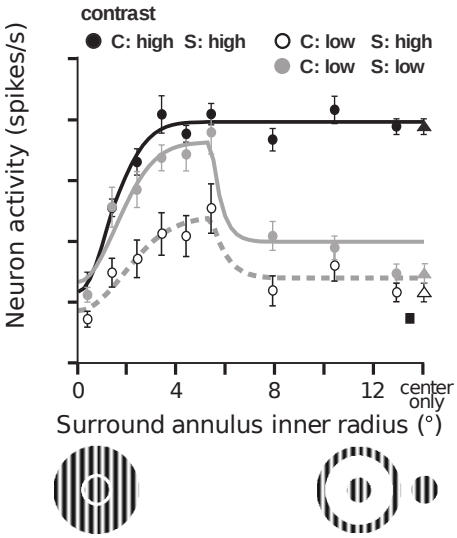


Figure 3.4 – The influence of surround on the centre, adapted from [14]. Response of a V1 cell in an anaesthetised macaque as a function of the inner radius of the surround annular grating. The triangles are responses to centre-only stimulation. The square indicates response to a surround stimulus of the smallest inner radius presented alone.

Physiological studies [239] also revealed that cortical RFs systematically increase their diameters by approximately a factor of three along the ventral stream, *i.e.* the visual pathway specialised in high-level tasks such as object recognition and form representation. This is due to the pooling mechanism of RFs from preceding areas, which combines signals from the central region as well as neighbouring spatial

locations. This suggests that local visual stimuli is processed in the lower cortical areas and the scope becomes increasingly global as the signal progresses throughout the pathway.

3.2.2 A model of contrast-dependent colour constancy

Surround modulation has been incorporated to biologically-inspired computer vision models with encouraging results, *e.g.* visual attention [128], saliency [172], tone mapping [192], and boundary detection [9]. However, in the field of computational colour constancy this important physiological finding seems to have been largely overlooked. In this section we investigate the implications of contrast-dependent centre-surround modulation on illuminant estimation by incorporating them into a simple and fully automatic model.

Primary visual cortex (V1)

We recreated a typical RF and its surround using two overlapping asymmetric Gaussian functions which have been reported to adequately fit neuronal responses, *e.g.* [45, 126, 203]. These functions, referred in our modelling context as the spatially "narrower" and "broader" Gaussians, represent the centre and surround respectively. The width of the narrower Gaussian varies between $[\sigma, 2\sigma]$ and is inversely proportional to the centre contrast. This mimics the changes in size that occur when the centre is exposed to high or low contrast and is similar to incorporating the dual-role region of Figure 3.3. Therefore, prior to convolving an image I with a Gaussian kernel, we compute local contrast C at every pixel through the local standard deviation of I as

$$C_{c,d}(x,y;\sigma) = \sqrt{\left(I_c(x,y) - I_c(x,y) * \mu_d(\sigma)\right)^2 * \mu_d(\sigma)}, \tag{3.4}$$

where c indexes each colour channel $\{R,G,B\}$; d is the spatial orientation $\{h,v,i\}$ (horizontal, vertical, and isotropic) over which contrast is measured; (x,y) are the spatial coordinates of a pixel; μ is the average kernel with size σ in the direction d and $*$ is the convolution operator. In the case of horizontal contrast, μ is a column vector; in the case of vertical contrast, μ is a row vector; and in the case of isotropic contrast, μ is a square matrix.

In our model, the receptive field's centre response CR is computed by convolution of the original image I at every chromatic channel c with the narrower Gaussian as follows:

$$CR_c(x,y) = I_c(x,y) * g_c(x,y;s_{c,h}(x,y),s_{c,v}(x,y)). \tag{3.5}$$

In Eq. 3.5, g is the two-dimensional Gaussian kernel defined as

$$g(x,y;\sigma_h,\sigma_v) = \frac{1}{2\pi\sigma_h\sigma_v}\exp\left(-0.5\left(\frac{x^2}{\sigma_h^2}+\frac{y^2}{\sigma_v^2}\right)\right), \tag{3.6}$$

where σ_d is the size of the Gaussian kernel in the direction d. The values of $s_{c,h}(x,y)$ and $s_{c,v}(x,y)$ in Eq. 3.5 represent the vertical and horizontal dimensions of the Gaussian kernel respectively. Since in our formulation the size of the RF's centre is inversely proportional to its local contrast (see Figure 3.3), we compute it from the values obtained in Eq. 3.4:

$$s_{c,d}(x,y) \propto C_{c,d}^{-1}(x,y;\sigma), \tag{3.7}$$

inversely linking the size of the RF's central kernel to its contrast. In theory, $s_{c,d}$ can be calculated for each individual pixel, however, in practice convolving an image with a unique Gaussian kernel at every pixel is extremely expensive from a computational point of view. For this reason, we approximated $s_{c,d}$ through its uniform quantisation into l different levels, effectively limiting the number of executed convolutions to l. We computed this uniform quantisation by finding the range of local contrasts through the difference between the two extrema of $s_{c,d}$ and dividing it into an arbitrary number of contrast levels. For example, let's assume that local contrasts are in the range $[0,1]$ and the arbitrary number of contrast levels is 4: pixels with local contrast between $[0.00,0.25]$ are convolved with a Gaussian of 2σ; pixels in the range $(0.25,0.50]$ with a Gaussian of 1.66σ; pixels in the range $(0.50,0.75]$ with a Gaussian of 1.33σ; and pixels in the range $(0.75,1.00]$ with a Gaussian of σ.

To summarise, we calculated the centre response CR by convolving low contrast image pixels with large Gaussians and high contrast image pixels with small Gaussians. It is worth noting that σ_h and σ_v in Eq. 3.6 are not identical (a common assumption in computer vision) due to the fact that the local interactions in V1 are not always organised in a symmetric fashion [233].

The RF's surround response, SR, was computed by convolution of the original image in every $\{R,G,B\}$ channel with the broader symmetric Gaussian kernel as follows:

$$SR_c(x,y) = I_c(x,y) * g_c(x,y;5\sigma,5\sigma), \tag{3.8}$$

where kernel size is constant in both directions regardless of local contrast. The decision of keeping the size of the SR kernel fixed was made after considering the much smaller variations that occur in the surround RFs of neurons under different

contrast levels [207].

The final RF response RR, was computed by combining centre and surround modulations as follows:

$$RR_c(x,y) = \lambda_c(x,y)CR_c(x,y) + \kappa_c(x,y)SR_c(x,y), \tag{3.9}$$

where λ and κ are the weights of centre and surround in each spatial location. These parameters model the fact that the strength of centre response and surround suppression depend of the contrast and relative orientations of the centre and surround stimuli (see Figure 3.4 and the work of Shushruth *et al.* [207]). We modelled λ and κ as inversely proportional to the oriented contrast of centre and surround respectively, which was computed as

$$\begin{aligned} \lambda_c(x,y) &\propto C^{-1}_{c,i}(x,y;\sigma); \\ \kappa_c(x,y) &\propto C^{-1}_{c,i}(x,y;5\sigma), \end{aligned} \tag{3.10}$$

where i denotes the spatial direction. We modelled the fact that suppression can turn into facilitation when the centre is exposed to low contrast or when centre and surround stimuli are orthogonal from each other [14]. This can be done by allowing the sign of κ to change from minus (suppressive surround) to the occasional plus (facilitatory surround) transforming our model from a DoG to Sum-of-Gaussians (SoG). Although our proposed model allows the possibility of a positive κ, we should note that the boundary between suppression and facilitation is cell specific and there is no universal contrast level or surround stimulus size that triggers facilitation across the entire cell population [14]. Due to this, and the fact that numerical surround suppression figures in macaque V1 neurons were reported to be all negative [207], the results we present in this chapter were all obtained with a negative κ value.

Area V4

Up to this point we implemented a model of RR based on well known properties of V1 neurons. In the next processing stage, the visual signal is pooled and sent to higher cortical areas whose exact location is unknown. A number of studies [51, 94] have proposed area V4 as the most likely candidate for a colour constancy site. We hypothesised the existence of V4 neurons that perform operations on the outputs of those in V1. From the physiology, we know that cortical RFs increase their diameter systematically by approximately a factor of three from lower to higher areas [239]. This means that V4 RFs are about nine times larger than those in V1 (which is 0.26°, see Figure 3.3). Therefore, the centre and surround of a typical V4 RF subtend

approximately to 2.3° and 11.7° of visual angle respectively, which are equivalent to 117 and 585 pixels, respectively, on a standard monitor viewed from a $100cm$ distance.

The exact pooling mechanism applied to these V1 signals is unknown, however "winner-takes-all" and "sparse coding" kurtotical behaviour are common to large groups of neurons all over the visual cortex [43, 175] and it is not infeasible to assume that a small group of neurons with the largest activation dominate most of the process. Some have modelled this "winner-takes-all" mechanism as a max-pooling operation [91]. However, one important flaw of this simple approach is that a single activated neuron can misrepresent the whole illuminant. Similar problem has been reported in the traditional White-Patch algorithm that may fail in the presence of noise or clipped pixels in the image due to the limitations of the max-pooling operator [87]. One approach to address these issues is to account for a larger set of "white" points by pooling a small percentage of the brightest pixels (*e.g.* the top 1%) [63], an operation referred as *top-x-percentage-pooling*. In this manner the pooling mechanism is collectively computed considering a group of pixels rather than an single one. Within this formulation it is very cumbersome to define a universal, optimally-fixed percentage of pixels to be be pooled [63] and consequently the free variable x requires specific tuning for each image or dataset individually. Naturally this limitation restricts the usability of the *top-x-percentage-pooling* operator.

We approximated this hypothetical behaviour of V4 neurons by selecting a small percentage of "winner neurons" whose RFs are highly activated. To simulate contrast adaptation behaviour in our hypothetical V4 neurons similar to those in V1, we inversely linked the percentage of pooled signals to the variability of the signal collected by their receptive field. In other words, when the "contrast" applied to V4 RF is high, a smaller percentage of signals from V1 is pooled and vice versa. As before, contrast was calculated as the local standard deviation of the input. Figure 3.2 summarises the whole feedforward process in a flowchart. The first stage of the model simulates the operation of the typical V1 neuron with contrast-dependent RFs and the second stage simulates the V4 sparse-coding pooling of a small percentage of highly activated V1 neurons.

Indeed such behaviour has been discovered across a population of cells in the cat visual cortex [142] and interestingly the activation level of cells with max-like behaviour was reported to vary depending on the contrast of visual stimuli. Results reported by [142] hint to an inverse relationship between the contrast of a stimulus and the percentage of the signal pooled. When pooling neurons were exposed to low contrast stimuli their responses shifted slightly away from pure max-pooling (selecting the highest activation response within a region) towards integrating over a larger number of highly activated neurons. In the language of computer vision,

this can be regarded as *top-x-percentage-pooling*, where x assumes a smaller value in high contrast and a larger value in low contrast. It is important to point out that the pooling of those neurons remained always much closer to max-pooling than to the linear integration of all neurons (sum-pooling) [142]. Mathematically, this can be interpreted as having a very small x value in the process of *top-x-percentage-pooling*.

In practice RR – V1 output – is an image composed by three chromatic channels (RR_c). We implemented the above explained "winner-takes-all" behaviour via a histogram-based clipping mechanism [63, 75] as follows. Let H_c denotes the histogram of RR_c values obtained by applying Eq. 3.9 to each colour channel c of the input image. In this histogram, the neural response of the cells contained in an individual bin b is represented by $RR_c(b_c)$. We estimate the scene illuminant by computing:

$$L_c = RR_c(b_c), \tag{3.11}$$

with b_c chosen such that only the most activated ("winner") units contribute to the pooling (sum). To calculate b_c we started by estimating the average local contrast of all inputs to V4 in a given colour channel c using

$$p_c = \frac{1}{n} \sum_{x,y} F_c(x,y), \tag{3.12}$$

where F is the standard deviation of the pixels of RR_c computed using the average V4 neuron receptive field (nine times larger than that of a V1 neuron), *i.e.* $F_c(x,y) \approx C_{c,i}(x,y;9\sigma)$. Bear in mind that "contrast" is just a fraction in the range $[0,1]$. Instead of choosing a fix percentage of neurons with the largest activation for each colour channel (as in [75]), we chose an adaptive activation level such that all neurons with activations higher than the one chosen account for fraction p_c of the total number of pixels. In other words, we computed b_c as the threshold activation level that defines a number of highly activated neurons equal to the contrast value calculated in Eq. 3.12 as follows:

$$p_c n \leq \sum_{k=b_c}^{n_b} H_c(k) \quad \text{and} \quad p_c n \geq \sum_{k=b_c+1}^{n_b} H_c(k), \tag{3.13}$$

where n is the total number of RR_c response units; and n_b represents the total number of bins in histogram H_c. This effectively links the number of highly activated neurons in our scene's illuminant estimation to the average contrast of the input to area V4.

We illustrated this contrast-dependent mechanism of V4 pooling in Figure 3.5,

where RR_c is represented by the red, green and blue signals corresponding to each chromatic channel. Dashed vertical lines show b_c, *i.e.* cells (bins) on the right side of these lines are pooled by our hypothetical V4 neuron and their sum for each colour channel is the estimated source of light. In this example we can observe that contrast is higher for the red signal and therefore a smaller percentage of cells are pooled in the red channel.

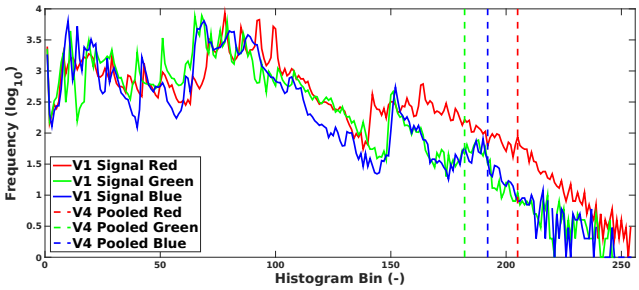


Figure 3.5 – V4 "winner-takes-all" mechanism. Each colour depicts its chromatic channel. Straight lines show which portion of V1 signals is pooled into V4. The ordinates are shown as logarithms to base 10 due to the large variations in counts of different bins.

Mathematically, there is a direct relation between the fraction of "winner" pixels, p in Eq. 3.12, and value of the Minkowski norm in Eq. 3.2. When the fraction of "winner" pixels is equal to unity (*i.e.* 100% pooling) our calculation in Eq. 3.11 includes the responses of all V1 neurons, resembling the Grey-World assumption. Recalling from earlier, this happens when the exponential term of the Minkowski sum in Eq. 3.2 is equal to unity. Correspondingly, when the percentage of "winner" pixels tends to zero, only the most activated V1 response is pooled, resembling the White-Patch algorithm.

3.3 Experiments and results

The issue of observer's performance evaluation in colour constancy tasks using naturalistic stimuli is still an open problem [81, 195]. In the case of algorithms, popular measures consist of some kind of angular distance in chromatic space between the estimated illuminant and that of the ground truth. Although intuitively simple, psychophysical experiments have shown that these error measures do not always correspond to observer preferences [229]. However, despite their shortcomings, angular errors are a convenient way to compare results among algorithms

and for this reason their use in the literature is widespread, being perhaps the most common the *recovery angular error* defined as

$$\epsilon^{\circ}_{recovery}(e_e, e_t) = cos^{-1}\left(\frac{e_e.e_t}{\|e_e\|.\|e_t\|}\right), \tag{3.14}$$

where $e_e.e_t$ is the dot product of the estimated illuminant e_e and the ground truth e_t, and $\|.\|$ represents the Euclidean norm of a vector. This simple measure has recently been the subject of criticism from Finlayson *et al.* [78] since it arguably produces different recovery errors for identical scenes viewed under two different coloured illuminants. For this reason, they proposed an improved version (termed *reproduction angular error*):

$$\epsilon^{\circ}_{reproduction}(e_e, e_t) = cos^{-1}\left(\frac{(e_t/e_e)}{\|e_t/e_e\|}.w\right), \tag{3.15}$$

where $w = \frac{e_t/e_t}{\sqrt{3}}$ is the true colour of the white reference.

In order to compare our results with those of state-of-the-art algorithms, we present the mean, median and trimean of both recovery and reproduction angular errors. The later two measures are considered to be more appropriate to assess the performance of colour constancy algorithms, because of their robustness to outliers [101, 118].

We evaluated our method on four benchmark datasets[1] without adjusting free parameters since ASM is fully automatic (*i.e.* dataset-independent) in contrast to most other algorithms whose results were acquired after adjusting their parameters to the optimum value for each dataset. Additionally, in order to better understand the contribution of the different components of our model, we conducted three extra experiments, which are explained later in this section.

3.3.1 Single-illuminant scenes

We tested our colour constancy model on three single-illuminant benchmark datasets, (i) SFU Lab [20], (ii) Colour Checker [205], and (iii) Grey Ball [49]. Our results for single-illuminant scenes were obtained under four contrast levels, $l = 4$, with $\sigma = 1.5$. This σ is equivalent to 13 pixels or 0.26° of visual angle when viewed from $100cm$ in a standard monitor, which is also the size of sRF_{high} (see Figure 3.3). We set the range of surround suppression to $\kappa = -[0.67, 0.77]$, considering the surround suppression index of macaque V1 neurons reported at [207]. The centre weight was retrieved directly from the contrast of pixels, $\lambda_c(x, y) = 1 + C^{-1}_{c,i}(x, y; \sigma)$.

[1] All source code and experimental materials are available under this link https://goo.gl/nQUenN.

SFU Lab

The SFU Lab dataset [20] consists of 321 images of size 637 × 468 captured in a controlled environment under eleven different sources of light. The scenes are partitioned into four categories: (a) minimal specularities, (b) non-negligible dielectric specularities, (c) metallic specularities, and (d) at least one fluorescent surface. We report the results of our method and several others on this dataset in Table 3.1. Our model's results show a clear improvement in the median and trimean angular errors (both reproduction and recovery) compared to state-of-the-art for the SFU Lab dataset.

Colour Checker

The Colour Checker dataset [95, 205] consists of 568 indoor and outdoor images of size 2041 × 1359. Each image contains a MacBeth colour-checker as a reference to retrieve the chromaticity of the actual source of light. We followed the best practices and guidelines of this dataset by masking out MacBeth colour-checker boards prior to processing an image with our model. The original images are non-linear due to gamma and tone curve correction. Shi and Funt [205] reprocessed the raw data and generated 12-bit images. We report the results of our method on this dataset along with several others in Table 3.2. The results show that our model is in par with the state-of-the-art for this dataset.

Grey Ball

The Grey Ball dataset [49] consists of 11,346 non-linear images of size 360 × 240 extracted from two hours of video recorded under a large variety of conditions in both indoor and outdoor environments. In every image there is a grey sphere at the bottom right corner from which the ambient illuminant can be estimated. We followed the best practices and guidelines of this dataset by masking out the grey spheres prior to processing an image with our model. We report the recovery and reproduction angular errors of our method on this dataset along with several others in Table 3.3. These results suggest that our model is in par with the learning-based state-of-the-art for this dataset, while it outperforms all other low-level features-driven methods.

3.3.2 Testing the role of each model component

We studied contribution of each component of our colour constancy model (*i.e.* , adaptive centre, dynamic surround and p estimation) by conducting three experiments and analysing their results in terms of median and trimean angular errors, proposed by Hordley and Finlayson [118] and Gijsenij *et al.* [101] as robust

		Recovery Error		
	Method	Mean	Median	Trimean
	Do Nothing	17.3	15.6	16.9
Low-level features	Inverse-Intensity Chromaticity Space [131]	15.5	8.2	10.7
	Grey-World [40]	9.8	7.0	7.6
	White-Patch [144]	9.1	6.5	7.5
	Shades of Grey [77]	6.4	3.7	4.6
	General Grey-World [77]	5.4	3.3	3.8
	First-order Grey-Edge [225]	5.6	3.2	3.7
	Second-order Grey-Edge [225]	5.2	2.7	3.3
	Local Surface Reflectance Statistics [92]	5.7	2.4	-
	Edge-based Grey Pixel [245]	5.3	2.3	-
	Double-Opponency [91]	4.8	2.4	3.5
Learning-based	Pixel-based Gamut Mapping [80]	3.7	2.3	2.5
	Edge-based Gamut Mapping [102]	3.9	2.3	2.7
	Spectral Statistics [46]	5.6	3.5	4.3
	Weighted Grey-Edge [103]	5.6	2.4	2.9
	Regression [89]	-	2.2	-
	Thin-plate Spline Interpolation [206]	-	2.4	-
	ASM	**4.7**	**1.8**	**2.3**

		Reproduction Error		
	Method	Mean	Median	Trimean
	Do Nothing	17.3	15.6	16.9
Low-level features	Inverse-Intensity Chromaticity Space [131]	15.1	9.3	11.5
	Grey-World [40]	10.1	7.5	8.3
	White-Patch [144]	9.7	7.4	8.2
	Shades of Grey [77]	6.9	3.9	4.8
	General Grey-World [77]	6.0	3.9	4.3
	First-order Grey-Edge [225]	6.3	3.6	4.2
	Second-order Grey-Edge [225]	5.8	3.0	3.8
Learning-based	Pixel-based Gamut Mapping [80]	4.2	2.8	3.0
	Edge-based Gamut Mapping [102]	4.5	2.7	3.2
	Weighted Grey-Edge [103]	6.1	3.6	4.3
	ASM	**5.2**	**2.3**	**2.7**

Table 3.1 – Angular error of several methods on SFU Lab [20] benchmark dataset. Table on top corresponds to the recovery angular errors. Table on bottom corresponds to the reproduction angular error. Lower figures indicate better performance.

			Recovery Error		
	Method		Mean	Median	Trimean
	Do Nothing		13.7	13.6	13.5
Low-level features	Grey-World	[40]	6.4	6.3	6.3
	White-Patch	[144]	7.5	5.7	6.4
	Shades of Grey	[77]	4.9	4.0	4.2
	General Grey-World	[77]	4.7	3.5	3.8
	First-order Grey-Edge	[225]	5.3	4.5	4.7
	Second-order Grey-Edge	[225]	5.1	4.4	4.6
	Random Sample Consensus	[88]	3.2	2.3	-
	Edge-based Grey Pixel	[245]	4.6	3.1	-
	Double-Opponency	[91]	4.0	2.6	-
Learning-based	Pixel-based Gamut Mapping	[80]	4.2	2.3	2.9
	Edge-based Gamut Mapping	[102]	6.5	5.0	5.4
	Regression	[89]	8.1	6.7	7.2
	Bayesian	[95]	4.8	3.5	3.9
	Natural Image Statistics	[100]	4.2	3.1	3.5
	Exemplar-based method	[131]	2.9	2.3	2.4
	CNN Fine Tuned	[29]	2.6	2.0	-
	Deep Learning Colour Constancy	[155]	3.1	2.3	-
	ASM		**3.8**	**2.4**	**2.7**

			Reproduction Error		
	Method		Mean	Median	Trimean
	Do Nothing		13.7	13.6	13.5
Low-level features	Grey-World	[40]	7.0	6.8	6.9
	White-Patch	[144]	8.1	6.5	7.1
	Shades of Grey	[77]	5.8	4.4	4.9
	General Grey-World	[77]	5.3	4.0	4.4
	First-order Grey-Edge	[225]	6.4	4.9	5.3
	Second-order Grey-Edge	[225]	6.0	4.8	5.2
Learning-based	Pixel-based Gamut Mapping	[80]	4.8	2.7	3.4
	Edge-based Gamut Mapping	[102]	8.0	5.9	6.6
	Regression	[89]	8.8	7.4	7.9
	Bayesian	[95]	5.6	3.9	4.4
	Natural Image Statistics	[100]	4.8	3.5	3.9
	Exemplar-based method	[131]	3.4	2.6	2.9
	ASM		**4.9**	**3.0**	**3.4**

Table 3.2 – Angular error of several methods on Colour Checker [205] benchmark dataset. Lower figures indicate better performance.

