Contribution of colour in guiding visual attention and in a computational model of visual saliency

Shahrbanoo Talebzadeh Shahrbabaki

UNIVERSITÉ GRENOBLE ALPES

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Signal and Image Processing**

Arrêté ministériel : 7 août 2006

Présentée par

Shahrbanoo Talebzadeh Shahrbabaki

Thèse dirigée par **Dominique Houzet**

préparée au sein du **GIPSA-lab**
et de l'École Doctorale **Electronique, Electrotechnique, Automatique & Traitement du Signal**

Contribution of colour in guiding visual attention and in a computational model of visual saliency

Thèse soutenue publiquement le **16 Octobre, 2015**
devant le jury composé de:

Mr. Alain Tremeau
Université Jean Monnet, Saint-Etienne, France, Président
Mme. Christine Fernandez-Maloigne
Université de Poitiers, France, Rapporteur
M. Vincent Courboulay
Université de La Rochelle, France, Rapporteur
M. Denis Pellerin
Université Grenoble Alpes, France, Examinateur
Mme. Nathalie Guyader
, , Examinateur
M. Dominique Houzet
Université Grenoble Alpes, France, Directeur de thèse

Abstract

The studies conducted in this thesis focus on the role of colour in visual attention. We tried to understand the influence of colour information on the eye movements while observing videos, to incorporate colour information into a model of visual saliency. For this, we analysed different characteristics of eye movements of observers while freely watching videos in two conditions:colour and grayscale videos. We also have compared the main regions of regard of colour videos with those of grayscale. We observed that colour information influences only moderately, the eye movement characteristics such as the position of gaze and duration of fixations. However, we found that colour increases the number of the regions of interest in video stimuli. Moreover, this varies across time. Based on these observations, we proposed a method to compute colour saliency maps for videos. We have incorporated colour saliency maps in an existing model of saliency.

Résumé

Les études menées dans cette thèse portent sur le rôle de la couleur dans l'attention visuelle. Nous avons tenté de comprendre l'influence de l'information couleur dans les vidéos sur les mouvements oculaires, afin d'intégrer la couleur comme un attribut élémentaire dans un modèle de saillance visuelle. Pour cela, nous avons analysé différentes caractéristiques des mouvements oculaires d'observateurs regardant librement des vidéos dans deux conditions: couleur et niveaux de gris. Nous avons également comparé les régions principalement regardées dans des vidéos en couleur avec celles en niveaux de gris. Il est apparu que les informations de couleur modifient légèrement les caractéristiques de mouvement oculaire comme la position des fixations et la durée des fixations. Cependant, nous avons constaté que la couleur augmente le nombre de régions regardées. De plus, cette influence de la couleur s'accroît au cours du temps. En nous appuyant sur ces résultats expérimentaux nous avons proposé un modèle de saillance visuelle intégrant la couleur comme attribut. Ce modèle reprend le schéma de base d'un précédent modèle (sans couleur) développé au laboratoire et intègre des cartes de couleur. Nous avons proposé une méthode de calcul de ces cartes de couleur permettant de reproduire au mieux nos résultats expérimentaux.

Contents

List of Figures

List of Tables

Page

Publications

Some ideas and figures presented in the thesis have appeared previously in the following publications:

Refereed Journals

[Ham+15a] S. Hamel, N. Guyader, D. Pellerin, and D. Houzet. "Contribution of color in saliency model for videos". English. In: *Signal, Image and Video Processing* (2015), pp. 1–7. DOI: 10.1007/s11760-015-0765-5.

[Ham+15b] S. Hamel, D. Houzet, D. Pellerin, and N. Guyader. "Does color influence eye movements while exploring videos?" In: *Journal of Eye Movement Research* 8 (2015), pp. 1 –10 (see p. 77).

Conferences

[Ham+14] S. Hamel, N. Guyader, D. Pellerin, and D. Houzet. "Color information in a model of saliency". In: *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. 14785751. IEEE, 2014, pp. 226 –230.

1 Introduction

1.1 Context

Face to the the huge amount of the visual information that surrounds us, our visual system has limited biological and sensorial resources. However, human visual system (HVS) performs a rather efficient visual perception of our environment. Visual perception corresponds to the faculty of human visual system in interpreting and exploring the raw visual information, from acquisition of image by retina to the cortical processing. To deal with the huge amount of visual information, our visual system, is able to select the most pertinent information that achieves to retina from the whole stimuli located in the visual field. This ability is referred as visual attention.

Visual attention is correlated to the eye movements. A sequence of saccadic movements of eye and gazes brings a particular zone of the visual scene to the fovea, where the sensorial dispositions of eye are concentrated to perform a proper process of the gazed location. The selection of the location, that is to be gazed, involves two mechanism of selective attention: an unconscious, exogenous mechanism called also bottom-up attention and a concious endogenous mechanism also known as top-down attention. Bottom-up attention, which is stimulated by low-level features of the stimuli, allows the primarily processing of visual information rapidly and without involving all attentional resources. Top-down selection attention is concious, controlled, task dependent and involves most of the attentional and cognitive resources.