			Recovery Error		
	Method		Mean	Median	Trimean
	Do Nothing		8.3	6.7	7.2
Low-level features	Inverse-Intensity Chromaticity Space	[131]	6.6	5.6	5.8
	Grey-World	[40]	7.9	7.0	7.1
	White-Patch	[144]	6.8	5.3	5.8
	Shades of Grey	[77]	6.1	5.3	5.5
	General Grey-World	[77]	6.1	5.3	5.5
	First-order Grey-Edge	[225]	5.9	4.7	5.1
	Second-order Grey-Edge	[225]	6.1	4.8	5.3
	Local Surface Reflectance Statistics	[92]	6.0	5.1	-
	Edge-based Grey Pixel	[245]	6.1	4.6	-
Learning-based	Pixel-based Gamut Mapping	[80]	7.1	5.8	6.1
	Edge-based Gamut Mapping	[102]	6.8	5.8	6.0
	Spectral Statistics	[46]	10.3	8.9	9.1
	Natural Image Statistics	[100]	5.2	3.9	4.3
	Exemplar-based method	[131]	4.4	3.4	3.7
	Deep Learning Colour Constancy	[155]	4.8	3.7	-
	ASM		**4.7**	**3.8**	**4.0**

			Reproduction Error		
	Do Nothing		8.3	6.7	7.2
Low-level features	Inverse-Intensity Chromaticity Space	[131]	7.0	6.0	6.2
	Grey-World	[40]	8.7	7.6	7.9
	White-Patch	[144]	7.1	5.5	6.0
	Shades of Grey	[77]	6.5	5.6	5.8
	General Grey-World	[77]	7.1	6.2	6.4
	First-order Grey-Edge	[225]	6.3	4.8	5.4
	Second-order Grey-Edge	[225]	6.5	5.0	5.6
Learning-based	Pixel-based Gamut Mapping	[80]	7.5	5.9	6.3
	Edge-based Gamut Mapping	[102]	7.3	5.8	6.3
	Natural Image Statistics	[100]	5.5	4.3	4.7
	Exemplar-based method	[131]	4.8	3.7	4.0
	ASM		**5.0**	**4.1**	**4.3**

Table 3.3 – Angular error of several methods on Grey Ball [49] benchmark dataset. Table on top corresponds to the recovery angular errors. Table on bottom corresponds to the reproduction angular error. Lower figures indicate better performance.

measures to evaluate colour constancy algorithms.

Experiment 1 – constant vs. adaptive centre size

In order to measure contribution of the adaptive size of the narrower Gaussian, we kept all other parameters fixed (*i.e.* the centre-surround influence, $\lambda = 1.00; \kappa = -0.77$, and the percentage of pooled signal, $p = \infty$). We investigated two scenarios: (a) all pixels were convolved with a constant Gaussian of width σ (essentially the Double-Opponency algorithm [91]), whereas, in (b) this width was varied in the range of $[\sigma, 2\sigma]$ and computed for each pixel. These two conditions were called "Constant Gaussian Width" (CGW) and "Adaptive Gaussian Width" (AGW). Additionally, since the Grey-Edge hypothesis captures high-order image features similar to the DoG, we tested whether this centre adaptation can improve the first and second order Grey-Edge algorithm with a Minkowski norm $p = \infty$.

The results of experiment 1 (see Figure 3.6) show that both criteria of median and trimean angular errors are always smaller in the adaptive case (AGW) in comparison to the constant one (CGW). This is true for both measures of recovery and reproduction angular errors. The largest and smallest improvements are achieved in the SFU Lab (about 19% on average) and Grey Ball (about 6% in average) datasets, respectively.

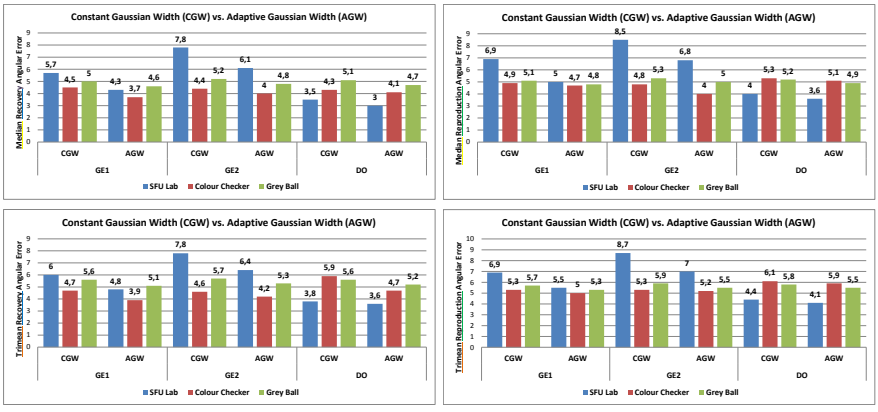


Figure 3.6 – Influence of contrast-dependent RF size on illuminant estimation. Panels on the left correspond to the recovery angular error while those on the right are reproduction angular error. Panels on the top show median and those on the bottom the trimean angular error.

Experiment 2 – constant vs. adaptive surround

In order to measure contribution of the adaptive surround modulation, we kept all other parameters fixed (*i.e.* the centre adaptation, $l = 1$, and the percentage of pooled signal, $p = \infty$). We tested three scenarios, the first and second were computed under a constant surround influence, $\kappa = -0.67$ and $\kappa = -0.77$, respectively (both extrema of our adaptive κ), as well as constant centre weight, $\lambda = 1.00$. In the third scenario, the centre-surround influence was adaptive, $\lambda = 1 + C^{-1}_{c,i}(x, y; \sigma)$ and $\kappa = -[0.67, 0.77]$, under four contrast levels $l = 4$.

Figure 3.7 shows the results for Experiment 2, where the median and trimean errors (both recovery and reproduction) obtained with a dynamic surround suppression, $\kappa = -[0.67, 0.77]$, are always lower in comparison to the constant κ. The gain across datasets appear to be similar (around 3% for both error measures).

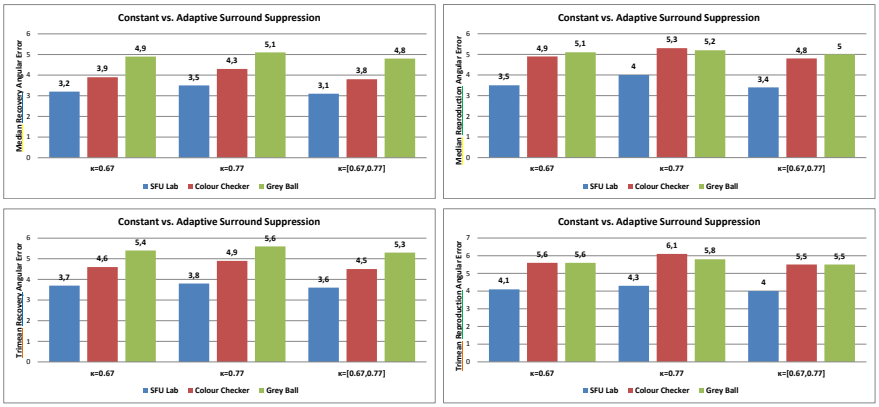


Figure 3.7 – Influence of contrast-dependent surround suppression on illuminant estimation. Panels on the left correspond to the recovery angular error while those on the right are reproduction angular error. Panels on the top show median and those on the bottom the trimean angular error.

Experiment 3 – constant vs. adaptive "winners" percentage

In order to measure contribution of the adaptive clipping, we examined five different scenarios. In the first four, histograms (see Eq. 3.13) were clipped with constant percentages, $p = \{5, 1, 0.5, 0.1\}\%$, *i.e.* a fixed set of V1 cells were pooled into V4. In the fifth case, value of p was adaptive and computed as the average contrast of RR (see Eq. 3.12).

The results of Experiment 3 (see Figure 3.8) show that using a contrast-adaptive pooling mechanism reduces the recovery/reproduction angular errors in all cases considered in the SFU Lab dataset (blue bar with $p = \bar{c}$ is smaller than all the others). In the Colour Checker and Grey Ball datasets (red and green bars respectively), estimating p adaptively yields angular errors very close to the best constant p values. Among the constant clipping percentages $p = 0.5\%$ performs best: moving towards a Grey-World pooling deteriorates the results ($p = 5\%$ obtain the highest angular errors) and moving towards a White-Patch solution also worsens angular errors ($p = 0.5\%$ always performs better than $p = 0.1\%$). This suggests the optimal pooling mechanism is close to our proposal of pooling a set of highly activated cells. A comparison of the best fixed p ($= 0.5\%$) and adaptive p ($= \bar{c}$) shows a 4% improvement of median and trimean errors (average of all three datasets) in the case of adaptive p.

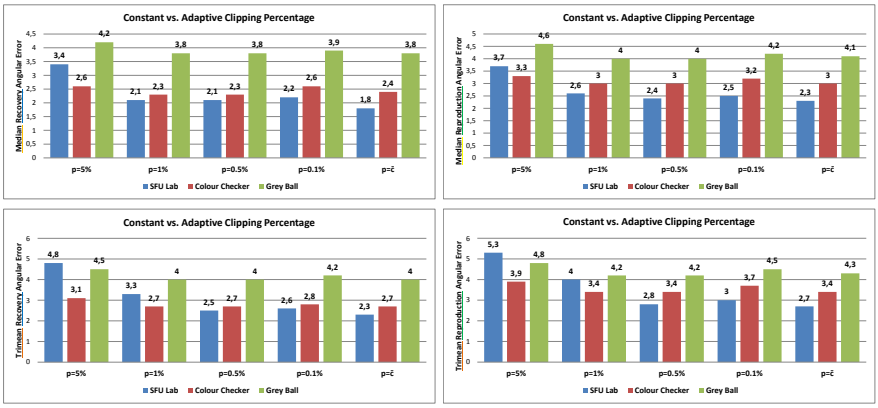


Figure 3.8 – Influence of "winners" percentage p on illuminant estimation. Panels on the left correspond to the recovery angular error while those on the right are reproduction angular error. Panels on the top show median and those on the bottom the trimean angular error.

3.3.3 Multi-illuminant scenes

We also tested the proposed model on one multi-illuminant benchmark dataset [22] which consists of 78 images. Each image is captured under the illumination of two different artificial sources of light. This dataset contain two set of images: (a) laboratory (58 images of size 452×260) and (b) real-world images (20 images of size

452 × 302).

The extension of our model to multi-illuminant scenes is straightforward by modelling each region or pixel with a similar contrast-dependent pooling mechanism (Eq. 3.11, 3.12,3.13 will be region or pixel dependent). This solution is biologically-plausible as different V4 neurons pool signals from different V1 neurons. For this multi-illuminant dataset we used the exact parameters as single-illuminant datasets (refer to Section 3.3.1). Here we defined four simple image regions (by halving the image in both horizontal and vertical directions) and computed the source of light in each region accordingly. These results are reported alongside several others in Table 3.4. Since other methods have not reported their respective trimean and reproduction angular errors in this dataset, we only report the mean and median recovery angular error. Our results are competitive with the state-of-the-art.

		Laboratory		Real-world	
Method		Mean	Median	Mean	Median
Do Nothing		10.6	10.5	8.9	8.8
Grey-World	[40]	3.2	2.9	5.2	4.2
White-Patch	[144]	7.8	7.6	6.8	5.6
First-order Grey-Edge	[225]	3.1	2.8	5.3	3.9
Second-order Grey-Edge	[225]	3.2	2.9	6.0	4.7
Gijsenij *et al.*	[104]	4.8	4.2	4.2	3.8
Double-Opponency	[91]	4.6	4.4	7.8	4.9
STD-based Grey Pixel	[245]	2.9	2.2	5.7	3.5
MI Random Field	[22]	2.6	2.6	4.1	3.3
ASM		**2.7**	**2.5**	**5.1**	**3.5**

Table 3.4 – Recovery angular error of several methods on Multi-illuminant [22] benchmark dataset. Lower figures indicate better performance.

3.4 Discussion

Figure 3.9 illustrates results of our *Adaptive Surround Modulation (ASM)* colour constancy model alongside three other algorithms on four exemplary images (one from each of the benchmark datasets considered) captured under different illumination sources: “synthetic indoor”, “natural daylight”, “dim evening”, and “multi-illuminant”. The qualitative results demonstrate that ASM can efficiently estimate the present source of light in synthetic and natural images, bright and dark environ-

ments, and in both single- and multi-illuminant scenes. We believe that self-similar dynamical properties of ASM, both at local and global level explain why our *fully automatic* model, with no training required, can adapt itself to each environment and therefore recover the illuminant in a wide range of scenarios and illumination conditions.

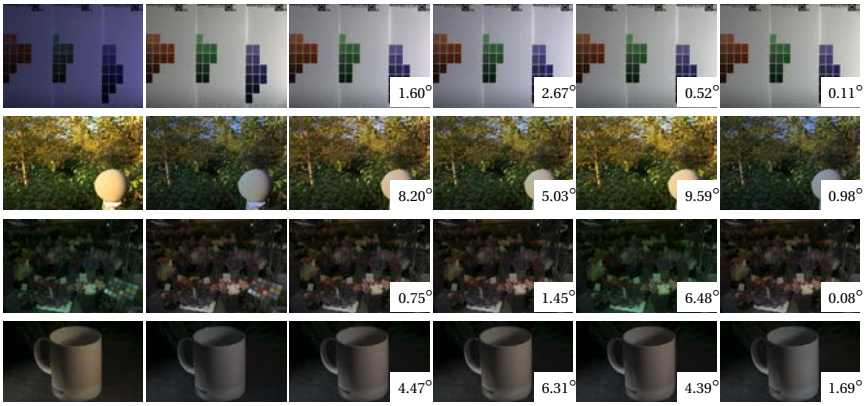


Figure 3.9 – Colour constancy results of several methods. The recovery angular error is indicated on the right bottom corner. The first row shows results for a picture from the SFU Lab dataset, the second row from the Grey Ball dataset, the third row from the Colour Checker dataset, and the last row from the Multi-illuminant dataset.

The quantitative results in Table 3.1 show that ASM outperforms all other state-of-the-art algorithms in the SFU Lab dataset. In the Grey Ball dataset (Table 3.3), ASM performs the best amid methods driven by low-level features and obtains comparable results to the learning-based techniques. In the Colour Checker and Multi-illuminant datasets (Tables 3.2 and 3.4 respectively), our results are highly competitive with the best learning ones. Considering the fact that, unlike our competitors, we are using a fix set of parameters for all four datasets, our results look promising indeed.

A quick comparison among Tables 3.1–3.3 and Figure 3.6, shows that the colour constancy methods driven by the higher-order image statistics (*e.g.* Grey-Edge and Double-Opponency), are highly sensitive to their choice of parameters. For example, in the SFU Lab dataset, the median recovery angular error of the second order Grey-Edge (GE2) escalates from 2.7° (Table 3.1) to 7.8° (Figure 3.6) under the

optimum ($p = 7, \sigma = 4$) and non-optimum parameters ($p = 1, \sigma = 1$) respectively. This is not the case for our fully automatic method. The angular error of ASM across datasets is less variable than that of most of its competitors. This is a yet another sign of robustness and implies that ASM adapts based on the contrast of an image independently of previous history, much in the same way as the HVS does.

The results of experiment 1 (see Figure 3.6) show that the performance of colour constancy methods driven by the high-order image statistics (*e.g.* Grey-Edge and Double-Opponency) can be improved, as much as 21%, by adapting their Gaussian width σ based on local contrast at pixel level. As discussed in the introduction, this does not come as a surprise, given that the high-order derivatives are similar to those of the centre-surround mechanism present in biological visual systems, where the RF size expands in the presence of low contrast and shrinks in high contrast. The improvement originated from the AGW appears to be largest for the Grey-Edge (about 13% on average) than for the Double-Opponency (about 7% on average). This could be explained by the fact that the centre-surround contrast adaptation requires both dynamic centre and dynamic surround. In the Grey-Edge centre-surround is modelled in one operation, whereas in the Double-Opponency neither the surround size nor its contribution change according to the contrast level.

The results of experiment 2 (see Figure 3.7) demonstrate that contrast-dependent surround modulation can improve the angular errors up to 15%, however the average improvement is a more modest figure of about 3%. This is explained by the fact that surround modulation depends on number of other parameters in addition to the local contrast of stimuli, such as spatial frequency and orientation. In this chapter, we limited our studies to the role of contrast on surround modulation and therefore the range of surround suppression we could explore was rather limited to $\kappa = -[0.67, 0.77]$. However, we believe our results can be improved even further by taking into account the orientation selectivity of surround suppression and consequently allowing a larger range of κ values. This way ASM can oscillate between DoG to SoG to account for both surround inhibition and facilitation. This can be achieved for example by wavelet decomposition, which we propose as future work. Such pyramids of wavelets have been successfully used to model the operation of neurons in the visual cortex in the case of contrast induction [179] and saliency [172].

Interestingly, in both experiments 1 and 2, implementing a contrast-dependent centre-surround never deteriorates the results and it always systematically reduces angular errors, even if this reduction is minimal. Conceptually, our contrast-dependent centre-surround is intuitive: on homogeneous regions a larger window must be applied to represent true surround variation, whereas on heterogeneous regions a small neighbourhood suffices. Similar types of contrast-dependent mod-

ulation have shown to boost true edges while suppressing undesired textural information [7]. Theoretically, our variations of the Gaussian kernel width, σ, are resemblant of processing an image through a Gaussian pyramid (although not of fixed one-octave log increments in size, like those found in the cortex). Correspondingly, our variations of the influence of surround, κ, resembles a Laplacian pyramid.

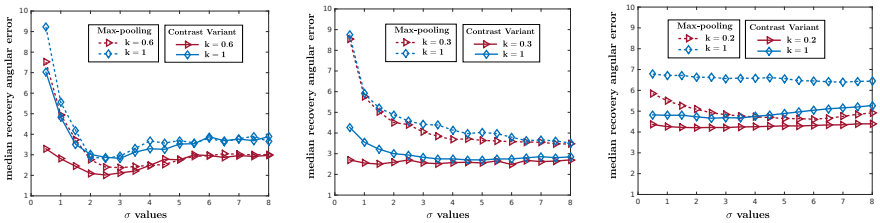
The results of experiment 3 (see Figure 3.8) also indicate that our "winner-takes-all" hypothesis appears to be correct. The lowest angular errors are obtained when only a small percentage of V1 signals are pooled into V4 and when this percentage is high (5%) the results deteriorate significantly. However, there is no unique p to minimise the angular errors across different datasets for both measures of median and trimean. Determining the "winners" according to the average contrast of V1 RFs ($p = \bar{c}$) produces the lowest angular errors across datasets. Conceptually, in a low contrast image a few bright pixels can hint the source of light, whereas in a high contrast image (*i.e.* with high variation of pixel values) more samples are required to determine the scene illuminant. This is in line with the results of Joze *et al.* [132], which indicate that bright pixels play a vital role in illuminant estimation. A better estimation of p might be obtained by a more thorough modelling of V4 neurons (for example by calculating p in different image regions, rather than the entire population of V1 neuron).

3.4.1 Contrast variant pooling colour constancy

Triggered by results of experiment 3 we further investigated the efficiency of the proposed "winner-takes-all" hypothesis when applied to three different colour constancy algorithms driven by the high-order image statistics: the first order Grey-Edge [225], the second order Grey-Edge [225], and Double-Opponency [91]. We simply replaced their max-pooling operator with our proposed contrast-variant-pooling (CVP) mechanism at area V4. We are assuming that features maps of those models can be interpreted as the V1 output.

For each free variable of investigated models we compared performance of max- to contrast-variant-pooling. In Figure 3.10 we have reported the impact of different σs (receptive filed size) on the Double-Opponency algorithm for the best and the worst results obtained by free variable k in each dataset. We can observe that almost in all cases contrast-variant-pooling obtains lower angular errors in comparison to max-pooling. The improvement is more tangible for the Colour Checker and Grey Ball datasets and in low σs.

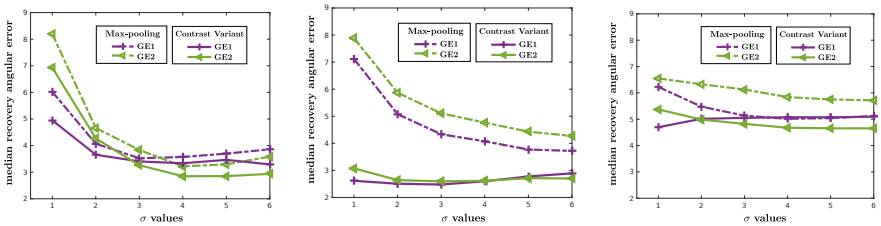
Figure 3.11 illustrates the impact of different σs (Gaussian size) on the first- and second-order Grey-Edge algorithm. We can observe similar patterns as with Double-Opponency (contrast-variant-pooling outperforms max-pooling practically in all



SFU Lab [20] Colour Checker [205] Grey Ball [49]

Figure 3.10 – The best and the worst results obtained by max- and contrast-variant-pooling mechanism for free variables of the Double-Opponency [91] algorithm (k and σ).

cases). This improvement is more significant for low σs, for the Colour Checker dataset and for the second-order derivative. It must be noted here that our objective is to study the performance of max-pooling and CVP on top of the Grey-Edge algorithm. In this respect, *CVP Grey-Edge* angular errors are on par with the best reported results for max-pooling Grey-Edge using Minkowski norm optimisation for each dataset [225], With the important caveat that CVP has no extra variables to be tuned, whereas in the Minkowski norm optimisation the value of p must be hand-picked for each dataset.



SFU Lab [20] Colour Checker [205] Grey Ball [49]

Figure 3.11 – Comparison of max- and contrast-variant-pooling mechanism for free variable σ of the Grey-Edge [225] algorithm (both first- and second-order derivatives).

From Figures 3.10 and 3.11 we can observe that the greatest improvement occurs in the Colour Checker dataset. We speculate that one of the reasons for this is the larger range of intensity values in the Colour Checker dataset (16-bit) in comparison to the other two datasets that contain 8-bit images, therefore, an

inaccurate max-pooling is greatly penalised.

Physiological evidence besides, the better performance of CVP can be explained intuitively by the fact that max-pooling relies merely on the peak of a function (or a region of interest), whereas in our model, pooling is defined collectively based on a number of elements near the maximum. Consequently those peaks that are outliers and likely caused by noise get normalised by other pooled elements. The rationale within our model is to pool a larger percentage at low contrast since in those conditions, peaks are not informative on their own, whereas at high contrast peaks are likely to be more informative and other irrelevant details must be removed (therefore a smaller percentage is pooled).

Although the importance of choosing an appropriate pooling type has been demonstrated both experimentally [130, 243], and theoretically [32], current standard pooling mechanisms lack the desired generalisation [171]. We believe that contrast-variant-pooling can address this problem by offering a more dynamic and general solution. In this chapter, we evaluated the performance of CVP on the colour constancy phenomenon as a proof-of-concept, however our formulation of CVP is generic (and based on local contrast) and in principle can be applied to a wider range of computer vision algorithms, such as deep-learning, where pooling is a decisive factor [198].

3.4.2 Mondrian images

To mitigate the influence of higher-level visual cues, we tested our algorithm with the exact same parameters on 1000 randomly generated Mondrian images under randomly generated illuminants. Median reproduction angular error of adaptive surround modulation (our full model) was 2.3; the same measure for our model in its constant form (*i.e.* no contrast-dependent V1 and V4 neurons, similar to Gao *et al.* [91]) was 3.8. In more than 77% of the images, adaptive surround modulation obtains better results in comparison to the constant one. In order to investigate whether level of cluttering in a scene is an important factor in our model, we repeated this experiment with different number of Mondrians in each image. We did not notice any correlation between number of Mondrians and performance of our model.

In Figure 3.12 we have illustrated one example of conducted experiment with Mondrian images. If we compare the pictures corresponding to “Constant V1” and “Adaptive V1”, we can observe that in case of constant centre-surround modulation the picture becomes blurrier at every pixels, whereas a contrast-dependent formulation allows for sharper edges. Similarly, in case of “Constant V4” the estimated illuminant is significantly greener than the actual illuminant and therefore the corrected image appears reddish.

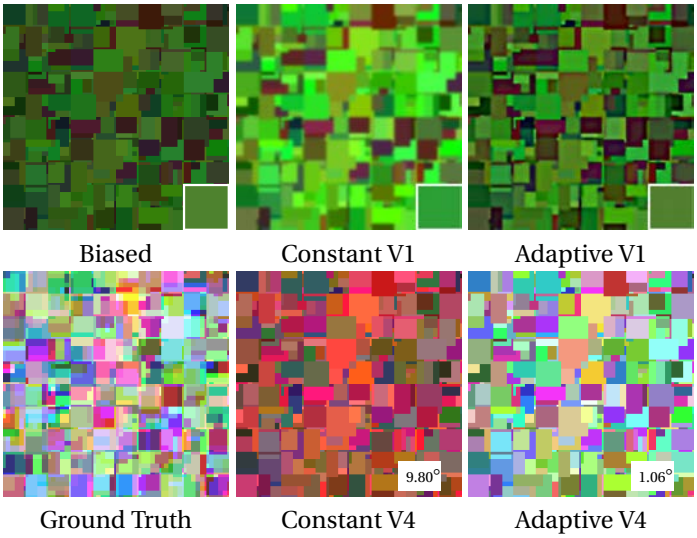


Figure 3.12 – Constant versus adaptive V1 and V4 modules. Colour patches on the right bottom corner of images in the first row depict the ground truth illuminant in case of biased image and estimated illuminants in case of constant and adaptive results.

3.4.3 Computational complexity

Computationally, the proposed colour constancy model is very efficient since no training is required. Furthermore, the backbone of ASM is only simple convolutional operators. The complexity of our algorithm is l (number of contrast levels, 4 in this chapter) times more expensive than a simple DoG. However, each level is 100% independent and their convolutions can easily run in parallel, as it is implemented in our source code.

3.5 Conclusion

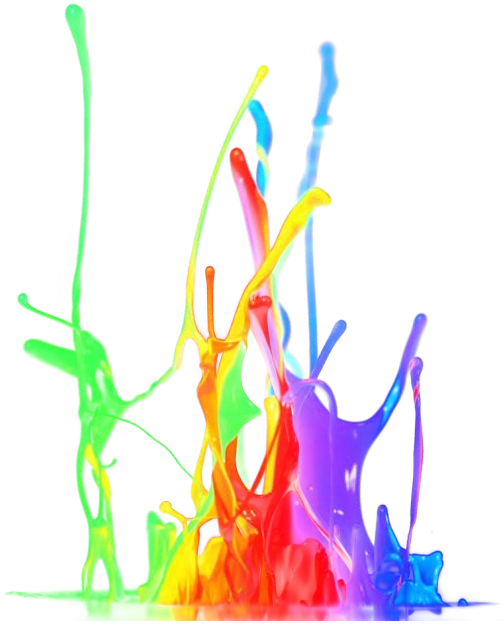
It has been demonstrated that global and local contrast greatly influences the appearance of colours in a scene [81, 125]. In this chapter, we show that adopting some of the computations that evolved in the human visual system after millions of years of evolution into a simple, functional colour constancy model allows us to obtain results on par to much more complex computational learning approaches.

The mechanisms in question are three: (i) adaptation of receptive field size depending on local contrast, (ii) influence of surround-on-centre also according to local contrast, and (iii) computation of global contrast in higher visual areas to produce the final illuminant estimation. Their particular contributions were quantified by performing additional experiments. We compared our results to current state-of-the-art algorithms in four benchmark datasets showing a significant improvement regarding other low-level feature-driven methods, while still highly competitive with respect to the best learning-based methods.

The significance of this performance is evident considering that our model is (a) *fully automatic* and *parameter-free* (*i.e.* it does not require learning the properties of each dataset since all its initial variables are set at the beginning) (b) *parsimonious* (it follows basic simplicity principles such as Occam's razor) and (c) *biologically-inspired* on well established findings within the neurophysiology and visual perception communities. These properties make it an excellent choice to be implemented in small image-gathering devices such as webcams and mobile phones. Furthermore, ASM does not only provides a good solution to the *engineering* problem of removing the illuminant in images, but, because of its close links to the properties of cortical neurons allows us to speculate on the *scientific* question regarding the evolutionary role of these properties of the visual system, something that other algorithms are unable to do.

As a final note, we would like to express our conviction that complex multi-dimensional problems such as colour constancy cannot be solved by one-fits-all solutions. In other words, the results of fully automatic solutions should not be interpreted the same as those of learning-based solutions. Our view is that these belong to different and sometimes orthogonal directions and should be considered according to their own particular merits.

Colour Names Part II



How do we categorise colours of a scene?

4 Colour Categorisation

During the last two chapters, we discussed how a visual system perceives colours of objects under different illuminants and the implications of this for higher-level visual tasks. In this chapter, we focus on colour names. Although, there are no discontinuities in the electromagnetic spectrum of the light reaching us from a rainbow, yet we see hues clearly separated by colour categories. Our brains categorise colours into distinct semantic categories that are used not only to describe objects but also to facilitate parts of the processing, *e.g.* in finding objects in cluttered conditions, such as, car parks and bookshelves. Here, we investigate whether adding knowledge from the other disciplines, such as, biology and psychophysics may help us to improve algorithms that obtain colour names from coloured pixels.

4.1 Introduction

Colour vision contributes significantly to our perception of the world by providing valuable information about properties of objects and facilitating their segmentation from each other and the background [36]. Its evolution might be guided by ecologically important tasks such as collecting ripe fruits and leaves or spotting predators. Besides that, our brains have further evolved to communicate the perception of colour through natural language. As a consequence of that, colour terms are extensively used in our day-to-day life in a wide range of scenarios. For instance, we tend to describe objects by their colour names (*e.g.* pass me the blue book; look at that orange house; *etc.*). Moreover, we explicitly benefit from colours to facilitate various tasks (*e.g.* software programmers colour-code their source code to aid interpretation; pedestrians and drivers rely on colour-coded city traffic lights to avoid chaos; *etc.*).

Consequently, any computer application seeking to intuitively interact with humans (*e.g.* visual searching, image labelling, content retrieval, *etc.*) may benefit from incorporating colour naming in its routine [227]. Furthermore, numerous computer vision algorithms (such as scene segmentation, high-dynamic-range imaging (HDR), target tracking, object recognition, texture classification, *etc.*) can greatly benefit from the segmentation of an image to its constituent colours: either by improving their accuracy or lowering their computational complexity [227].

Despite the omnipresence of colour in our lives and the prominent role played by our perceptual machinery, only a handful of computational colour naming models has been developed and even fewer of them attempt to incorporate our knowledge of the perceptual system into them.

Colour naming (also referred here as "colour categorisation") is a highly multidisciplinary topic. A large-scale linguistic survey by anthropologists Berlin & Kay [25] hinted at eleven basic colour terms – *i.e.* black, blue, brown, green, grey, orange, pink, purple, red, white, and yellow – that are shared across most evolved languages and cultures. Universality of these colour terms has been challenged by the role of linguistic contexts [111]. Nevertheless, they have been reconfirmed in various other studies [33, 137, 217] and to a certain extent explained by physiological evidence that demonstrate low-level mechanisms contribute to colour categorical perception prior to language acquisition [210]. Present general consensus favours an intermediate free-from-language low-level colour perception stage supported by non-verbal cognitive experiments [127].

Colour naming at first might appear to be fully deterministic (indeed a few computational models have taken this approach [151, 222]). However, Kay & McDaniel [136] suggested that the determining perceptual input comes from the language-processing part of the brain. Therefore the underlying visual mechanism behind colour naming must be modelled by continuous mathematics, *i.e.* fuzzy logic. This insight (also supported by psychophysics) implies that in practice every pixel has a value of "belongingness" (from zero to hundred per cent) to each colour category which is directly computed from the measured reflectance spectrum of a surface at that point.

Initial works on fuzzy models started with Lammens [64], who fitted the data collected by Berlin & Kay into some variations of Gaussian functions. Mojsilovic [168] followed this approach with a new perceptual colour metric. Seaborn *et al.* [199] clustered psychophysical colour points with a k-means algorithm while Benavente *et al.* [24] tackled the problem by means of a triple-sigmoidal parametric model, with a few lightness planes sliced into different colour categories and the rest approximated through interpolation. Contrary to previous algorithms that are based on fitting colour categories to psychophysically obtained *focal colours*, van de Weijer *et al.* [228] proposed a new procedure to learn colour names from real-world images using probabilistic latent semantic indexing.

Our proposal in this dissertation to capture colour terms using simple geometrical shapes is fundamentally different from current methods: (i) we benefit from parametric modelling [24, 64] with the added advantage of partitioning the colour space directly into three-dimensional shapes rather than interpolating from two-dimensional planes; (ii) unlike some algorithms that learn every pixel independently through histograms with no explicit constraints on colour regions [228],

we impose ellipsoidal shapes that function as natural restrictions to such colour regions.

Acknowledging the fact that concept of colour is a product of our brain, it naturally follows that the best way to address colour naming is to model what we know physiologically and psychophysically about the human cortical machinery. For example, it is widely accepted that colour categorisation has been shaped by evolution and neonatal adaptation to break down an extremely complex world into cognitively tractable entities, reducing the nearly two million colours that can be distinguished perceptually [187] to about thirty categories than can be recalled by average subjects [55]. In particular, the elven universal colour categories [25] are unlikely to be arbitrary and possibly reflect ideal divisions of an irregularly shaped perceptual color space [191]. In our conducted psychophysical experiment [6, 186], we observed that in chromatically opponent space categorical frontiers between these eleven universal colours form ellipsoidal shapes in agreement with the elliptical isorresponses of V1 neurons reported in a physiological study by Horwitz & Hass [119].

Following this rationale, in this chapter we present a biologically-inspired colour naming model based on an "ideal" partitioning of colour-opponent space (as suggested by Regier *et al.* [191]) through parsimonious ellipsoidal shapes (as revealed by psychophysics [186] and physiology [119]). We extend the work of [186] by: (i) demonstrating that parameters of ellipsoids and growth ratio can be learnt more ecologically from segmented images; (ii) accounting for rotation along each axis and all ellipsoids; (iii) showing that it is straightforward to incorporate new colour terms within the new framework; (iv) prototyping the means of ellipsoids adaptation to the image contents in order to account for the phenomenon of colour constancy; and (v) testing our model on real-world images.

4.2 Ellipsoidal colour categorisation model

In this section: (i) we review relevant physiological and psychophysical facts about colour vision and colour naming; (ii) we detail theory of proposed model; and (iii) we explain different means of obtaining parameters of our colour categorisation model.

4.2.1 Colour perception

At present, we have a fairly rigorous understanding of cone photoreceptors that initiate colour vision by absorbing light at the back of retina. Signals produced by these cells are combined in an antagonistic manner to from the opponent channels

that convey information to the visual cortex through the lateral geniculate nucleus (LGN) [68, 184]. As we advance deeper inside cortical areas, our knowledge of cerebral mechanisms involved in colour vision becomes less clear. In the primary visual cortex (V1), there is population of specialised neurons called single- and double-opponent cells that respond non-linearly to chromatic stimulus [203]. A recent study by Horwitz & Hass [119] analysed neurons in V1 in terms of their uniform responses to three-dimensional shapes in colour-opponent space. A large subset of these neurons (termed Neuron-3) responded best to ellipsoids whose major and minor axes are aligned to perceptual cardinal directions (see the schematics in Figure 4.1). These findings by [119] show how neurons of V1 can act jointly to process colour.

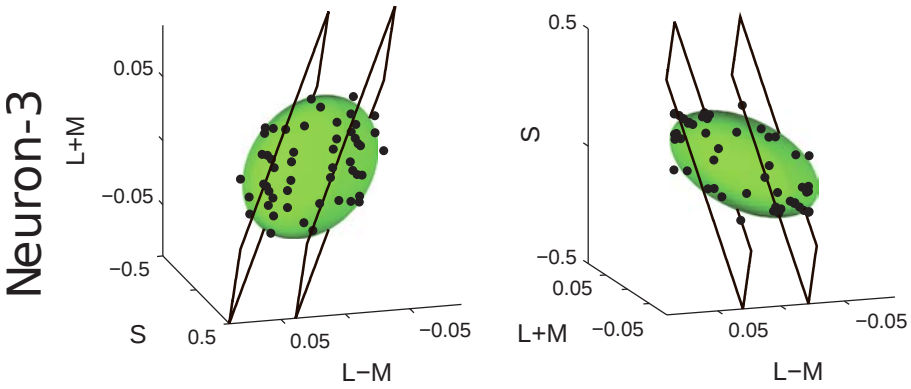


Figure 4.1 – Two projections of Neurons-3 fitted to quadratic surfaces (green ellipsoids) in colour-opponent space, adapted from [119]. Black lines represent the best fitting planes.

Similar ellipsoidal shapes have also emerged in our psychophysical measures of colour boundaries where subjects were asked to produce the intermediate colour between two basic colour terms on a calibrated cathode ray tube (CRT) monitor [186]. This does not appear as a great surprise since colour categories tend to occupy connected regions of colour space [191]. However, these results could in turn explain the organisation of universal colour terms around foci with perceptual constraints governing their position and shape, *i.e.* supporting the hypothesis that colour naming reflects optimal partitions of colour space [191].

4.2.2 Ellipsoidal partitioning of colour space (EPCS)

We modelled each colour category as an ellipsoid in three-dimensional colour-opponent space following these rationale:

1. The presence of neurons like Neuron-3 in V1 [119] shows the plausibility of complex colour-opponent processing at low cortical levels, *i.e.* ellipsoids are parsimonious shapes that can be implemented by low-level visual neurons.

2. Contours of ellipsoids provide an appropriate fit to the psychophysical experiments we conducted in which colour categorical boundaries were measured in a controlled environment [186].

3. In the context of colour categorisation, the centre of an ellipsoid can be interpreted as the focal colour and its geometrical properties determine the optimal partitioning [191].

An ellipsoid aligned to the axes of a Cartesian coordinate system is defined as:

$$\left(\frac{x-x_0}{a}\right)^2+\left(\frac{y-y_0}{b}\right)^2+\left(\frac{z-z_0}{c}\right)^2=1, \tag{4.1}$$

where (x_0, y_0, z_0) are the coordinates of the ellipsoid-centre; and (a, b, c) represent the length of the semi-axes. To account for any rotations around the axes of the coordinate system, we defined our complete set of ellipsoid parameters s with nine parameters:

$$s = [(x_0, y_0, z_0), (a, b, c), (\theta, \phi, \gamma)], \tag{4.2}$$

where (θ, ϕ, γ) are the rotational angles around each of the colour-opponent axes.

A naïve procedure to categorise pixels into different colour terms can be described as a simple binary test: when a pixel is inside an ellipsoid, it belongs to that category, otherwise it does not. However, there are two major flaws with this approach: (i) pixels outside of all ellipsoids will be categorised with neither of the colour terms; and (ii) the colour categorisation will lack the fuzziness proposed by [136] as its underlying visual mechanism. Thus, to simulate the large variability present in the categorisation decision we utilised the sigmoid curve that is a spacial case of the logistic function, given as

$$S(g) = \frac{1}{1+e^{-g}}, \tag{4.3}$$

where g is the steepness of the curve (also knows as the growth ratio). Larger values of g results in a more binarised categorisation, whereas smaller values of g increase

the fuzziness of our model.

There are various ways to model the steepness of each colour category. The simplest is to set g as a constant number. Another strategy is to establish a relationship between the steepness of each category and size of its ellipsoid. We favoured a more flexible solution in which g is set as a free variable for each colour category. This allows our model to vary its level of fuzziness for different colour names. Therefore, in our model each colour term, t, consists of ten parameters:

$$t = [s; g]. \tag{4.4}$$

We defined "belongingness" of a pixel to a colour category as:

$$B_t(x) = \frac{1}{1 + e^{g_t(|p-c_t|-h)}}, \tag{4.5}$$

where B is the likelihood of pixel p belonging to colour term t; g_t represents the steepness; c_t is the centre of its ellipsoid; h is the position of the half-height transition point, which in our model is defined as the distance from the centre of an ellipsoid to its surface in the direction joining c_t and p. It must be noted that in order to obtain a probability distribution B_t must be divided by the sum of all Bs.

Although trivial, it is worth mentioning that when a pixel falls inside an ellipsoid, $|p - c_t|$ is smaller than h, as a result the input of the natural exponential function becomes a negative value. Consequently, the entire natural exponential term becomes smaller than 1. The belongingness of a pixel to a colour category increases as $|p - c_t| - h$ tends towards $-\infty$ and it reaches its maximum value at the centre of an ellipsoid, where the exponential term drops to 0.

Deterministic colour naming requires a unique term for every pixel. This can be achieved through different strategies of combining probabilities of all colour categories, for instance considering the perceptually neighbouring colour (*i.e.* red and orange, or pink and purple). However, this is beyond the scope of this chapter and we adopted a simple maximum pooling mechanism: the highest probability among all colour categories is assigned as the colour term C of that pixel:

$$C(x) = \underset{t}{\operatorname{argmax}}\, B_t(x) \tag{4.6}$$

4.2.3 Acquiring model parameters

Colour space

The first prerequisite for modelling the processes that occur in the visual cortex is to represent the chromatic signal in a colour-opponent space (resembling the

signal arriving from the retina). We selected the CIE L*a*b* colour space because is considered to be (approximately) perceptually uniform [48] and is widely used in computer vision and visual sciences. Nevertheless, since in our model we employ ellipsoids to partition a given colour space into different colour categories, our model is not dependant on the CIE L*a*b* and should work equally well in other colour-opponent spaces, such as CIE L*u*v*, lsY and DKL.

Parameters optimisation

The proposed colour ellipsoids are parsimonious geometrical shapes whose parameters can be determined by different procedures. The simplest option would be to draw those ellipsoids manually and set the steepness to a constant value. Alternatively, the surface of each ellipsoid can be fitted into data points that represent boundaries of a colour term; and the steepness of a category can be defined as the average length of its ellipsoid semi-axes, $g_t = \frac{a_t+b_t+c_t}{3}$, similar to [186]. The most comprehensive solution would probably be to construct a ground truth for every point in a canonical colour-opponent space by means of psychophysical experiments. From this ground truth all the ten parameters of our model can simultaneously be learnt in an optimisation framework.

However, in practice collecting such an exhaustive ground truth from a large set of subjects is extremely time consuming. To overcome this issue we simulated the ground truth from the validation set of the Ebay colour naming dataset presented in [228] (8 images per each of the eleven basic colour names, making a total of 88). Given pixel p, we counted the number of times it was categorised as each of the eleven basic colour names. Dividing this by the total number of times pixel p was categorised resulted in the degree of membership to each colour term.

We learnt the parameters of our model with a sequential quadratic programming optimisation method (10^3 number of iterations and 10^{-3} as tolerance constraint) with the error function

$$\operatorname{argmin} \sum_{x=1}^{N} B_t(x) - G_t(x), \tag{4.7}$$

where N is number of pixels in the ground truth set; B_t is defined in Eq. 4.5; and $G_t(x)$ is the ground truth value of pixel x belonging to category t. We simply initialised each colour ellipsoid, i_t, as follows

$$i_t = [(\mu_{L*t}, \mu_{a*t}, \mu_{b*t}), (10,10,10), (0,0,0); 1], \tag{4.8}$$

where $(\mu_{L*t}, \mu_{a*t}, \mu_{b*t})$ are the average coordinates (in CIE L*a*b* colour space) of all the pixels whose ground truth value of category t is non-zero. We did not set

any constraints on the optimisation of ellipsoid centres. Naturally, we restricted the length of semi-axes to positive values and the rotational angles to the range of $[0, \pi)$. Steepness of sigmoidal function was limited to the range of $(0, 1]$.

Figure 4.2 illustrates the eleven colour ellipsoids learnt from our simulated ground truth. One can highlight a few aspects of the zenithal view that express high congruence with our very own colour perception as follow:

- The achromatic categories, *i.e.* white, grey and black, are placed at the centre of all other ellipsoids in line with the hue circle, which was first proposed by Newton [173].
- The ellipsoids corresponding to opponent colours, *i.e.* red-green and yellow-blue, do not overlap. This is in line with Hering's colour theory which states that these colour cannot be perceived together.

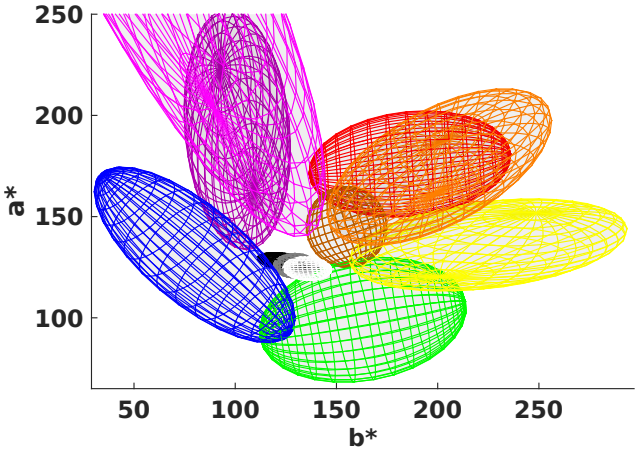


Figure 4.2 – Zenithal view (the $L = 0$ plane) of the learnt colour ellipsoids corresponding to each of the eleven basic colour terms in the CIE L*a*b* colour space.

4.3 Experiments and results

We learnt parameters of our model – termed Ellipsoidal Partitioning of Colour Space (EPCS) – from two different ground truths:

- *EPCS [Rw]* – learnt only from real-world images by extracting the ground truth from validation set of [228].

- *EPCS [Ps]* – to account for colour naming experiments we averaged pixel probabilities of real-world ground truth with the psychophysical results of [186].

We quantitatively evaluated the proposed model by conducting experiments on two different kinds of datasets: (i) colour chips categorised by psychophysical experiments; and (ii) colour segmented objects in real-world images.

4.3.1 Munsell colour chart

The left panel of Figure 4.3 shows the Munsell chart that contain 330 different colour chips (*i.e.* eight chromatics rows, each consisting of 40 hues in increments of 2.5, and one column of 10 achromatic lightness). A large number of colour naming studies have compared their categorisation results to the psychophysical experiments of Berlin & Kay [25] (*i.e.* 24 native speakers from 110 languages were asked to name each Munsell chip) and Sturges & Whitfield [217] (*i.e.* 20 English speakers named each Munsell sample twice). Our segmentation of the Munsell chart is illustrated on the middle and right panels of Figure 4.3. Segmentation obtained by EPCS [Rw] perfectly matches with the psychophysical experiment of Sturges & Whitfield and only vary on five points (all caused by the white colour) comparing to the survey of Berlin & Kay.

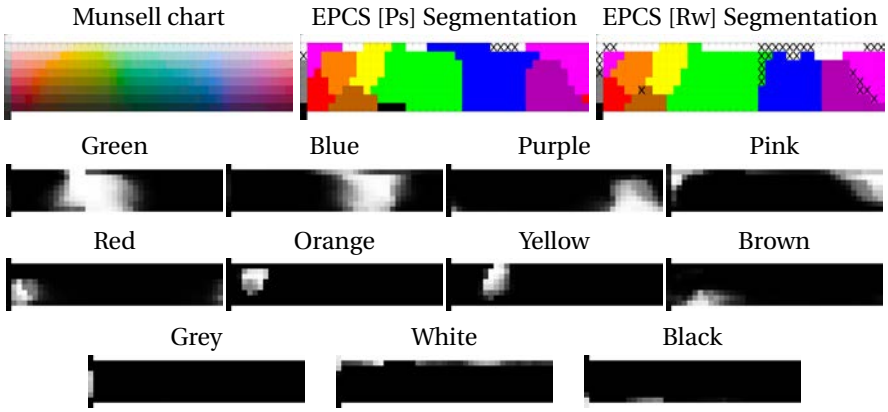


Figure 4.3 – Result of EPCS applied to the Munsell colour chart. Black crosses indicate the mismatches to either Berlin & Kay [25] or Sturges & Whitfield [217] data. Black and white images correspond to probability of each pixel to different colour categories, illustrated in the form of heat map, brighter pixels represent higher probabilities.

Table 4.1 quantitatively compares the accuracy of our model to seven state-of-the-art algorithms that have also reported theirs results on the Munsell chart. In comparison to the colour naming survey of Berlin & Kay [25] EPCS practically matches the best results reported in the literature (NICE), far ahead of the third best colour naming models (SFKM, TSEM). With respect to the psychophysical experiment of Sturges & Whitfield [217] our model along with SFKM, TSEM and NICE obtains the perfect accuracy.

	Berlin & Kay	Sturges & Whitfield
LGM [64]	0.77	0.83
MES [159]	0.87	0.96
TSM [23]	0.88	0.97
SFKM [199]	0.92	1.00
TSEM [24]	0.92	1.00
PLSA [228]	0.89	0.98
NICE [186]	0.98	1.00
EPCS [Ps]	**0.98**	**1.00**
EPCS [Rw]	**0.87**	**0.98**

Table 4.1 – The true positive ratio of several colour naming models on psychophysical experiments of Berlin & Kay [25] and Sturges & Whitfield [217]. Lammens's Gaussian model (LGM) [64], MacLaury's English speaker model (MES) [159], Benavente & Vanrell's triple sigmoid model (TSM) [23], Seaborn's fuzzy k-means model (SFKM) [199], Benavente *et al.* 's triple sigmoid elliptic model (TSEM) [24], van de Weijer *et al.* 's probabilistic latent semantic analysis (PLSA) [228], Parraga & Akbarinia's neural isoresponsive colour ellipsoids (NICE), and the proposed ellipsoidal partitioning of colour space (EPCS).

Referring to Table 4.1, we can observe a large difference between two variations of our model mainly caused by white pixels. EPCS [Rw] (learnt only from real-world images) categorises pixels with a faint colour as white, whereas EPCS [Ps] (learnt by influence of colour naming experiments in controlled environment) categorise those pixels into chromatic categories. This is an issue noted by [228] as well.

4.3.2 Real-world images

We evaluated the proposed model on two datasets of real-world images[1]. Along with our model we tested three state-of-the-art methods (whose source codes are

[1]The source code and all the experimental materials are available at https://github.com/ArashAkbarinia/ColourCategorisation.

publicly available) : Benavente *et al.* 's triple sigmoid elliptic model (TSEM) [24], van de Weijer *et al.* 's probabilistic latent semantic analysis (PLSA) [228], and Parraga & Akbarinia's neural isoresponsive colour ellipsoids (NICE) [186]. We assessed each algorithm based on their true positive ratio, *i.e.* $\frac{TP}{TP+FN}$, where TP represents pixels whose colour names are correctly labelled and FN are those that are mislabelled. Due to the nature of the available ground truths, which primarily contain one colour category per image, other evaluation metrics were inappropriate. Images of tested datasets are of various size and in order to avoid the bias for smaller images, we first computed the true positive ratio for each image and reported results are averaged over all.

Ebay dataset

The Ebay dataset [228] consists of four sets of man-made objects, *i.e.* cars, dresses, pottery and shoes. Every set contains 110 images, *i.e.* ten images for each of the eleven basic colour terms. The ground truth masks are based on semi-automatic segmentation algorithms. In order to compensate for absence of natural objects (such as, fruits, vegetables, flowers, *etc.*, that colour information arguably plays an important role in their recognition) we extended this dataset by creating an extra set of images containing natural objects following the same procedure as the original authors.

We have reported true positive ratio of four methods on the Ebay dataset in Table 4.2. Evidently EPCS [Rw] outperforms all other methods with a large margin. We can also observe a large gap between performance of EPCS [Ps] and PLSA in comparison to TSEM and NICE in all five subcategories. In three sets (dresses, shoes and natural) EPCS [Ps] obtains higher true positive ratio compared to PLSA. Advantage of EPCS [Ps] over PLSA becomes more tangible by considering their respective performance on psychophysical data, where EPCS [Ps] performs notably better (see Table 4.1).

	Cars	Dresses	Pottery	Shoes	Natural
TSEM [24]	0.59	0.68	0.62	0.73	0.69
PLSA [228]	0.60	0.82	0.76	0.78	0.77
NICE [186]	0.52	0.69	0.54	0.67	0.67
EPCS [Ps]	**0.60**	**0.84**	**0.76**	**0.79**	**0.80**
EPCS [Rw]	**0.65**	**0.86**	**0.80**	**0.80**	**0.80**

Table 4.2 – True positive ratio of four colour naming models on Ebay dataset for each subcategory.

Small objects dataset

The small objects dataset [248] contains 300 images (in 16-bit format) of various materials (*e.g.* paper, plastic, metal, wood, fruits, *etc.*) that are captured under different types of illuminants. Each image is supplemented with a manual segmentation of its constituting regions according to their colour names. However, it is important to note that number of pixels for each of the eleven basic colour terms is not uniformly distributed.

We have reported true positive ratio of four methods on the small objects dataset in Table 4.3. We can observe comparable patterns similar to those in the Ebay dataset. EPCS [Rw] performs best among all others with a 4% margin. Correspondingly, both EPCS [Ps] and PLSA obtain better results in comparison to TSEM and NICE.

		Small Objects
TSEM	[24]	0.69
PLSA	[228]	0.73
NICE	[186]	0.52
EPCS	[Ps]	**0.73**
EPCS	[Rw]	**0.77**

Table 4.3 – True positive ratio of four colour naming models on the small objects dataset.

4.4 Discussion

In Figure 4.4 we have illustrated one exemplary image from the small object dataset. We can observe that EPCS [Ps] mislabels the white part of the wall as pink. This is the main reason that EPCS [Rw] clearly performs better than EPCS [Ps] in real-world images (refer to Tables 4.2 and 4.3). However, we would like to emphasise the results of the later one that in psychophysical data, where the environment is controlled and no noise is present, obtains almost perfect true positive ratio (similar to NICE), at the same time it does reasonably well on real-world images (contrary to NICE). We believe high accuracy on psychophysical experiments is essential because a colour naming model should first and foremost correctly categorise individual pixels. Other challenging tasks for colour naming models have to do with faint colours appearing as white in an image context. This is caused by different phenomena such as colour constancy and induction. These challenges should be solved by modelling colour naming in a dynamic fashion. In other words a model

that can adapt itself to the image or pixel content.

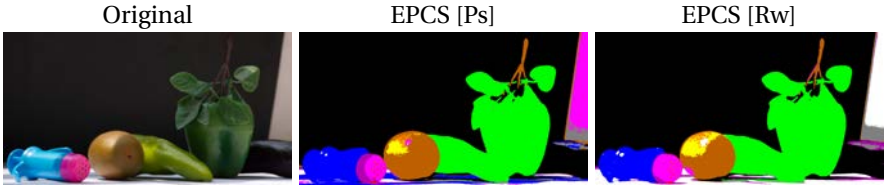


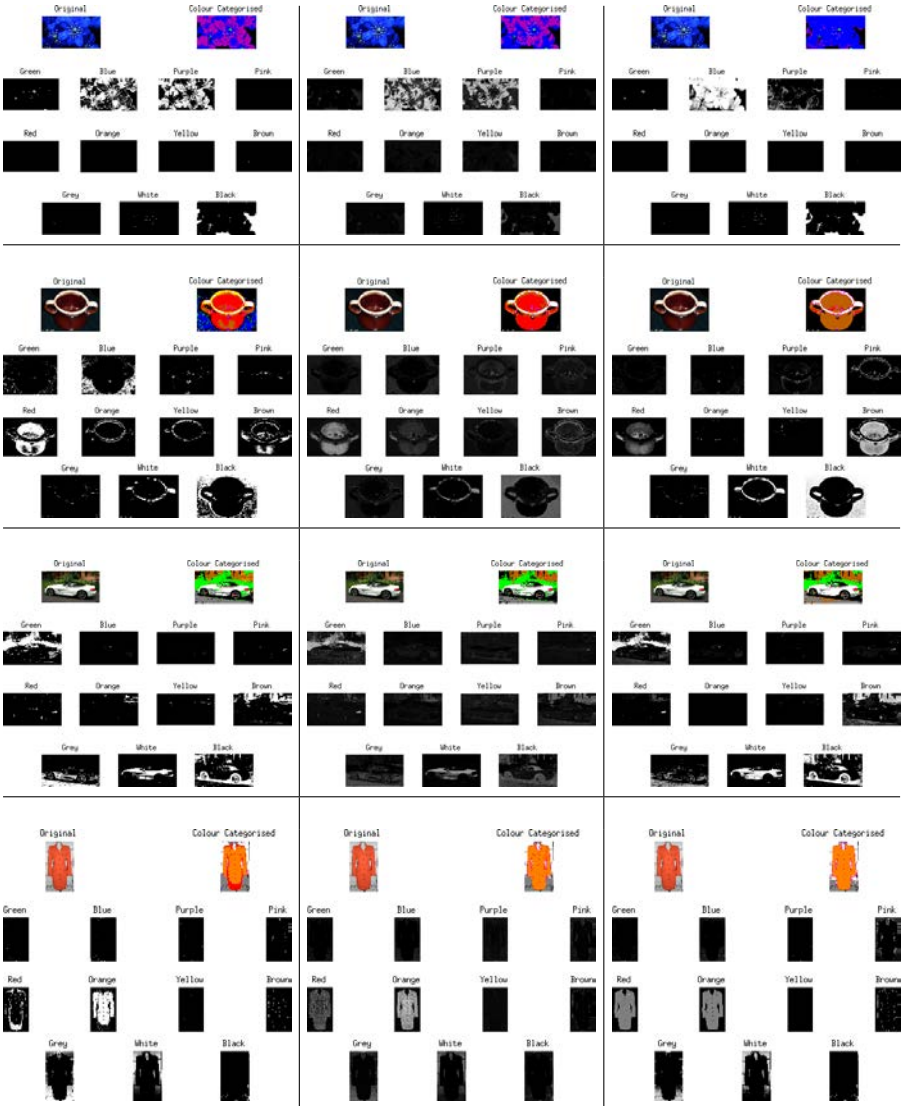
Figure 4.4 – EPCS [Ps] versus [Rw] in a real-world image.

Figure 4.5 shows four examples from the real-world datasets. In each panel the original image is displayed accompanied with its respective colour categorised results from each of the algorithms considered: TSEM [24], PLSA [228], and EPCS. We can observe in the first row that the blue flowers are largely misclassified as purple by TSEM and PLSA. However, they are correctly assigned to the blue category in our model. We detected a number of similar cases with other blue objects.

The brown pottery mug – present in the second row of Figure 4.5 – is almost entirely miscategorised as red by TSEM and PLSA. On the contrary, EPCS accurately labels it as brown. A closer inspection to the corresponding probability maps reveals that TSEM assigns pixels of the mug to the red category with a very high probability (almost 100%). PLSA labels them as red (with 60% probability) while granting some likelihood to the perceptually neighbouring colours (about 20% to orange and 10% to brown). However, this uncertainty spreads to the purple category as well with about 5% probability. EPCS's results show more consistency with about 60% probability on the brownness of the mug, while acknowledging that the neighbouring colour red is also probable (with about 40%). It is also worth paying extra attention to the background of this picture, where TSEM misassigns a great portion of it to the blue category. Contrary to this, PLSA and EPCS have no difficulties to correctly label it as black.

The white car – depicted in the third row of Figure 4.5 – is a difficult case due to the cast of green light over its body and surroundings. We can observe that all three methods, in general, accurately label the car as white. Nonetheless, there are some pixels near the back wheel and on the front door that are mistakenly categorised as green. This issue is more noticeable for TSEM and its minimal in EPCS.

The orange dress – displayed in the forth row of Figure 4.5 – is also correctly labelled by all three algorithms. However, the colour of dress is slightly ambiguous and certain individuals might label it as red. A review of the probability maps shows that TSEM considers the dress as orange with almost 100% probability, therefore not satisfactorily capturing the observers' opinion. In contrast to this, PLSA and



(a) TSEM [24] (b) PLSA [228] (c) **EPCS** [Ps]

Figure 4.5 – Detailed comparison of three different algorithms on real-world images. Individual panels consists of the original image, its colour categorised version, and elven probability maps corresponding to each basic colour term.

EPCS assign equally high probability to the dress being red and being orange.

4.4.1 Model extension

There are a number of situations where one might want to add extra categories to the eleven basic colour terms (*e.g.* , some languages contain two names for "blue" like Russian, Italian and Spanish from the River Plate area). Alternatively, there are many intermediate colour terms used in everyday language (such as, olive, turquoise, cream, *etc.*) that arguably deserve their own category. Furthermore, certain applications need more elaborated colour names (*e.g.* , those that are used by artists and painters).

New colour names are usually learnt by humans (both adults and children) after the presentation of a small handful of examples. Within our model we can simulate this process straightforwardly. As an illustration, we learnt the colour term "cream" from merely two images that are depicted in Figure 4.6. We followed the same procedure explained in section 4.2.3 by manually labelling the cream part of training images.

Figure 4.6 – Two images used in learning colour cream.

Figure 4.7 shows the impact of this newly introduced cream category on the colour segmentation of a sample image from the Pascal Project Dataset [67]. We can observe that by relying only on the eleven colour terms, EPCS incorrectly labels the wall on the back of the image as pink (although with low probability that is on average smaller than 20%). Colour segmentation with twelve categories allows our model to accurately classify the wall as cream. The flexibility of our algorithm can be further exploited to create a personalised colour naming model which reflects the individual variability present in the psychophysical data. This is very economical and can be achieved by segmenting a handful of images from a personal digital assistant (PDA), for example. Furthermore, an interactive application can allow subjects to manipulate the colour ellipsoids directly to achieve the colour categorisation they desire.

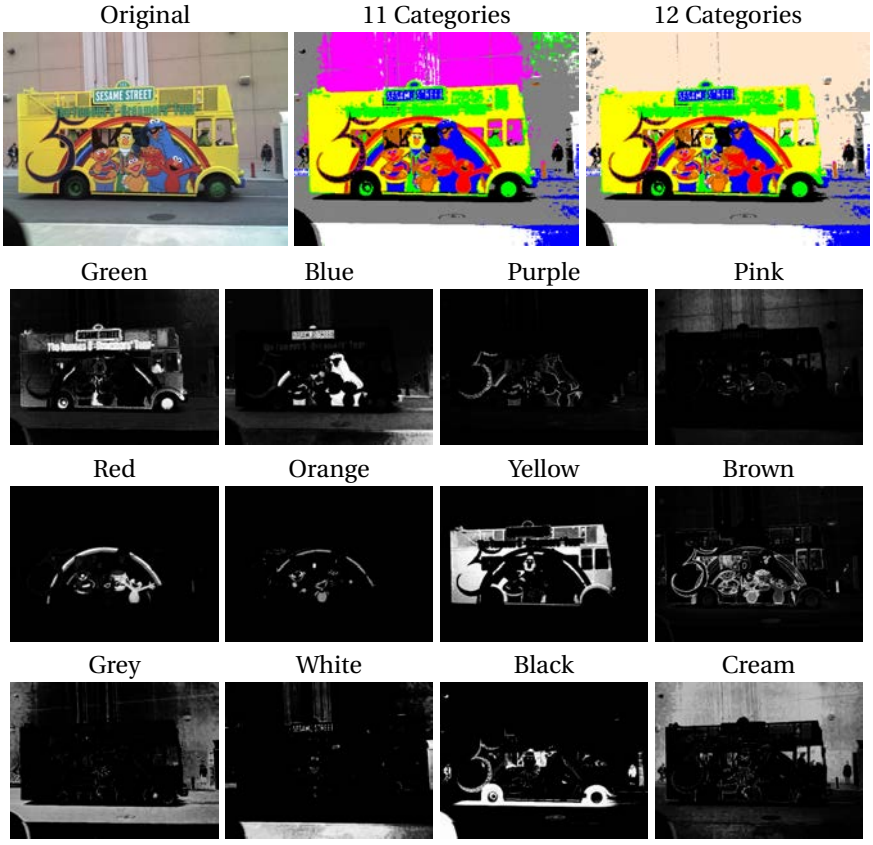


Figure 4.7 – Colour categorisation including an extra category for the cream colour. The top row shows the original image with is respective segmentation considering eleven or twelve colour categories.

4.4.2 Model adaptation

One important aspect of any colour naming model is its context adaptability. This is feasible within our model by dynamically adjusting the ellipsoids to the image or even the pixel being processed. One of the greatest challenges in colour naming algorithms is the frontier between chromatic and achromatic colours as we experienced in our experiments and mentioned by [228]. In a neutral background colours appear more saturated comparing to a colourful environment [38]. As a proof-of-

concept we attempted to address this issue by adapting achromatic ellipsoids to the level of colourfulness of an image. We stretched the chromatic semi-axes (a*b*) of achromatic ellipsoids on the direction that average pixels of an image differed from neutral grey. The results of this experiment are reported in Table 4.4.

Cars	Dresses	Pottery	Shoes	Natural	Small objects
0.60	0.84	0.76	0.80	0.81	0.74

Table 4.4 – The true positive ratio of our adaptive ellipsoids for EPCS [Ps] on the real-world datasets.

Our naïve adaptation increases the true positive ratio by 1% on three sets of real-world images (shoes and natural categories of Ebay dataset and small objects). This by no means is a finished adaptable model, rather a demonstration that our model is able capture a greater variation in image content with the addition of simple extensions. This can be further explored by adapting chromatic ellipsoids to the presence or absence of certain colour categories in the image, following reports that link them to the phenomenon of colour constancy [230]. For instance when the green signal is abundant one could shrink the green ellipsoid or translate its centre. The adaptability of the ellipsoids in our model in turn could offer a framework in which colour constancy and colour categorisation are addressed simultaneously.

4.5 Conclusion

In this chapter, we presented a biologically-inspired colour categorisation model where each colour term is represented by an ellipsoid in colour-opponent space. To capture the fuzzy nature of colour names and account for the non-linear operations performed by visual cortex neurons, we computed the final degree of membership to a category using a sigmoid curve. Theoretically, we justified our geometrical framework by linking it to physiological and psychophysical evidence. In practice, we showed that the parameters of our parsimonious model can be learnt from a simple optimisation procedure and conducted two kinds of experiments to verify its sanity. Results obtained on the Munsell chart are in excellent agreement with the psychophysical results of colour naming. We also perform better than other popular algorithms in real-world images. The advantage of the proposed model is more tangible by realising that, unlike all other state-of-the-art algorithms, it performs well on both types datasets. This shows that our model can both explain psychophysically-based colour naming results and perform an accurate categorisation of real-world images.

Biologically-inspired chromatic models have been successful in a wide range of colour computational tasks as shown in this dissertation and also in the literature, *e.g.* colour induction [180], colour constancy [7, 10, 185], saliency [172], colour descriptor [250] and boundary detection [4]. This is not surprising since colour is a sensation that originates from within our brains, which in turn is the product of millions of years of evolution, adaptation and "learning" from the visual environment. From this point of view, we believe our approach to colour categorisation can compete with other deterministic and learning-based approaches. In this line, we demonstrated that our model can be easily extended to incorporate more colour terms from few examples (as human infants do) and adapt itself to the content of image. Implicitly demonstrating the potential of biologically-inspired colour categorisation modelling for different applications such as image segmentation and image retrieval. Naturally, our model (as any other colour naming model) is likely to improve its accuracy in different environments when complemented with good colour constancy and colour induction algorithms and fundamentally with larger and better ground truths.

There are at this point a number of possible lines of investigation for the future. To mention a few of those: (i) Improving the assignment of colour names by considering more sophisticated rules than a simple max-pooling. For example, by the contrast-variant-pooling mechanism introduced by us in this dissertation can be considered. (ii) Making the model dynamically responsive to context (either by rearranging the ellipsoids according to image content, or alternatively, supplementing the model with a centre-surround adaptation mechanism similar to those we implemented for the colour constancy phenomenon). In this way we can account for the well known colour phenomena of induction and constancy. (iii) Converting the colour ellipsoids to three-dimensional Gaussian envelopes which are biologically more plausible and also mathematically more tractable. This might allow an easier adaptation of our model at pixel level.

Object Edges Part III

How do we separate objects from each other?

5 Boundary Detection

In the previous chapters, we discussed the perception of colours under different illuminants and how our brain categorises them into meaningful colour terms. We know that colour and form are linked inextricably by sharing the same cortical areas for their respective processing, and there is abundant evidence suggesting that contours and shapes contribute significantly to the appearance of colour [203]. Therefore, it is reasonable to speculate that a surround modulation mechanism similar to the one we presented in section 3 can be involved in the process of shape perception. Consequently in this chapter, we explore the role that surround modulation plays in detecting the boundaries of objects.

5.1 Introduction

Our ability to recognise objects is completely entangled with our ability to perceive contours [181, 234]. It has been shown that the primary and secondary visual cortices – *i.e.* V1 and V2 – play a crucial role in the process of detecting lines, edges, contours, and boundaries [152], to such extent that an injury to these areas can impair a person's ability to recognise objects [249]. Furthermore, edges (a form of image gradient sometimes also referred to as "boundaries" or "contours") are indispensable components of computer vision algorithms in a wide range of applications (such as, colour constancy [226], image segmentation [15], document recognition [147], human detection [54], *etc.*).

Given their importance, many computational models have been proposed to detect edges – for a comprehensive review refer to [181]. In its earliest form Prewitt [189] proposed a convolutional-based image gradient to capture local changes. Marr [163] suggested a correspondence between edges and zero-crossing points. Canny [41] improved on previous algorithms by incorporating non-maximum suppression and hysteresis thresholding. The greatest challenge faced by these classical methods is the distinction between authentic boundaries and undesired background textures. This issue was partially addressed by local smoothing techniques, such as bilateral filtering [221] and mean shift [50]. Thereafter, graph-based models emerged, *e.g.* [52, 70], allowing for closure to be taken into account. More recent frameworks extract relevant cues (*e.g.* colour, brightness, texture, *etc.*) feed-

ing them to machine learning algorithms, such as probabilistic boosting tree [60], gradient descent [15] and structured forest [61]. Currently, state-of-the-art algorithms [27, 28, 138, 204, 242] rely heavily on deep-learning techniques.

Despite their success, learning methods suffer from three major drawbacks: (a) their performance might be dataset dependant; (b) they are computationally demanding since for every single pixel a decision must be made (in both training and testing stages) on whether it corresponds to an edge or not; and (c) they require extremely large amounts of data for an effective training procedure. In addition to these, there is no biological or behavioural evidence that edge detection is the result of such a laboriously supervised learning process. On the contrary, biological systems compute edges in an unsupervised manner, starting from low-level features that are modulated by feedback from higher-level visual areas, *e.g.* those responsible for global shape [152].

In line with this, a number of biologically-inspired edge detection models have been recently proposed with promising results. Spratling [216] proposed a predictive coding and biased competition model based on the sparsity coding of neurons in V1. Wei *et al.* [235] presented a butterfly-shaped inhibition model based on non-classical receptive fields operating at multiple spatial scales. Further improvement came from Yang *et al.* [247] who explored imbalanced colour opponency to detect luminance boundaries. The same authors demonstrated employing the spatial sparseness constraint, typical to V1 neurons, helps to reserve desired fine boundaries while suppressing unwanted textures [244]. Another improvement in contour detection originated from introducing multiple features to the classical centre-surround inhibition common to most cortical neurons [246]. The introduction of feedback connections has also been beneficial. Díaz-Pernas *et al.* [56] extracted edges through oriented Gabor filters accompanied with top-down and region enhancement feedback layers.

In this chapter, we propose a biologically-inspired edge detection model that incorporates recent knowledge of the physiological properties of cortical neurons. Our work is novel compared to the methods mentioned above in four main aspects: (i) we incorporate a more sophisticated set of cortical interactions which includes four types of surround, *i.e.* full, far, iso- and orthogonal-orientation; (ii) we account for contrast variation of surround modulation; (iii) we model V2 neurons that pool signals from V1 responses over a larger region corresponding to the centre and neighbouring spatial locations; and (iv) we consider a fast-conducting feedback connection from higher visual areas to the lower ones.

Figure 5.1 illustrates the flowchart of our framework, which follows the functional structure of the human ventral pathway. Our processing starts in the retina, where the input image is convolved by single opponent cells and sent though the lateral geniculate nucleus (LGN) in the form of colour opponent channels [203].

These channels are processed by double-opponent cells in V1 – known to be responsive to colour edges [203] – whose receptive field (RF) are modelled through the first derivative of a Gaussian function [41]. To consider the RF surround: we define a short range circular (isotropic) region corresponding to full surround [152], long range iso- and orthogonal-orientation surrounds along the primary and secondary axes of the RF [71], and we model far surround via feedback connections to enhance the saliency of edge features. All these interactions are inversely dependant on the contrast of the RF centre [208]. The output signal from V1 is pooled at V2 by a contrast-variant centre-surround mechanism applied orthogonally to the preferred direction of the V1 RF [188]. Finally, to account for the impact of global shapes on local contours [152], we feed the output of V2 layer back into V1.

5.2 Surround Modulation Edge Detection

5.2.1 Retina and lateral geniculate nucleus (LGN)

The retina is the starting point of visual processing in humans. Cone photoreceptor cells located at the back of the retina absorb photons at every spatial location. Their output is processed in an antagonistic manner by further layers of single-opponent cells (ganglion cells) and sent to the cortex through the LGN in the form of a luminance and two chromatically-opponent channels [203], usually modelled as

$$\begin{aligned} SO_{lu}(x,y) &= S_r(x,y) + S_g(x,y) + S_b(x,y), \\ SO_{rg}(x,y) &= \kappa_r S_r(x,y) - \kappa_g S_g(x,y), \\ SO_{yb}(x,y) &= \kappa_b S_b(x,y) - \kappa_{rg}\left(\frac{S_r(x,y) + S_g(x,y)}{2}\right), \end{aligned} \tag{5.1}$$

where SO represents the response of single-opponent cells, $\{lu, rg, yb\}$ denotes the luminance, red-green and yellow-blue opponent-channels, (x,y) are the spatial coordinates, and $\{r,g,b\}$ are the original red, green and blue cone signals. S is the spectral response function of each cone and can be approximated by a two dimensional Gaussian function as follows

$$S_h(x,y) = I_h(x,y) * G(x,y,\sigma), \tag{5.2}$$

where I is the input image, $h \in \{r,g,b\}$ is the index of colour channels, $*$ denotes the convolution operator and G is the circular two-dimensional Gaussian kernel,

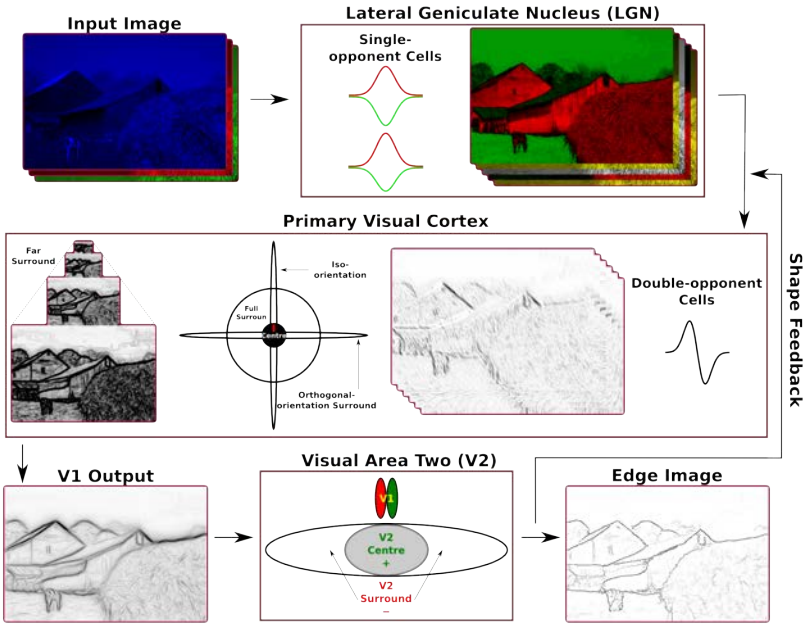


Figure 5.1 – The flowchart of our model. Balanced and imbalanced colour opponent channels are created in the retina and sent through the LGN. Orientation information is obtained in V1 by convolving the signal with a derivative of Gaussian at twelve different angles. We model four types of orientation-specific surround: full, far, iso- and orthogonal-orientation. In V2 the signal is further modified by input from surrounding areas in a directional orthogonal to that of the original RF. Shape feedback is sent to V1 as an extra channel.

defined as

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)}, \tag{5.3}$$

with variance σ. This Gaussian convolution is equivalent of a smoothing preprocessing stage in computer vision which has been demonstrated to play an important role in the successive edge detection [181].

When the chromatically-opponent input to single-opponent cells in Eq. 5.1 is in equilibrium, parameter κ is equal to one for all channels. However, there is

physiological evidence showing that some types of single-opponent cells combine chromatic information in an imbalanced fashion [203]. The significance of these cells has also been shown in practice through many computer vision algorithms, *e.g.* edge detection [244, 247] and colour constancy [93, 185]. Following this insight, we included two imbalanced chromatic opponent-channels: $SO_{rg'}$ with $\kappa_g = 0.7$ and $SO_{yb'}$ with $\kappa_{rg} = 0.7$.

5.2.2 Primary visual cortex (V1)

Once the visual signal is pre-processed in the retina and the LGN, it is sent for further processing into the visual cortex. Early neurophysiological evidence established that the feedforward arrays coming from the LGN interact dynamically in the visual cortex, creating various gain control pools across all spatial orientations which can be modelled as "divisive normalisation" [113]. In this configuration, each cortical neuron computes a rectified combination of its inputs, followed by a normalisation where the neuron's response is divided by the pooled activity of its neighbours. The overall effect of this gain normalisation is to both alter the contrast response of neural units, making them more responsive to boundaries, and to narrow their orientation bandwidths. Another mechanism contributing to orientation tuning stems from the long-range connections between neurons with similar orientation tuning ("collinear facilitation") [161]. Both mechanisms are thought to enhance contour continuity, altering the effective orientation tuning of cells.

Although divisive normalisation and collinear facilitation are powerful mechanisms, recent studies have shown that they are likely to be oversimplifications, since stimuli outside of the classical receptive field of a cortical neuron can also modulate that neuron's activity in various ways. The origin of this modulation is feedforward, feedback and lateral, stemming from previous connections, later connections and from neighbouring neurons in the visual pathway. However, it was not until the mid-1980s that the concept of non-classical (surround-modulated) receptive field became established and characterised using circular or annular gratings of varying characteristics.

Now we understand that *SO* channels arriving at the cortex are processed by a number of double-opponent cells in V1 that are responsive to boundaries [203], but also modulated by regions beyond their RF centres, with facilitation predominantly at low contrast and inhibition at high contrast [14, 135, 208].

As a consequence of the above, we defined the receptive field of our orientation-tuned double-opponent cells *DO* as

$$DO_c(x,y,\theta) = CR_c(x,y,\theta) + \zeta_c^{-1}(x,y)SR_c(x,y,\theta), \tag{5.4}$$

where c is the index of SO channels, θ is the preferred orientation of the RF (set to twelve evenly distributed angles in the range $[0, 2\pi)$ [188]), CR and SR are the centre and surround responses respectively, and ζ is the contrast of the RF centre approximated by the local standard deviation of its constituent pixels. Double-opponent cells are typically modelled in biologically-inspired algorithms by Gabor filters, [56, 216, 246], or the first derivative of a Gaussian function, [244, 247]. We settled for the later one originally proposed by Canny [41], therefore, we defined the DO centre response, CR, as

$$CR(x, y, \theta) = SO * \left| \frac{\vartheta G(x, y, \sigma)}{\vartheta \theta} \right|, \tag{5.5}$$

where σ is the RF size (set to 1.5 in our model corresponding to the typical RF size of foveally-connected neurons in V1 or 0.25° of visual angle [14], which is equivalent to approximately 13 pixels when viewed from $100cm$ in a standard monitor).

Surround modulation

We defined the surround response, SR, as follows

$$SR(x, y, \theta) = LS(x, y, \theta) + IS(x, y, \theta) + OS(x, y, \theta) + FS(x, y, \theta), \tag{5.6}$$

where LS is full surround referring to the isotropic region around the RF; IS denotes iso-orientation surround along the RF preferred axis; OS is orthogonal-orientation surround in the direction perpendicular to the RF preferred axis; and FS denotes far surround.

Because the full surround is an isotropic region (*i.e.* stimulus occupying the entire surrounding region rather than isolated flanking lines [152]) it can be modelled as the average response of a circular window around the cell's RF centre. This surround is inhibitory when it shares the same orientation as the centre and strongly facilitatory when its orientation is perpendicular to the centre [152]. Thus, we defined the full surround LS as

$$LS(x, y, \theta) = \lambda \zeta^{-1}(x, y)\Big(CR(x, y, \theta_\perp) * \mu\Big) - \zeta(x, y)\Big(CR(x, y, \theta) * \mu\Big), \tag{5.7}$$

where $\theta_\perp = \theta + \frac{\pi}{2}$, μ is the circular average kernel and λ determines the strength of orthogonal facilitation in comparison to the iso inhibition. The former facilitation is reported to be stronger than the later inhibition [152], therefore λ must be larger than one.

The iso-orientation surround, IS, extends to a distance two to four times larger than the RF size [71]. Within this region elements parallel to the RF preferred

orientation are facilitatory while orthogonal ones are inhibitory [71, 152], therefore, we modelled IS as

$$IS(x,y,\theta) = \zeta^{-1}(x,y)\Big(CR(x,y,\theta) * E(\sigma_x,\theta)\Big) - \zeta(x,y)\Big(CR(x,y,\theta_\perp) * E(\sigma_x,\theta)\Big), \tag{5.8}$$

where E is an elliptical Gaussian function elongated in the direction θ, defined as follows:

$$E(x,y,\sigma_x,\sigma_y,\theta) = e^{-(ax^2-2bxy+cy^2)}, \quad \text{with}$$
$$a = \frac{\cos^2\theta}{2\sigma_x^2} + \frac{\sin^2\theta}{2\sigma_y^2}, \quad b = -\frac{\sin 2\theta}{4\sigma_x^2} + \frac{\sin 2\theta}{4\sigma_y^2}, \quad c = \frac{\sin\theta^2}{2\sigma_x^2}\frac{\cos\theta^2}{2\sigma_y^2}.$$

We set $\sigma_y = 0.1\sigma_x$ and $\sigma_x = 3\sigma$ corresponding to physiological measurements [71].

The orthogonal-orientation surround, OS, projects to a distance half of the iso-orientation surround [71]. In the orthogonal-surround elements parallel to the RF preferred orientation are inhibitory while perpendicular ones are facilitatory [71, 152], thus, we modelled OS as

$$\begin{aligned} OS(x,y,\theta) = \zeta^{-1}(x,y)\Big(CR(x,y,\theta_\perp) * E(\sigma_x,\theta_\perp)\Big) - \\ \zeta(x,y)\Big(CR(x,y,\theta) * E(\sigma_x,\theta_\perp)\Big). \end{aligned} \tag{5.9}$$

The far surround could extend to regions up to 12.5° of visual angle [208] which is approximately equivalent to 673 pixels when viewed from $100cm$ in a standard monitor. Consequently the feedforward and horizontal connections in V1 that mediate interactions between the RF and its near surround are too slow to account for the fast onset of far surround. Due to this, it has been suggested that far surround is operated through a different mechanism via inter-areal feedback connections [14, 209]. We speculate that parts of these inter-areal connections come from spatial scale layers in V1 [115], and assume their influence to be facilitatory when image elements in this region share the same orientation as the centre [126]. Therefore, we defined FS as

$$FS(x,y,\theta) = \zeta^{-1}(x,y)\sum_{s=2}^{4}\frac{CR_s(x,y,\theta)}{s} \tag{5.10}$$

where s is the index of the corresponding spatial frequency scale. This processing is analogous to the multi-scale processing common to both visual sciences and computer vision, with the distinction that we account for both contrast and distance,

since surround modulation has been reported to be stronger in the near than in the far regions [14].

5.2.3 Visual area two (V2)

Visual processing becomes more global along the brain's ventral pathway, where neurons in each consecutive area seem to pool information from increasingly larger spatial regions (i.e. exponentially larger receptive fields). This allow them to process increasingly complex image features, such as curved arcs, angles, and line intersections and eventually shapes and objects. The next interconnected adjacent area to V1 is V2, where many neurons have been reported to respond to curvatures and extended arcs [239]. It has been proposed that RFs in area V2 extract curvature information by pooling signals from V1 using a centre-surround mechanism in the direction orthogonal to the V1 orientations [188, 239]. In order to model this, first, we defined the V1 response, $V1R$, as the most activated DO orientation. This operation is assumed to be realised by complex cells pooling the maximum value of DO cells [219], modelled as

$$V1R_c(x,y) = \underset{\theta\in[0,2\pi)}{\operatorname{argmax}}\big(DO_c(x,y,\theta)\big). \tag{5.11}$$

The V2 RFs show similar contrast-variant surround modulation as those of V1 [208]. Therefore, we modelled the V2 response, $V2R$, through a Difference-of-Gaussians (DoG) as

$$V2R_c(x,y) = V1R_{c,\theta}(x,y) * E(\sigma_x,\theta_\perp) - \upsilon_c(x,y)V1R_{c,\theta}(x,y) * E(5\sigma_x,\theta_\perp) \tag{5.12}$$

where υ is the contrast of $V1R$ computed by its local standard deviation, the index θ at $V1R_\theta$ shows the preferred orientation of that RF. Cortical RFs increase their diameters systematically by approximately a factor of three from lower to higher areas [239]. Therefore, we set the size of V2 RF, σ_x, to three times the size of a V1 RF. In Eq. 5.12 surround is five times larger than the centre according to physiological findings [208].

5.2.4 Feedback connections

In the primate visual system there are generally massive feedback connections from higher visual areas into lower ones [14]. For instance, the majority of the LGN inputs are feedback connections from other areas of the brain, in particular the visual cortex. The functional role of this cortical feedback in visual processing is

still poorly understood, although new evidence shows that these projections are organised into parallel streams and their effects include tune-sharpening, gain-modulation and various adjustments to behavioural demands [37].

In our model we accounted for only a fraction of the feedback from V2 to V1 corresponding to the well established fact that global shape influences local contours [152]. We simulated this global shape by averaging the V2 outputs of all channels and sending it as feedback to V1. This feedback is processed only one time same as all other inputs to V1. The final edge map is computed as a sum of all V2 output channels:

$$edge(x,y) = \sum_{c} V2R_c(x,y), \quad \text{with} \quad c \in \{lu, rg, yb, rg', yb', feedback\}. \tag{5.13}$$

5.3 Experiments and results

We tested our model – termed *Surround-modulation Edge Detection (SED)* – on three datasets[1]: (i) the Berkeley Segmentation Dataset and Benchmark (BSDS) [15, 165], (ii) the Multi-cue Boundary Detection Dataset (MBDD) [167], and (iii) the Contour Image Database (CID) [107]. Each image of all three datasets is supplemented with a ground truth that is created from manually-drawn edges by number of human subjects. We evaluated our algorithm in the standard precision-recall curve based on its harmonic mean (referred to as F-measure) on three criteria: optimal scale for the entire dataset (ODS) or per image (OIS) and average precision (AP). Naturally, ODS is the most representative of all to measure the performance since it uses a fixed threshold for all images in the dataset [15]. The results we report in this chapter were obtained with a fixed set of parameters (see details in Section 5.2) for all datasets much in the same way as the human visual system.

5.3.1 Berkeley Segmentation Dataset and Benchmark (BSDS)

The BSDS [15, 165] contains two sets of colour images BSDS300 (100 test images and 200 training images) and BSDS500 (200 test images). This dataset contains a wide range of natural and man-made objects. Size of each image is 481×321 pixels. Arguably BSDS is considered as the benchmark dataset for boundary detection in the field of computer vision.

Table 5.1 compares the results of our model to several other state-of-the-art edge detection algorithms that have also reported theirs results on the BSDS dataset.

[1] The source code and all the experimental materials are available at https://github.com/ArashAkbarinia/BoundaryDetection.

From this table we can observe that in BSDS500 our model improves the ODS of methods driven by low-level and biological features by 4%. This improvement is 3% in BSDS300. It must be noted that deep-learning methods often use BSDS300 as the training set for their learning procedure and therefore they do not report their results on this fragment of BSDS.

				BSDS300			BSDS500		
		Method		ODS	OIS	AP	ODS	OIS	AP
		Human		0.79	0.79	–	0.80	0.80	–
Biological	Low-level features	Canny	[41]	0.58	0.62	0.58	0.60	0.63	0.58
		Mean Shift	[50]	0.63	0.66	0.54	0.64	0.68	0.56
		Felz-Hutt	[70]	0.58	0.62	0.53	0.61	0.64	0.56
		Normalised Cuts	[52]	0.62	0.66	0.43	0.64	0.68	0.45
		PC/BC	[216]	0.61	–	–	–	–	–
		CO	[247]	0.64	0.66	0.65	0.64	0.68	0.64
		MCI	[246]	0.62	–	–	0.64	–	–
		dPREEN	[56]	0.65	–	–	–	–	–
		SCO	[244]	0.66	0.68	0.70	0.67	0.71	0.71
Deep-learning	Machine-learning	BEL	[60]	0.65	–	–	0.61	–	–
		gPb	[15]	0.70	0.72	0.66	0.71	0.74	0.65
		DeepNets	[138]	–	–	–	0.74	0.76	0.76
		DeepEdge	[27]	–	–	–	0.75	0.75	0.80
		DeepContour	[204]	–	–	–	0.76	0.77	0.80
		HFL	[28]	–	–	–	0.77	0.79	0.80
		HED	[242]	–	–	–	0.78	0.80	0.83
		SED (Proposed)		**0.69**	**0.71**	**0.71**	**0.71**	**0.74**	**0.74**

Table 5.1 – Results of several edge detection algorithms on the BSDS300 and BSDS500 [15, 165].

In order to study the robustness of different edge detection algorithms in achromatic scenes, we conducted a further experiment on the grey-scale version of BSDS images. The results of this experiment for our model along with five other algorithms driven by low-level features (whose source code were publicly available) are presented in Table 5.2. We can observe similar patterns as chromatic images: the proposed model offers a 3% ODS enhancement in both BSDS300 and BSDS500. A

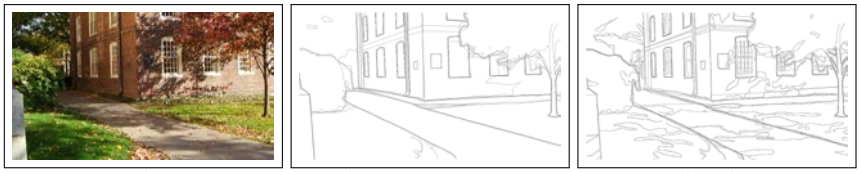
similar improvement can be observed for measures of OIS and AP.

		BSDS300			BSDS500		
Method		ODS	OIS	AP	ODS	OIS	AP
Canny	[41]	0.58	0.62	0.53	0.60	0.63	0.54
PC/BC	[216]	0.61	0.63	0.40	0.64	0.65	0.41
CO	[247]	0.60	0.63	0.60	0.61	0.64	0.61
MCI	[246]	0.62	0.64	0.55	0.64	0.66	0.56
SCO	[244]	0.62	0.64	0.64	0.63	0.67	0.66
SED (Proposed)		**0.65**	**0.67**	**0.68**	**0.67**	**0.70**	**0.70**

Table 5.2 – Results of several edge detection algorithms on the grey-scale images of BSDS300 and BSDS500 [15, 165].

5.3.2 Multi-cue Boundary Detection Dataset (MBDD)

The MBDD [167] is composed of short binocular video sequences in real world environments. This dataset contains challenging scenes for boundary detection by framing a few dominant objects in each shot under a large variety of appearances. Size of each image is 1280×720 pixels. The dataset contains 100 scenes and offers two sets of hand-annotations: one for *object boundaries* and another for *lower-level edges*.



Original Image | Object Boundaries | Low-level Edges

Figure 5.2 – Comparison of *object boundaries* and *low-level edges* annotations of MBDD [167].

We have reported the results of our proposed model along with five algorithms driven by low-level features for both types of annotations in Table 5.3. We can observe that SED offers a 3% ODS improvement in case of *object boundaries* and 2% for *lower-level edges*. We believe that the *object boundaries* annotation is more relevant for the problem we are addressing in this chapter since the *low-level edges*

annotation contains many uninformative line segments from small objects (*e.g.* leaves and grass) as it can be observed from an exemplary image illustrated in Figure 5.2.

		Object Boundaries			Low-level Edges		
Method		ODS	OIS	AP	ODS	OIS	AP
Canny	[41]	0.61	0.65	0.54	0.75	0.78	0.76
PC/BC	[216]	0.69	0.70	0.43	0.80	0.81	0.70
CO	[247]	0.64	0.67	0.66	0.77	0.80	0.83
MCI	[246]	0.69	0.70	0.70	0.77	0.77	0.66
SCO	[244]	0.68	0.71	0.72	0.79	0.82	0.86
SED (Proposed)		**0.72**	**0.74**	**0.77**	**0.81**	**0.83**	**0.86**

Table 5.3 – Results of several edge detection algorithms on the MBDD [167], for two ground truth annotations of *object boundaries* and *low level edges*.

Similar to BSDS, in order to study the role of colour on each algorithm, we performed an experiment on the grey-scale images of MBDD. Table 5.4 shows the results of this experiment. SED still performs better that other algorithms by 1% ODS improvement in both types of annotations. A surprising detail emerges when the results of CO or SCO for colour images is compared to the grey-scale ones; both algorithms perform slightly better in absence of colour (see Tables 5.3 and 5.4). This suggests unbalanced *colour opponency* require more careful implementation. We speculate this might also be the reason that our improvement in the grey-scale images of MBDD falls to minimal. This issue can be addressed in future studies.

		Object Boundaries			Low-level Edges		
Method		ODS	OIS	AP	ODS	OIS	AP
Canny	[41]	0.60	0.65	0.53	0.74	0.78	0.76
PC/BC	[216]	0.68	0.69	0.43	0.79	0.82	0.69
CO	[247]	0.65	0.67	0.67	0.77	0.80	0.83
MCI	[246]	0.69	0.70	0.67	0.73	0.73	0.59
SCO	[244]	0.69	0.71	0.73	0.79	0.82	0.83
SED (Proposed)		**0.70**	**0.71**	**0.74**	**0.80**	**0.82**	**0.86**

Table 5.4 – Results of several edge detection algorithms on the grey-scale images of MBDD [167], for two ground truths of *object boundaries* and *low level edges*.

5.3.3 Contour Image Database (CID)

The CID [107] contains 40 grey-scale images of natural scenes and animal wildlife. Size of each image is 512 × 512 pixels. Table 5.5 compares the results of SED to five algorithms driven by low-level features on this dataset. We can observe that SED exceeds other methods by 5% ODS improvement.

	CID		
Method	ODS	OIS	AP
Canny [41]	0.56	0.64	0.57
PC/BC [216]	0.58	0.62	0.42
CO [247]	0.55	0.63	0.57
MCI [246]	0.60	0.63	0.53
SCO [244]	0.58	0.64	0.61
SED (Proposed)	**0.65**	**0.69**	**0.68**

Table 5.5 – Results of six edge detection algorithms on the CID dataset [107].

5.3.4 Component analysis

In our algorithm we have modelled different areas and aspects of the visual cortex. In order to investigate the contribution of each component of our model, we conducted four additional experiments on the BSDS dataset:

- **Gaussian Derivative** – In this scenario, we accounted neither for the surround modulation in V1, nor for the V2 pooling and feedback. Essentially only convolving the single-opponent cells with the first derivative of Gaussian function similar to CO [247].
- **Only V1 Surround** – In this case, we excluded V2 pooling and feedback. We only included full, far, iso- and orthogonal-orientation surround modulation for V1 RFs.
- **No V2 Feedback** – In this scenario, we excluded the shape feedback sent from area V2 to V1, *i.e.* $c \in \{lu, rg, yb, rg', yb'\}$ in Eq. 5.13.
- **No Far surround** – In this case, we did not account for far surround modulation, *i.e.* $FS = 0$ in Eq. 5.6.

The precision-recall curves of these experiments for BSDS300 and BSDS500 are shown in Figure 5.4. Edge outputs of different components of our algorithm along

with the full model on a few exemplary images are illustrated in Figure 5.5.

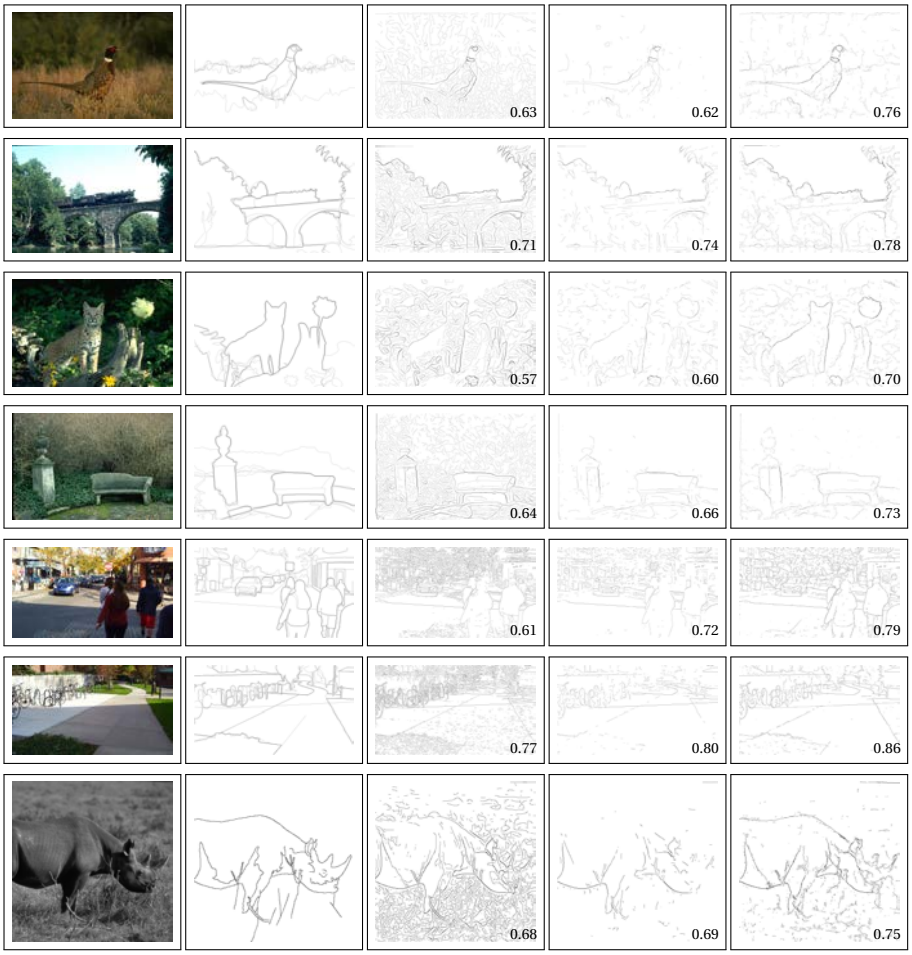


Figure 5.3 – Edge detection results of three biologically-inspired methods. The F-measure is indicated on the right bottom corner. The first two rows are images from BSDS300, the third and forth from BSDS500, the sixth and seventh from MBDD, and the last row from CID.

5.4 Discussion

Results of conducted experiments on three benchmark datasets of edge detection, *i.e.* BSDS, MBDD and CID, demonstrate a systematic quantitative improvement (approximately 4%) for SED over state-of-the-art. Our proposed model outperforms other methods driven by low-level features and biologically-inspired algorithms in all three criteria of ODS, OIS and AP (see Tables 5.1, 5.3 and 5.5). This improvement is also qualitatively pronounced in Figure 5.3. On the one hand, our model shows greater robustness in textural areas in comparison to CO [247], on the other hand, thanks to its surround modulation, SED performs better at detecting continuous lines, compared to SCO [244]. For instance, in the first row of Figure 5.3, it is evident that CO is strongly troubled with the textural information originating from the background vegetation, however SED successfully suppresses a large amount of them. At the same time, it is apparent that SCO blends the contours of the present bird with the straws, however SED correctly extracts the boundaries of the bird from the grassland. We can observe similar patterns in the rest of the pictures of the Figure 5.3.

Our improvements over state-of-the-art originates from a combination of different reasons corresponding to each component of the proposed model. Below, we have discussed each of them separately.

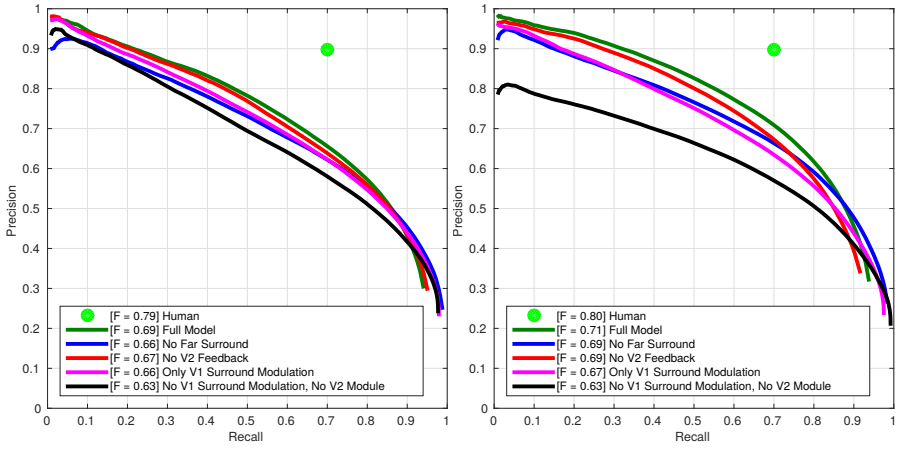


Figure 5.4 – Precision-recall curves of different components of our model on the BSDS300 (left) and BSD500 (right). In the legends the ODS F-measures are indicated.

5.4.1 Surround modulation

The precision-recall curves in Figure 5.4 shows that excluding surround modulation and the V2 module all together drops the ODS F-measure to 0.63 (black curves in both BSDS300 and BSDS500). This is in full agreement with the results of CO [247] which is essentially the same as our model in the absence of both V1 surround modulation and the V2 module. Including surround modulation in its entirety (*i.e.* full, far, iso- and orthogonal-orientation regions) contributes to a significant enhancement of results by boosting the ODS F-measure to 0.66 and 0.67 in BSDS300 and BSDS500, respectively (pink curves). This is clearly an indication that surrounding regions play a crucial role in the process of edge detection. This is to be expected as psychophysical experiments have demonstrated similar phenomenon in our visual perception [152].

Qualitative comparison of the second and third columns of Figure 5.5 suggests that although V1 surround modulation does not contribute to texture suppression, it strengthens continues contours (we have marked a few examples by the red and blue ovals, for instance the exterior borders of bricks in the last row are more continuous in the *"Only V1 Surround"* column in comparison to the *"Gaussian Derivative"* one, at the same time the intermedial borders are correctly suppressed in *"Only V1 Surround"* as a result of accounting for iso- and orthogonal-orientation surround modulation).

5.4.2 V2 module

Comparison of *"Only V1 Surround"* and *"No V2 Feedback"* pictures in Figure 5.5 reveals that the V2 module strongly assist the process of eliminating textural and noisy patches. This is consistent with physiological findings that suggest texture statistics might be encoded in V2 [83, 145]. The robustness of our model to noisy scenes and undesired background textures could be explained by the fact that V2 RFs are large and therefore suppress small discontinuities across neighbouring pixels. Although V2 centre-surround suppression is beneficial in general with 1% (BSDS300) and 2% (BSDS500) improvements in F-measures (the red curves versus the pink ones in Figure 5.4), it causes occasional over-smoothing and consequently in high recalls the precisions of the pink curves are higher than the red ones. We postulate that this problem can partially be addressed by accounting for a mechanism similar to the visual cortex where suppression can turn to facilitation at low contrast levels [14]. Modelling this phenomenon is onerous since the threshold between suppression and facilitation is cell specific and there is no universal contrast level or surround stimulus size that triggers facilitation across the entire cell population [14]. Furthermore, neural recordings of macaque demonstrates that the activation level of V2

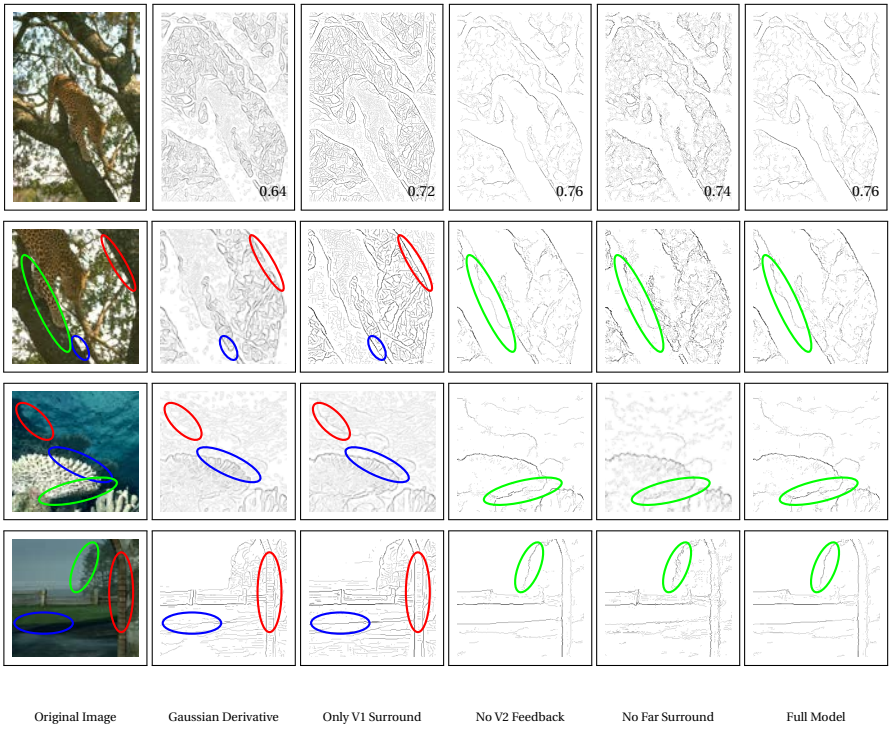


Figure 5.5 – Evaluation of the different components of SED. The images show the result of our full model on one exemplary image along with the four experiments we conducted. F-measures are on the right bottom corner of images.

neurons are higher when exposed to naturalistic texture in comparison to spectrally matched noise [83]. This feature was not present among V1 neurons. This indicates a more complex V2 model is required to treat noise and texture distinctively. We propose this as a line of future work.

5.4.3 Shape feedback

Excluding the global shape feedback from our model lowers the ODS F-measure by 2% (compare the green and red curves in Figure 5.4). It is difficult to appreciate the influence of this feedback connection qualitatively in Figure 5.5, however a close comparison of the green ovals in the *"No V2 Feedback"* and *"Full Model"* columns

suggests that shape feedback re-enforce the true edges (the intensity of pixels along edges are higher in *"Full Model"* in comparison to their corresponding pairs in *"No V2 Feedback"*). This is in line with previously stabilised neurophysiological findings that show one of the functional roles of feedback connections is amplification and focus of neuronal activities in subsequent lower areas [124].

5.4.4 Far surround

The precision-recall curves in Figure 5.4 shows that excluding far surround modulation reduces the ODS F-measure to 0.66 and 0.69 in BSDS300 and BSDS500 respectively (blue curves), which still is better than other non-learning state-of-the-art algorithms. A qualitative comparison of *"No Far Surround"* and *"Full Model"* results in Figure 5.5 reveals that far surround appears to contribute in enhancing continuous edges while suppressing broken ones (*e.g.* the contours marked with green ovals in *"No Far Surround"* contain more abrupt alternate right and left turns in comparison to the *"Full Model"*, at the same time *"No Far Surround"* contains larger number of line fragments). Quantitatively, we observe a similar issue in far surround modulation to the V2 surround modulation: in high recalls *"No Far Surround"* has a higher precision than *"Full Model"* (blue versus green curves in Figure 5.4). Resolving this is a subject for further investigation.

5.4.5 Computational Complexity

In principle, our model ought to be computationally very low cost since its building blocks are simple Gaussian convolutions. With this in mind, we reported the average computational time of six algorithms on the BSDS500 in Table 5.6 and to our surprise, the Matlab implementation of SED is rather slow. After a careful analysis of the different components of our model, we discovered that the *imfilter* function of Matlab is substantially slower when an image is convolved with an oriented elliptical Gaussian kernel across right angles. This is presumably due to the fact that *imfilter* is optimised for separable two-dimensional kernels and behaves significantly slower for non-separable ones. This turned out to be an important issue for our V2 RF surround modulation which uses a kernel of size 157×157 pixels computed for twelve orientations. Since OpenCV *filter2D* does not suffer from this problem, the C++ implementation of our model offers real-time processing. It is worth mentioning that we did not take advantage of any GPU programming in the C++ implementation. We believe our model can greatly benefit from the GPU parallel architecture due to the fact that its basic units are matrix operations.

	Canny [41]	PC/BC [216]	CO [247]	MCI [246]	SCO [244]	**SED (Proposed)**
Time(s)	0.54	$\gg 1800$	0.73	21.00	2.27	7.45 (0.60)*

Table 5.6 – Average computational time (in seconds) of six edge detection algorithms driven by low-level features on the BSDS500 under the Matlab framework with Intel(R) Xeon(R) CPU E5-1620 v2 @ 3.70GHz. *C++ Implementation of our algorithm.

5.5 Conclusion

In this chapter, we presented a biologically-inspired edge detection model grounded on physiological findings of the visual cortex and psychophysical knowledge of visual perception. Our main contributions can be summarised as follows: (i) modelling a contrast-dependant surround modulation of V1 receptive fields by accounting for full, far, iso- and orthogonal-orientation surround; (ii) introducing a V2 centre-surround mechanism to pool V1 signals in their perpendicular orientations; and (iii) accounting for shape feedback connections from V2 to V1. We quantitatively compared the proposed model to current state-of-the-art algorithms on three benchmark datasets (on both colour and grey scale images) and our results show a significant improvement compared to the other non-learning and biologically-inspired models while being competitive to the learning ones. Detailed analysis of different components of our model suggest that V1 surround modulation strengthen edges and continues lines while V2 module contributes to the suppression of undesired textural elements.

Within our framework we treat different surrounding regions disjointly as individual entities with no interactions between them. There are two limitations with this simplification. Firstly, psychophysical studies show that the non-linear interactions between surround and central regions depend on the configurations of both inducers and targets [152]. Secondly, it is well established that perception of shape is significantly influenced by points where multiple edges meet, *e.g.* corners [139]. Consequently, in a more comprehensive model these short and long interactions must be unified under one mechanism by considering configurational settings of full, far, iso- and orthogonal-orientation surrounds. The details of such combinations are still to be investigated. In addition to this, we can further improve our model by accounting for the complex shape processing occurring in V4, for example, by concentric summation of V2 signals [239].

Clausula Part IV

Conclusion

6 Conclusions

Recalling from one of our computer vision master lectures, a teacher once said half-jokingly "if you find a solution to the problem of thresholds, you have solved computer vision". Although, naturally this statement is exaggerated, however, we believe that the core concept carries an important message. Most computer vision and image processing algorithms (including those of convolutional neural networks) consist of a list of parameters which tend to be hard-coded before runtime. The values of these parameters are either established manually or learnt through an optimisation process. At the same time, we are aware that it is rather cumbersome (if not impossible) to set a parameter with a fix value that is optimum for all images and scenarios. Therefore, within the community there is a general consensus [47] that a greater amount of effort is required in order to *dynamically* adapt algorithms' parameters to image or even pixel content. Correspondingly, in this dissertation, we have strived to step towards this direction by proposing more *dynamic* procedures through consideration of *surrounding region* and in particular according to its *local contrast*. There is abundant evidence suggesting that *contrast* plays a crucial role in biological vision to the extent that the human visual system is much more sensitive to contrast than to absolute luminance [21]. Perhaps, it is reasonable to imagine that a similar approach could be of benefit to machine vision as well.

6.1 Summary

We started this doctorate dissertation by discussing the importance of a scene's illuminant in visual perception and image processing. We conducted extensive experiments on a large set of reflectance spectra (gathered from surfaces of real world objects). We investigated their tristimulus values under the illumination of many natural and artificial lights. Based on the results obtained, we reached the conclusion that *metameric pairs* are infrequent in real world scenarios. Therefore, it can be argued that they doe not pose any hindering issue to other high-level visual tasks, such as, the phenomenon of colour constancy.

We continued with the topic of scene's illuminant by proposing a *biologically-inspired colour constancy* model grounded on known properties of neurons in the visual cortex. The framework we developed is constructed over a simple Difference-

of-Gaussians (DoG) kernel with subtle yet influential novelties in order to *dynamise* the DoG computations: (i) the width of the narrower Gaussian (centre) varies between σ and 2σ according to the local contrast of the pixel; (ii) the height of the broader Gaussian (surround) depends on the contrast of the region. Essentially, the main idea is to adjust the band-pass DoG filter according to the local contrast of the centre and surround. Intuitively, in our formulation, the convolutional kernel shifts towards a high-pass filter when applied over specularities and edges resulting in the enhancement of informative pixels, whereas it tends towards a low-pass filter in homogeneous areas in order to represent surround variation.

The next stage of our proposed colour constancy model (and many other models driven by low-level features) is to pool relevant information from the computed feature map (in this case the output of DoG convolution). Once again, inspired by the contrast variability of biological neurons, we proposed a *contrast-variant-pooling* (CVP) mechanism in which a higher percentage of pixels are pooled at low contrast and a smaller percentage at high contrast. This *dynamic* formulation allows for more informative peaks to be pooled while normalising outliers and suppressing irrelevant detail. We examined the efficiency of our proposed *colour constancy* model on four benchmark datasets of single- and multi-illuminant scenes. Our results significantly improved over methods driven by low-level features while being competitive compared to the learning solutions.

We carried on with our study of our colour vision and specifically investigated one of its most common applications, *i.e. colour naming*. We demonstrated that it is possible to capture each colour term through an ellipsoid – a parsimonious geometrical shape. The results of experiments conducted on two benchmark datasets demonstrated that our model can explain colour names in both real world images and psychophysical data. We further argued that the simplicity of the proposed ellipsoids offers a number of benefits, such as: (i) a simple learning subroutine, (ii) a straightforward procedure to add extra colour terms, and (iii) a feasible framework to adapt those ellipsoids to image or pixel contents.

In the last part of this dissertation, we enquired into the impact of surround modulation on *edge detection* because it is well recognised that perception of colour and form are closely entangled. In our proposed model, four types of surrounds are accounted for: (i) full-surround – an isotropical region around the receptive field (pixel), (ii) iso-orientation-surround – an elongated narrow region along the preferred orientation of a receptive field, (iii) orthogonal-orientation-surround – an elongated narrow region perpendicular to the main axis, and (iv) far-surround – corresponding to the inter-areal feedback connections. Similar to the model of colour constancy we proposed, all surround modulations in our edge detection model are also contrast-dependent, *i.e.* more facilitatory at low contrast. Intuitively, the objective of these surround modulations is to enhance collinear edges in order

to form continuous lines, while suppressing undesired textural and noisy areas. We further showed the benefits of two different higher level operations: (i) pooling edges over a large neighbourhood in a direction orthogonal to their preferred orientation (similar to the receptive field of cells at area V2), and (ii) feeding the global shape back to the area V1. We evaluated the proposed *edge detection* model on three benchmark datasets and our results showed a significant overall improvement in comparison to other non-learning methods.

6.2 Contribution

The contributions made by *biologically-inspired* solutions such as those presented throughout this dissertation are twofold:. On the one hand, engineering vision applications can boost their performance and push forward the state-of-the-art by incorporating the large body of knowledge gathered in physiological and psychophysical studies. On the other hand, theoretical models of human visual system can be tested under controlled or realistic conditions, offering a feedback framework for scientific advancements.

To make a concrete example, the proposed colour constancy model surpasses state-of-the-art, including the the results obtains by machine learning solutions, in one benchmark dataset, while at the same time computationally explaining the perceptual findings of Brown *et al.* [38]. This suggests that colour constancy does not depend merely on the average colour of the surround, but also on the distribution of surround colours about the mean. Although further experiments are required, this model potentially allow us to close the missing gap in the circle of "brains", "minds" and "machines" with respect to the role of contrast in the phenomenon of colour constancy.

Broadly speaking, it is believed that the more we learn about the properties of the human visual system the better we can explain visual behaviour and consequently more efficiently transfer them into practical applications. Therefore, within current limitations (both in knowledge and resources) we have tried to keep our modelling decisions as close as possible to what we know about the physiology and psychology of our visual system. This is achieved in three main respects:

1. The architecture of the proposed models reflect the low-level features that are common to mammalian cortical architecture and emerged after millions of years of evolution (*i.e.* are not ad-hoc or dataset-dependant).

2. The parameters of the proposed models are kept identical in all the experiments we conducted across the different datasets, which is a feature of how the human visual system operates.

3. Our low-level models exclude supervised learning from large datasets, which is also a feature of how biological systems operate (their low-level features learning tends to be largely unsupervised).

6.3 Future work

There are plenty of research opportunities (besides those that we mentioned at the end of each chapter) which can be pursued following this dissertation. The models of *surround modulation* we have proposed only incorporate a small fraction of what we know physiologically and psychophysically about the human visual system. We concentrated primarily on the role of *contrast* and to a lesser extent on the *orientation selectivity* of surround modulation. However, there are other parameters (*e.g.* phase, spatial frequency, *etc.*) that must be accounted for in order to achieve a complete surround modulation.

For instance, let us consider the envelope of receptive fields in the primary visual cortex (V1). In our *colour constancy* algorithm we modelled a population of double-opponent cells through DoG, which is an even symmetric function. In our *edge detection* algorithm we modelled another population with the first derivative of Gaussian, which is an odd symmetric function. There is physiological evidence for either type, therefore, a framework must be constructed to combine both types faithfully. In this way the phase sensitivity of the surround can be modelled more appropriately. Furthermore, this might open the door for symmetry detection [178].

Gaetano Kanizsa, a prominent psychologist of the twentieth century, once wrote [133]:

> "… space and colour are not distinct elements but, rather, are interdependent aspects of a unitary process of perceptual organisation."

To the best of our knowledge, there is no work to computationally model this strong coupling of colour and form. This is a testimony to the strenuous challenges involved. In this dissertation, we addressed each separately and our models lack the adjoining component. We believe an attempt towards this direction might result in finding a common solution for the related phenomena known as "colour appearance" (which includes colour induction, colour constancy, naming, *etc.*).

Last but not least, the essence of the proposed *contrast-dependent surround modulation* is task irrelevant, and therefore its implications can be extended to a substantially wider range of applications. This might appear as an extremely bold claim, however, we believe it is worth examining *contrast-dependent Gaussian envelopes* for any arbitrary image processing algorithm that uses constant Gaussians. Along the same line, our contrast-variant-pooling mechanism is a suitable candi-

date to address the shortcomings of the typical max-pooling operator. Therefore, it certainly is interesting to investigate whether our contrast-variant-pooling can improve convolutional neural networks.

Bibliography

[1] Vivek Agarwal, Andrei V Gribok, and Mongi A Abidi. Machine learning approach to color constancy. *Neural Networks*, 20(5):559–563, 2007.

[2] Arash Akbarinia and Karl R Gegenfurtner. Metameric mismatching in natural and artificial reflectances. *Journal of Vision*, 2017.

[3] Arash Akbarinia and Karl R Gegenfurtner. Metamers in real world scenarios. In *International Colour Vision Society*, 2017.

[4] Arash Akbarinia and Alejandro Parraga. Dynamically adjusted surround contrast enhances boundary. In *European Conference on Visual Perception (ECVP)*, volume 45, pages 254–254, 2016.

[5] Arash Akbarinia, Alejandro Parraga, Marta Expósito, Bogdan Raducanu, and Xavier Otazu. Can biological solutions help computers to detect symmetry? In *European Conference on Visual Perception (ECVP)*, 2017.

[6] Arash Akbarinia and C Alejandro Parraga. Biologically plausible colour naming model. In *European Conference on Visual Perception (ECVP)*, volume 44, pages 115–115, 2015.

[7] Arash Akbarinia and C. Alejandro Parraga. Biologically plausible boundary detection. In *British Machine Vision Conference (BMVC)*, September 2016.

[8] Arash Akbarinia and C. Alejandro Parraga. Colour constancy beyond the classical receptive field [under review]. *IEEE transactions on pattern analysis and machine intelligence*, 2017.

[9] Arash Akbarinia and C. Alejandro Parraga. Feedback and surround modulated boundary detection. *International journal of computer vision [Accepted]*, 2017.

[10] Arash Akbarinia, Raquel Gil Rodríguez, and C. Alejandro Parraga. Colour constancy: Biologically-inspired contrast variant pooling mechanism. In *British Machine Vision Conference (BMVC)*, September 2017.

[11] Alexandre Alahi, Georges Goetz, and Emmanuel D'Angelo. Biologically inspired keypoints. *Biologically Inspired Computer Vision: Fundamentals and Applications*, 2015.

[12] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. Freak: Fast retina keypoint. In *Computer Vision and Pattern Recognition, (CVPR)*, pages 510–517, 2012.

[13] John Allman, Francis Miezin, and EveLynn McGuinness. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual review of neuroscience*, 8(1):407–430, 1985.

[14] A Angelucci and S Shushruth. Beyond the classical receptive field: Surround modulation in primary visual cortex. *The new visual neurosciences*, pages 425–444, 2013.

[15] Pablo Arbelaez, Michael Maire, Charless Fowlkes, and Jitendra Malik. Contour detection and hierarchical image segmentation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):898–916, 2011.

[16] Lawrence E Arend, Adam Reeves, James Schirillo, and Robert Goldstein. Simultaneous color constancy: papers with diverse munsell values. *JOSA A*, 8(4):661–672, 1991.

[17] Sarah EJ Arnold, Vincent Savolainen, and Lars Chittka. Fred: the floral reflectance spectra database. *Nature Precedings*, 2008.

[18] Kobus Barnard. Improvements to gamut mapping colour constancy algorithms. In *Computer Vision-ECCV 2000*, pages 390–403. Springer, 2000.

[19] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms. i: Methodology and experiments with synthesized data. *Image Processing, IEEE Transactions on*, 11(9):972–984, 2002.

[20] Kobus Barnard, Lindsay Martin, Brian Funt, and Adam Coath. A data set for color research. *Color Research & Application*, 27(3):147–151, 2002.

[21] Peter GJ Barten. *Contrast sensitivity of the human eye and its effects on image quality*, volume 72. SPIE press, 1999.

[22] Shida Beigpour, Christian Riess, Joost Van de Weijer, and Elli Angelopoulou. Multi-illuminant estimation with conditional random fields. *IEEE Transactions on Image Processing*, 23(1):83–96, 2014.

[23] Robert Benavente and Maria Vanrell. Fuzzy colour naming based on sigmoid membership functions. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2004, pages 135–139. Society for Imaging Science and Technology, 2004.

[24] Robert Benavente, Maria Vanrell, and Ramon Baldrich. Parametric fuzzy sets for automatic color naming. *JOSA A*, 25(10):2582–2593, 2008.

[25] Brent Berlin and Paul Kay. *Basic color terms: Their universality and evolution*. Univ of California Press, 1991.

[26] R. S Berns. Artist paint spectral database. https://www.rit.edu/cos/colorscience/mellon/Publications/Artist_Spectral_Database_CIC2016.pdf. Accessed: 2016-12-01.

[27] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. Deepedge: A multi-scale bifurcated deep network for top-down contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4380–4389, 2015.

[28] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 504–512, 2015.

[29] Simone Bianco, Claudio Cusano, and Raimondo Schettini. Color constancy using cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 81–89, 2015.

[30] Ali Borji and Laurent Itti. Human vs. computer in scene and object recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 113–120, 2014.

[31] Dr PJ Bouma. Physical aspects of colour. 1947.

[32] Y-Lan Boureau, Jean Ponce, and Yann LeCun. A theoretical analysis of feature pooling in visual recognition. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 111–118, 2010.

[33] Robert M Boynton and Conrad X Olson. Salience of chromatic basic color terms confirmed by three measures. *Vision research*, 30(9):1311–1317, 1990.

[34] David H Brainard. Color appearance and color difference specification. *The science of color*, 2:191–216, 2003.

[35] David H Brainard and William T Freeman. Bayesian color constancy. *JOSA A*, 14(7):1393–1411, 1997.

[36] David H Brainard and A Radonjic. Color constancy. *The visual neurosciences*, 1:948–961, 2004.

[37] Farran Briggs and Usrey Martin. Functional properties of cortical feedback to the primate lateral geniculate nucleus. *Werner JS Chalupa L.(Eds.) The new visual neurosciences*, pages 315–322, 2014.

[38] Richard O Brown and Donald IA MacLeod. Color appearance depends on the variance of surround colors. *Current Biology*, 7(11):844–849, 1997.

[39] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.

[40] Gershon Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1):1–26, 1980.

[41] John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):679–698, 1986.

[42] M. Carandini. Area V1. *Scholarpedia*, 7(7):12105, 2012. revision #126411.

[43] Matteo Carandini and David J Heeger. Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13(1):51–62, 2012.

[44] Vlad C Cardei, Brian Funt, and Kobus Barnard. Estimating the scene illumination chromaticity by using a neural network. *JOSA A*, 19(12):2374–2386, 2002.

[45] James R Cavanaugh, Wyeth Bair, and J Anthony Movshon. Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of neurophysiology*, 88(5):2530–2546, 2002.

[46] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler. Color constancy with spatio-spectral statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8):1509–1519, 2012.

[47] Andrzej Cichocki and Shun-ichi Amari. *Adaptive blind signal and image processing: learning algorithms and applications*, volume 1. John Wiley & Sons, 2002.

[48] CIE. *Colorimetry*, volume 15. CIE Publication, 3 edition, 2004.

[49] Florian Ciurea and Brian Funt. A large image database for color constancy research. In *Color and Imaging Conference*, volume 2003, pages 160–164. Society for Imaging Science and Technology, 2003.

[50] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

[51] Bevil R Conway, Soumya Chatterjee, Greg D Field, Gregory D Horwitz, Elizabeth N Johnson, Kowa Koida, and Katherine Mancuso. Advances in color science: from retina to behavior. *The Journal of Neuroscience*, 30(45):14955–14963, 2010.

[52] Timothee Cour, Florence Benezit, and Jianbo Shi. Spectral segmentation with multiscale graph decomposition. In *Computer Vision and Pattern Recognition, (CVPR)*, volume 2, pages 1124–1131, 2005.

[53] Tim Crane. *The mechanical mind: A philosophical introduction to minds, machines and mental representation.* Routledge, 2015.

[54] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, (CVPR)*, volume 1, pages 886–893, 2005.

[55] Gunilla Derefeldt and Tiina Swartling. Colour concept retrieval by free colour naming. identification of up to 30 colours without training. *Displays*, 16(2):69–77, 1995.

[56] Francisco J Díaz-Pernas, Mario Martínez-Zarzuela, Míriam Antón-Rodríguez, and David González-Ortega. Double recurrent interaction v1–v2–v4 based neural architecture for color natural scene boundary detection and surface perception. *Applied Soft Computing*, 21:250–264, 2014.

[57] James J DiCarlo, Davide Zoccolan, and Nicole C Rust. How does the brain solve visual object recognition? *Neuron*, 73(3):415–434, 2012.

[58] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, pages 1–6. IEEE, 2016.

[59] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. *arXiv preprint arXiv:1705.02498*, 2017.

[60] Piotr Dollar, Zhuowen Tu, and Serge Belongie. Supervised learning of edges and object boundaries. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1964–1971. IEEE, 2006.

[61] Piotr Dollár and C Lawrence Zitnick. Fast edge detection using structured forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(8):1558–1570, 2015.

[62] J. Dowling. Retina. *Scholarpedia*, 2(12):3487, 2007. revision #91715.

[63] Marc Ebner. *Color constancy*, volume 6. John Wiley & Sons, 2007.

[64] Johan Maurice Gis ele Lammens. *A computational model of color perception and color naming*. PhD thesis, Citeseer, 1994.

[65] Christina Enroth-Cugell and John G Robson. The contrast sensitivity of retinal ganglion cells of the cat. *The Journal of physiology*, 187(3):517–552, 1966.

[66] Azim Eskandarian. *Handbook of intelligent vehicles*. Springer, 2014.

[67] Mark Everingham, Luc Van Gool, Christopher KI Williams, John Winn, and Andrew Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88(2):303–338, 2010.

[68] Mark D Fairchild. *Color appearance models*. John Wiley & Sons, 2013.

[69] Daniel J Felleman and David C Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral cortex (New York, NY: 1991)*, 1(1):1–47, 1991.

[70] Pedro F Felzenszwalb and Daniel P Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[71] David J Field, James R. Golden, and Anthony Hayes. Contour integration and the association field. *Werner JS Chalupa L.(Eds.) The new visual neurosciences*, pages 627–638, 2014.

[72] Greg D Field, Jeffrey L Gauthier, Alexander Sher, Martin Greschner, Timothy A Machado, Lauren H Jepson, Jonathon Shlens, Deborah E Gunning, Keith Mathieson, Wladyslaw Dabrowski, et al. Functional connectivity in the retina at the resolution of photoreceptors. *Nature*, 467(7316):673–677, 2010.

[73] Graham Finlayson and Steven Hordley. Improving gamut mapping color constancy. *Image Processing, IEEE Transactions on*, 9(10):1774–1783, 2000.

[74] Graham D. Finlayson. Color in perspective. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(10):1034–1038, 1996.

[75] Graham D Finlayson, Steven D Hordley, and Mark S Drew. Removing shadows from images. In *Computer Vision—ECCV 2002*, pages 823–836. Springer, 2002.

[76] Graham D Finlayson and Peter Morovic. Metamer sets. *JOSA A*, 22(5):810–819, 2005.

[77] Graham D Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[78] Graham D Finlayson and Roshanak Zakizadeh. Reproduction angular error: An improved performance metric for illuminant estimation. *perception*, 310(1):1–26, 2014.

[79] François Fleuret, Ting Li, Charles Dubout, Emma K Wampler, Steven Yantis, and Donald Geman. Comparing machines and humans on a visual categorization test. *Proceedings of the National Academy of Sciences*, 108(43):17621–17625, 2011.

[80] David A Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1):5–35, 1990.

[81] David H Foster. Color constancy. *Vision research*, 51(7):674–700, 2011.

[82] David H Foster, Kinjiro Amano, Sérgio MC Nascimento, and Michael J Foster. Frequency of metamerism in natural scenes. *JOSA A*, 23(10):2359–2372, 2006.

[83] Jeremy Freeman, Corey M Ziemba, David J Heeger, Eero P Simoncelli, and J Anthony Movshon. A functional and perceptual signature of the second visual area in primates. *Nature neuroscience*, 16(7):974–981, 2013.

[84] William T. Freeman and Edward H Adelson. The design and use of steerable filters. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (9):891–906, 1991.

[85] Itzhak Fried, Ueli Rutishauser, Moran Cerf, and Gabriel Kreiman. *Single neuron studies of the human brain: probing cognition.* MIT Press, 2014.

[86] Kunihiko Fukushima and Sei Miyake. Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets*, pages 267–285. Springer, 1982.

[87] Brian Funt, Kobus Barnard, and Lindsay Martin. Is machine colour constancy good enough? In *Computer Vision—ECCV'98*, pages 445–459. Springer, 1998.

[88] Brian Funt and Milan Mosny. Removing outliers in illumination estimation. In *Color and Imaging Conference*, volume 2012, pages 105–110. Society for Imaging Science and Technology, 2012.

[89] Brian Funt and Weihua Xiong. Estimating illumination chromaticity via support vector regression. In *Color and Imaging Conference*, volume 2004, pages 47–52. Society for Imaging Science and Technology, 2004.

[90] Brian V Funt, Mark S Drew, and Jian Ho. Color constancy from mutual reflection. *International Journal of Computer Vision*, 6(1):5–24, 1991.

[91] Shao-Bing Gao, Kai-Fu Yang, Chao-Yi Li, and Yong-Jie Li. Color constancy using double-opponency. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 37(10):1973–1985, 2015.

[92] Shaobing Gao, Wangwang Han, Kaifu Yang, Chaoyi Li, and Yongjie Li. Efficient color constancy with local surface reflectance statistics. In *Computer Vision–ECCV 2014*, pages 158–173. Springer, 2014.

[93] Shaobing Gao, Kaifu Yang, Chaoyi Li, and Yongjie Li. A color constancy model with double-opponency mechanisms. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 929–936, 2013.

[94] Karl R Gegenfurtner. Cortical mechanisms of colour vision. *Nature Reviews Neuroscience*, 4(7):563–572, 2003.

[95] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *Computer Vision and Pattern Recognition, (CVPR)*, pages 1–8, 2008.

[96] Stuart Geman, Elie Bienenstock, and René Doursat. Neural networks and the bias/variance dilemma. *Neural computation*, 4(1):1–58, 1992.

[97] Theo Gevers, Arjan Gijsenij, Joost Van de Weijer, and Jan-Mark Geusebroek. *Color in computer vision: fundamentals and applications*, volume 23. John Wiley & Sons, 2012.

[98] Theo Gevers and Arnold WM Smeulders. Color-based object recognition. *Pattern recognition*, 32(3):453–464, 1999.

[99] Theo Gevers and Arnold WM Smeulders. Pictoseek: combining color and shape invariant features for image retrieval. *Image Processing, IEEE Transactions on*, 9(1):102–119, 2000.

[100] Arjan Gijsenij and Theo Gevers. Color constancy using natural image statistics and scene semantics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(4):687–698, 2011.

[101] Arjan Gijsenij, Theo Gevers, and Marcel P Lucassen. Perceptual analysis of distance measures for color constancy algorithms. *JOSA A*, 26(10):2243–2256, 2009.

[102] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Generalized gamut mapping using image derivative structures for color constancy. *International Journal of Computer Vision*, 86(2-3):127–139, 2010.

[103] Arjan Gijsenij, Theo Gevers, and Joost Van De Weijer. Improving color constancy by photometric edge weighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(5):918–929, 2012.

[104] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *IEEE Transactions on Image Processing*, 21(2):697–707, 2012.

[105] E Bruce Goldstein and James Brockmole. *Sensation and perception*. Cengage Learning, 2016.

[106] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006.

[107] Cosmin Grigorescu, Nicolai Petkov, and Michel A Westenberg. Contour detection based on nonclassical receptive field inhibition. *Image Processing, IEEE Transactions on*, 12(7):729–739, 2003.

[108] C. G. Gross. Inferior temporal cortex. *Scholarpedia*, 3(12):7294, 2008. revision #91373.

[109] J Haanpalo. Paper spectra. http://www.uef.fi/en/web/spectral/paper-spectra. Accessed: 2016-12-01.

[110] Thorsten Hansen, Maria Olkkonen, Sebastian Walter, and Karl R Gegenfurtner. Memory modulates color appearance. *Nature neuroscience*, 9(11):1367–1368, 2006.

[111] Clyde L Hardin and Luisa Maffi. *Color categories in thought and language*. Cambridge University Press, 1997.

[112] James V Haxby, Elizabeth A Hoffman, and M Ida Gobbini. The distributed human neural system for face perception. *Trends in cognitive sciences*, 4(6):223–233, 2000.

[113] David J Heeger. Normalization of cell responses in cat striate cortex. *Visual neuroscience*, 9(02):181–197, 1992.

[114] Jeanny Herault, Gabriel Cristobal, Laurent Perrinet, and Matthias S Keil. *Biologically Inspired Computer Vision: Fundamentals and Applications*. John Wiley & Sons, 2015.

[115] Robert F Hess. Spatial scale in visual processing. *Werner JS Chalupa L.(Eds.) The new visual neurosciences*, pages 595–615, 2014.

[116] J Hiltunen. Lumber spectra. http://www.uef.fi/en/web/spectral/lumber-spectra. Accessed: 2016-12-01.

[117] Steven D Hordley. Scene illuminant estimation: past, present, and future. *Color Research & Application*, 31(4):303–314, 2006.

[118] Steven D Hordley and Graham D Finlayson. Reevaluation of color constancy algorithm performance. *JOSA A*, 23(5):1008–1020, 2006.

[119] Gregory D Horwitz and Charles A Hass. Nonlinear analysis of macaque v1 color tuning reveals cardinal directions for cortical color processing. *Nature Neuroscience*, 15(6):913–919, 2012.

[120] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574–591, 1959.

[121] David H Hubel and Torsten N Wiesel. *Brain and visual perception: the story of a 25-year collaboration*. Oxford University Press, 2004.

[122] Paul M Hubel. The perception of color at dawn and dusk. *Journal of Imaging Science and Technology*, 44(4):371–375, 2000.

[123] Oliver J Hulme and S Zeki. The sightless view: neural correlates of occluded objects. *Cerebral Cortex*, 17(5):1197–1205, 2006.

[124] JM Hupe, AC James, BR Payne, SG Lomber, P Girard, and J Bullier. Cortical feedback improves discrimination between figure and background by v1, v2 and v3 neurons. *Nature*, 394(6695):784–787, 1998.

[125] Anya Hurlbert and Kit Wolf. Color contrast: a contributory mechanism to color constancy. *Progress in brain research*, 144:145–160, 2004.

[126] Jennifer M Ichida, Lars Schwabe, Paul C Bressloff, and Alessandra Angelucci. Response facilitation from the "suppressive" receptive field surround of macaque v1 neurons. *Journal of Neurophysiology*, 98(4):2168–2181, 2007.

[127] Tarow Indow. Multidimensional studies of munsell color solid. *Psychological Review*, 95(4):456, 1988.

[128] Laurent Itti and Christof Koch. Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3):194–203, 2001.

[129] Timo Jaaskelainen, R Silvennoinen, J Hiltunen, and Jussiu PS Parkkinen. Classification of the reflectance spectra of pine, spruce, and birch. *Applied Optics*, 33(12):2356–2362, 1994.

[130] Kevin Jarrett, Koray Kavukcuoglu, Yann LeCun, et al. What is the best multi-stage architecture for object recognition? In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2146–2153. IEEE, 2009.

[131] Hamid Reza Vaezi Joze and Mark S Drew. Exemplar-based color constancy and multiple illumination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 36(5):860–873, 2014.

[132] Hamid Reza Vaezi Joze, Mark S Drew, Graham D Finlayson, and Perla Aurora Troncoso Rey. The role of bright pixels in illumination estimation. In *Color and Imaging Conference*, volume 2012, pages 41–46. Society for Imaging Science and Technology, 2012.

[133] Gaetano Kanizsa. *Organization in vision: Essays on Gestalt perception*. Praeger Publishers, 1979.

[134] Mitesh K Kapadia, Minami Ito, Charles D Gilbert, and Gerald Westheimer. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in v1 of alert monkeys. *Neuron*, 15(4):843–856, 1995.

[135] Mitesh K Kapadia, Gerald Westheimer, and Charles D Gilbert. Dynamics of spatial summation in primary visual cortex of alert monkeys. *Proceedings of the National Academy of Sciences*, 96(21):12073–12078, 1999.

[136] Paul Kay and Chad K McDaniel. The linguistic significance of the meanings of basic color terms. *Language*, pages 610–646, 1978.

[137] Paul Kay and Terry Regier. Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15):9085–9089, 2003.

[138] Jyri J Kivinen, Christopher KI Williams, Nicolas Heess, and DeepMind Technologies. Visual boundary prediction: A deep neural prediction network and quality dissection. In *AISTATS*, volume 1, page 9, 2014.

[139] Jan J Koenderink and Andrea J Van Doorn. The shape of smooth objects and the way contours end. *Perception*, 11(2):129–137, 1982.

[140] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[141] Jonas Kubilius, Stefania Bracci, and Hans P Op de Beeck. Deep neural networks as a computational model for human shape sensitivity. *PLoS Comput Biol*, 12(4):e1004896, 2016.

[142] Ilan Lampl, David Ferster, Tomaso Poggio, and Maximilian Riesenhuber. Intracellular measurements of spatial integration and the max operation in complex cells of the cat primary visual cortex. *Journal of neurophysiology*, 92(5):2704–2713, 2004.

[143] Edwin 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*, 83(10):3078–3080, 1986.

[144] Edwin H Land et al. *The retinex theory of color vision*. Citeseer, 1977.

[145] Michael S Landy. Texture analysis and perception. *Werner JS Chalupa L.(Eds.) The new visual neurosciences*, pages 639–652, 2014.

[146] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.

[147] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[148] Hsien-Che Lee. Method for computing the scene-illuminant chromaticity from specular highlights. *JOSA A*, 3(10):1694–1699, 1986.

[149] Peter Lennie. Single units and visual cortical organization. *Perception*, 27(8):889–935, 1998.

[150] Martin D Levine. *Vision in man and machine*. McGraw-Hill College, 1985.

[151] H Lin, MR Luo, LW MacDonald, and AWS Tarrant. A cross-cultural colour-naming study: Part ii–using a constrained method. *Color Research & Application*, 26(3):193–208, 2001.

[152] Gunter Loffler. Perception of contours and shapes: Low and intermediate stage mechanisms. *Vision Research*, 48(20):2106–2127, 2008.

[153] Alexander D Logvinenko, Brian Funt, and Christoph Godau. Metamer mismatching. *IEEE Transactions on Image Processing*, 23(1):34–43, 2014.

[154] Alexander D Logvinenko, Brian Funt, Hamidreza Mirzaei, and Rumi Tokunaga. Rethinking colour constancy. *PloS one*, 10(9):e0135029, 2015.

[155] Zhongyu Lou, Theo Gevers, Ninghang Hu, and Marcel P. Lucassen. Color constancy by deep learning. In *British Machine Vision Conference (BMVC)*, pages 1–712, September 2015.

[156] M Ronnier Luo, Guihua Cui, and B Rigg. The development of the cie 2000 colour-difference formula: Ciede2000. *Color Research & Application*, 26(5):340–350, 2001.

[157] David L MacAdam. Visual sensitivities to color differences in daylight. *JOSA*, 32(5):247–274, 1942.

[158] David L MacAdam. *Sources of color science*. Mit Press, 1970.

[159] Robert E MacLaury, Gordon W Hewes, Paul R Kinnear, JB Deregowski, William R Merrifield, BAC Saunders, James Stanlaw, Christina Toren, J Van Brakel, and Roger W Wescott. From brightness to hue: An explanatory model of color-category evolution [and comments and reply]. *Current Anthropology*, pages 137–186, 1992.

[160] Mohammad Hossein Maghami, Amir Masoud Sodagar, Alireza Lashay, Hamid Riazi-Esfahani, and Mohammad Riazi-Esfahani. Visual prostheses: the enabling technology to give sight to the blind. *Journal of ophthalmic & vision research*, 9(4):494, 2014.

[161] R Malach, Y Amir, M Harel, and A Grinvald. Relationship between intrinsic connections and functional architecture revealed by optical imaging and in vivo targeted biocytin injections in primate striate cortex. *Proceedings of the National Academy of Sciences*, 90(22):10469–10473, 1993.

[162] Laurence T Maloney and Brian A Wandell. Color constancy: a method for recovering surface spectral reflectance. *JOSA A*, 3(1):29–33, 1986.

[163] David Marr and Ellen Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London B: Biological Sciences*, 207(1167):187–217, 1980.

[164] E Marszalec. Agfa it8.7/2 set. http://www.uef.fi/en/web/spectral/agfa-it8.7/2-set.

[165] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 416–423. IEEE, 2001.

[166] Yoshifumi Matsumoto, Chihiro Hiramatsu, Yuka Matsushita, Norihiro Ozawa, Ryuichi Ashino, Makiko Nakata, Satoshi Kasagi, Anthony Di Fiore, Colleen M Schaffner, Filippo Aureli, Amanda D. Melin, and Shoji Kawamura. Evolutionary renovation of l/m opsin polymorphism confers a fruit discrimination advantage to ateline new world monkeys. *Molecular ecology*, 23(7):1799–1812, 2014.

[167] David A Mély, Junkyung Kim, Mason McGill, Yuliang Guo, and Thomas Serre. A systematic comparison between visual cues for boundary detection. *Vision research*, 120:93–107, 2016.

[168] Aleksandra Mojsilovic. A computational model for color naming and describing color composition of images. *Image Processing, IEEE Transactions on*, 14(5):690–699, 2005.

[169] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi, and Pascal Frossard. Deepfool: a simple and accurate method to fool deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2574–2582, 2016.

[170] Milan Mosny and Brian Funt. Cubical gamut mapping colour constancy. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2010, pages 466–470. Society for Imaging Science and Technology, 2010.

[171] Naila Murray and Florent Perronnin. Generalized max pooling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2473–2480, 2014.

[172] Naila Murray, Maria Vanrell, Xavier Otazu, and C Alejandro Parraga. Saliency estimation using a non-parametric low-level vision model. In *Computer Vision and Pattern Recognition, (CVPR)*, pages 433–440, 2011.

[173] Isaac Newton. *Opticks, or, a treatise of the reflections, refractions, inflections & colours of light*. Courier Corporation, 1979.

[174] Anh Nguyen, Jason Yosinski, and Jeff Clune. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 427–436, 2015.

[175] Bruno A Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[176] Bruno A Olshausen and JD Field. What is the other 85 percent of v1 doing. *L. van Hemmen, & T. Sejnowski (Eds.)*, 23:182–211, 2006.

[177] J Orava. Munsell colors glossy (all) (spectrofotometer measured). http://www.uef.fi/en/web/spectral/munsell-colors-glossy-all-spectrofotometer-measured. Accessed: 2016-12-01.

[178] D Osorio. Symmetry detection by categorization of spatial phase, a model. *Proceedings of the Royal Society of London B: Biological Sciences*, 263(1366):105–110, 1996.

[179] Xavier Otazu, C Alejandro Parraga, and Maria Vanrell. Toward a unified chromatic induction model. *Journal of Vision*, 10(12):5–5, 2010.

[180] Xavier Otazu, Maria Vanrell, and C Alejandro Párraga. Multiresolution wavelet framework models brightness induction effects. *Vision Research*, 48(5):733–751, 2008.

[181] Giuseppe Papari and Nicolai Petkov. Edge and line oriented contour detection: State of the art. *Image and Vision Computing*, 29(2):79–103, 2011.

[182] Seymour A Papert. The summer vision project. 1966.

[183] J Parkkinen, T Jaaskelainen, and M Kuittinen. Spectral representation of color images. In *Pattern Recognition, 1988., 9th International Conference on*, pages 933–935. IEEE, 1988.

[184] C Alejandro Parraga. Color vision, computational methods for. *Encyclopedia of Computational Neuroscience, Ed. D. Jaeger and R. Jung, SpringerReference*, 10:58, 2013.

[185] C Alejandro Parraga and Arash Akbarinia. Colour constancy as a product of dynamic centre-surround adaptation. *Journal of Vision*, 16(12):214–214, 2016.

[186] C. Alejandro Parraga and Arash Akbarinia. Nice: A computational solution to close the gap from colour perception to colour categorization. *PLoS ONE*, 11:1–32, 03 2016.

[187] Michael R Pointer. On the number of discernible colours. *Color Research & Application*, 23(5):337–337, 1998.

[188] F Poirier and H R Wilson. A biologically plausible model of human radial frequency perception. *Vision research*, 46(15):2443–2455, 2006.

[189] Judith MS Prewitt. Object enhancement and extraction. *Picture processing and Psychopictorics*, 10(1):15–19, 1970.

[190] BC Regan, C Julliot, B Simmen, F Vienot, P Charles-Dominique, and JD Mollon. Frugivory and colour vision in alouatta seniculus, a trichromatic platyrrhine monkey. *Vision research*, 38(21):3321–3327, 1998.

[191] Terry Regier, Paul Kay, and Naveen Khetarpal. Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences*, 104(4):1436–1441, 2007.

[192] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 267–276. ACM, 2002.

[193] John-Paul Renno, Dimitrios Makris, Tim Ellis, and Graeme A Jones. Application and evaluation of colour constancy in visual surveillance. In *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, pages 301–308. IEEE, 2005.

[194] Maximilian Riesenhuber and Tomaso Poggio. Hierarchical models of object recognition in cortex. *Nature neuroscience*, 2(11):1019–1025, 1999.

[195] Jordi Roca-Vila, C Alejandro Parraga, and Maria Vanrell. Chromatic settings and the structural color constancy index. *Journal of vision*, 13(4):3–3, 2013.

[196] Charles Rosenberg, Alok Ladsariya, and Tom Minka. Bayesian color constancy with non-gaussian models. In *Advances in neural information processing systems*, page None, 2003.

[197] Multidimensional Scaling. Chapman & hall. *CRC, Boca Raton, FL*, 2001.

[198] Dominik Scherer, Andreas Müller, and Sven Behnke. Evaluation of pooling operations in convolutional architectures for object recognition. *Artificial Neural Networks–ICANN 2010*, pages 92–101, 2010.

[199] Matthew Seaborn, Lee Hepplewhite, and John Stonham. Fuzzy colour category map for the measurement of colour similarity and dissimilarity. *Pattern Recognition*, 38(2):165–177, 2005.

[200] Thomas Serre and Maximilian Riesenhuber. Realistic modeling of simple and complex cell tuning in the hmax model, and implications for invariant object recognition in cortex. Technical report, DTIC Document, 2004.

[201] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio. Robust object recognition with cortex-like mechanisms. *IEEE transactions on pattern analysis and machine intelligence*, 29(3), 2007.

[202] Thomas Serre, Lior Wolf, and Tomaso Poggio. Object recognition with features inspired by visual cortex. In *Computer Vision and Pattern Recognition, (CVPR)*, volume 2, pages 994–1000, 2005.

[203] Robert Shapley and Michael J Hawken. Color in the cortex: single-and double-opponent cells. *Vision research*, 51(7):701–717, 2011.

[204] Wei Shen, Xinggang Wang, Yan Wang, Xiang Bai, and Zhijiang Zhang. Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3982–3991, 2015.

[205] L. Shi and B. Funt. Re-processed version of the gehler color constancy dataset of 568 images. http://www.cs.sfu.ca/~colour/data/.

[206] Lilong Shi, Weihua Xiong, and Brian Funt. Illumination estimation via thin-plate spline interpolation. *JOSA A*, 28(5):940–948, 2011.

[207] S Shushruth, Jennifer M Ichida, Jonathan B Levitt, and Alessandra Angelucci. Comparison of spatial summation properties of neurons in macaque v1 and v2. *Journal of neurophysiology*, 102(4):2069–2083, 2009.

[208] S Shushruth, Jennifer M Ichida, Jonathan B Levitt, and Alessandra Angelucci. Comparison of spatial summation properties of neurons in macaque v1 and v2. *Journal of Neurophysiology*, 102(4):2069–2083, 2009.

[209] S Shushruth, Lauri Nurminen, Maryam Bijanzadeh, Jennifer M Ichida, Simo Vanni, and Alessandra Angelucci. Different orientation tuning of near-and far-surround suppression in macaque primary visual cortex mirrors their tuning in human perception. *The Journal of Neuroscience*, 33(1):106–119, 2013.

[210] Wai Ting Siok, Paul Kay, William SY Wang, Alice HD Chan, Lin Chen, Kang-Kwong Luke, and Li Hai Tan. Language regions of brain are operative in color perception. *Proceedings of the National Academy of Sciences*, 106(20):8140–8145, 2009.

[211] Orit Skorka and Dileepan Joseph. Toward a digital camera to rival the human eye. *Journal of Electronic Imaging*, 20(3):033009–033009, 2011.

[212] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.

[213] Elizabeth S Spelke. Principles of object perception. *Cognitive science*, 14(1):29–56, 1990.

[214] Hedva Spitzer and Yuval Barkan. Computational adaptation model and its predictions for color induction of first and second orders. *Vision Research*, 45(27):3323–3342, 2005.

[215] Hedva Spitzer and Sarit Semo. Color constancy: a biological model and its application for still and video images. *Pattern Recognition*, 35(8):1645–1659, 2002.

[216] Michael W Spratling. Image segmentation using a sparse coding model of cortical area v1. *Image Processing, IEEE Transactions on*, 22(4):1631–1643, 2013.

[217] Julia Sturges and TW Whitfield. Locating basic colours in the munsell space. *Color Research & Application*, 20(6):364–376, 1995.

[218] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *International Conference on Learning Representations (ICLR)*, 2014.

[219] Christian Thériault, Nicolas Thome, and Matthieu Cord. Cortical networks of visual recognition. *Biologically Inspired Computer Vision: Fundamentals and Applications*, 2015.

[220] Wiliam A. Thornton. Matching lights, metamers, and human visual response. *Journal of Color & Appear- ance*, 2(1):23–29, 1973.

[221] Carlo Tomasi and Roberto Manduchi. Bilateral filtering for gray and color images. In *Computer Vision, 1998. Sixth International Conference on*, pages 839–846. IEEE, 1998.

[222] Shoji Tominaga. A colour-naming method for computer color vision. In *Proceedings of the 1985 IEEE International Conference on Cybernetics and Society*, volume 573, page 577, 1985.

[223] Peter Ulric Tse. Volume completion. *Cognitive psychology*, 39(1):37–68, 1999.

[224] Philipp Urban and Roy S Berns. Paramer mismatch-based spectral gamut mapping. *IEEE Transactions on Image Processing*, 20(6):1599–1610, 2011.

[225] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007.

[226] Joost Van De Weijer, Theo Gevers, and Arjan Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9):2207–2214, 2007.

[227] Joost van de Weijer and Fahad Shahbaz Khan. An overview of color name applications in computer vision. In *Computational Color Imaging*, pages 16–22. Springer, 2015.

[228] Joost Van De Weijer, Cordelia Schmid, Jakob Verbeek, and Diane Larlus. Learning color names for real-world applications. *Image Processing, IEEE Transactions on*, 18(7):1512–1523, 2009.

[229] Javier Vazquez-Corral, C Párraga, Ramon Baldrich, and Maria Vanrell. Color constancy algorithms: Psychophysical evaluation on a new dataset. *Journal of Imaging Science and Technology*, 53(3):31105–1, 2009.

[230] Javier Vazquez-Corral, Maria Vanrell, Ramon Baldrich, and Francesc Tous. Color constancy by category correlation. *Image Processing, IEEE Transactions on*, 21(4):1997–2007, 2012.

[231] Johannes Von Kries. Chromatic adaptation. *Festschrift der Albrecht-Ludwigs-Universität*, 135:145–158, 1902.

[232] Johan Wagemans, Joeri De Winter, Hans Op de Beeck, Annemie Ploeger, Tom Beckers, and Peter Vanroose. Identification of everyday objects on the basis of silhouette and outline versions. *Perception*, 37(2):207–244, 2008.

[233] Gary A Walker, Izumi Ohzawa, and Ralph D Freeman. Asymmetric suppression outside the classical receptive field of the visual cortex. *The Journal of Neuroscience*, 19(23):10536–10553, 1999.

[234] Dirk B Walther, Barry Chai, Eamon Caddigan, Diane M Beck, and Li Fei-Fei. Simple line drawings suffice for functional mri decoding of natural scene categories. *Proceedings of the National Academy of Sciences*, 108(23):9661–9666, 2011.

[235] Hui Wei, Bo Lang, and Qingsong Zuo. Contour detection model with multi-scale integration based on non-classical receptive field. *Neurocomputing*, 103:247–262, 2013.

[236] Juyang Weng. *Natural and artificial intelligence*. BMI Press, Okemos, 2012.

[237] John Simon Werner and Leo M Chalupa. *The new visual neurosciences*. Mit Press, 2014.

[238] Stephen Westland, Julian Shaw, and Huw Owens. Colour statistics of natural and man-made surfaces. *Sensor Review*, 20(1):50–55, 2000.

[239] HR Wilson and F Wilkinson. Configural pooling in the ventral pathway. *New visual neurosciences*, pages 617–626, 2014.

[240] Christoph Witzel, Carlijn van Alphen, Christoph Godau, and J Kevin O'Regan. Uncertainty of sensory signal explains variation of color constancy. *Journal of Vision*, 16(15):8–8, 2016.

[241] Gunter Wyszecki and Walter Stanley Stiles. *Color science*, volume 8. Wiley New York, 1982.

[242] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1395–1403, 2015.

[243] Jianchao Yang, Kai Yu, Yihong Gong, and Thomas Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Computer Vision and Pattern Recognition, (CVPR)*, pages 1794–1801, 2009.

[244] Kai-Fu Yang, Shao-Bing Gao, Ce-Feng Guo, Chao-Yi Li, and Yong-Jie Li. Boundary detection using double-opponency and spatial sparseness constraint. *Image Processing, IEEE Transactions on*, 24(8):2565–2578, 2015.

[245] Kai-Fu Yang, Shao-Bing Gao, and Yong-Jie Li. Efficient illuminant estimation for color constancy using grey pixels. In *Computer Vision and Pattern Recognition, (CVPR)*, pages 2254–2263, 2015.

[246] Kai-Fu Yang, Chao-Yi Li, and Yong-Jie Li. Multifeature-based surround inhibition improves contour detection in natural images. *Image Processing, IEEE Transactions on*, 23(12):5020–5032, 2014.

[247] Kaifu Yang, Shaobing Gao, Chaoyi Li, and Yongjie Li. Efficient color boundary detection with color-opponent mechanisms. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2810–2817, 2013.

[248] Zejian Yuan, Badong Chen, Jianru Xue, Nanning Zheng, et al. Illumination robust color naming via label propagation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 621–629, 2015.

[249] Semir Zeki. *A Vision of the Brain*. Oxford Univ Press, 1993.

[250] Jun Zhang, Youssef Barhomi, and Thomas Serre. A new biologically inspired color image descriptor. In *Computer Vision–ECCV 2012*, pages 312–324. Springer, 2012.

[251] Xiandou Zhang, Brian Funt, and Hamidreza Mirzaei. Metamer mismatching in practice versus theory. *JOSA A*, 33(3):A238–A247, 2016.