Modelling the mechanism of selective visual attention is one of the active research areas in the field of computer vision as well as cognitive science. Because of the very high complexity of the visual attention due to the interactions and dependency between bottom-up and top-down attention, modelling the mechanism of visual attention is less realistic with the existing technologies. Hence, the researchers are leaded to divide the models of attention into bottom-up attention models and top-down attention models. At the basis of the models of attention there are theories such as the Filter Model [Bro58] and the Feature Integration Theory (FIT) [TG80]. The latter is one of the most cited theories of attention, and divides the processes of attention into two stages: a pre-attentive and a focused one. According to the FIT, elementary visual features such as intensity, colour and orientation are processed in parallel at the pre-attentive stage, and subsequently combined to drive the focus of attention.

Later in 1985, based on *Feature Integration Theory*, Koch and Ullman [KU85] have developed one of the first computational models of attention which was inspired from the biology of human visual system. For the first time the term of **saliency map** appeared in this work. A saliency map has been defined as a representation of the visual scene, in which the most attractive regions are enhanced.

The Feature Integration Theory and the computational architecture of this theory proposed by Koch and Ullman [KU85] were the inspiration for many other computational models of attention, such as the model proposed by Itti and colleagues [IKN98], which is a reference model in the field of computational models of attention. These models, mostly, compute a saliency map of the visual stimuli according to their low level features, such as, colour, intensity, orientation, frequency, motion, etc. The contribution of these features to the deployment of attention has been examined on the synthetic stimuli [WH04]. colour besides other features has been found to deploy the attention when performing a visual search, for example finding a horizontal red bar between green vertical bars. Yet, the guiding power of colour features when exploring natural scenes is being debated.

1.2 Challenges

We are interested, in this thesis, on the role of colour in the visual attention, from eye movements to the computational models.

The first challenge is to investigate the guiding power of colour features in the video stimuli, using eye-tracking experiments and evaluation. The main question is whether colour influences, in the least, the eye movements and the focus of attention when freely watching the video stimuli. There are also several questions regarding whether influence of colour on the visual attention is correlated to the content of the stimuli. Does colour deploy attention in natural video stimuli for example landscapes? Does the contribution of colour in guiding attention vary between man-made scenes, such as urban roads and indoor scenes, and landscapes? What about person-present scenes? Faces were found to guide the visual attention rapidly and independent from the task. Does differ the allocation of attention on faces in colour stimuli from in grayscale stimuli?

The second challenge is to incorporate the findings from the experiments and evaluations into a luminance-based computational model of attention. In this thesis, the bottom-up model of attention proposed previously by Marat and colleagues [Mar+09] is improved. The original model computes the visual saliency maps of a video through static and dynamic pathways for grayscale stimuli. We tend to improve the performance of the model using colour information.

1.3 Objectives

Regarding the challenges described above, the objective of this thesis is twofold. On the one hand, we study and compare the behaviour of observers when viewing colour and grayscale video stimuli. On the other hand, we would like to include colour features to a computational model of saliency.

The first step is to perform eye-tracking experiments using video stimuli with various contents. The experiments would allow us to identify various factors related to the impact of colour features on the visual attention.

The second step involves computational modelling, to incorporate colour features into a biologically inspired saliency model and modulate these features based on the factors

identified through eye-tracking experiments. A colour-wise saliency model could be beneficial for computer and machine vision, cognitive robots, recognition and quality control devices.

1.4 Main contributions

This thesis focuses on the contribution of colour information into the eye movements from one side and into the performance of a saliency model from the other side. These two objectives are accomplished through following main contributions made in this thesis.

- ❖ We identify the impact of colour information on eye movements when observing video stimuli, in terms of position of gaze, observers congruency, number of the regions of interest, fixation duration and amplitude of saccade in global and as a function of time.
- ❖ We incorporate a colour saliency map to an existing luminance-based saliency model. We evaluate the performance of the model in comparison to the existing models in the literature

1.5 Thesis organization

Chapters 2 and 3 establish the state of the art in the field of visual attention and provide background materials related to colour perception and representation systems. Chapter 2 introduces the mechanism of visual attention. First the human visual system is briefly introduced. Then eye movements and their relation to visual attention are described. Afterwards the main computational models of attention are described. Chapter 3 presents main notions related to colour as a psycho-physical phenomena and it follows by presenting the colour representation systems and colour measurements. Chapter 4 presents two eye-tracking experiments and analyses the eye-movement of observers regarding the colour features. Chapter 5 introduces a colour-wise saliency model and its evaluation against the eye position data obtained from experiments of chapter 4. Finally, in chapter 6 we conclude and mention several perspectives.

2

Visual attention and its computational models

In this chapter, we present the background of computational models of visual attention, specifically the models that are inspired by human visual system.

To this goal, first the basic mechanisms involved in processing the visual information by human visual system are briefly presented. Then we introduce the concept of visual attention as well as the psychological models of attention. Afterwards, several computational models of visual attention are presented. At the end we describe the contribution of different low-level features in deployment of attention and also in the performance of computational models of visual attention. The latter allows us to introduce the research direction of this thesis.

2.1 Human visual system

Eye is the foremost neuronal organism of human visual system. The outer parts of eye, such as pupil, iris and cornea, are visible when eye leads are open 2.1. The pupil is the black-looking hole located in the center that allows light to enter the eye. The size of pupil is controlled by the coloured circular muscles of the iris that surround pupil. The cornea is the transparent external surface that is the first powerful lens of the optical system of eye and with the crystalline provides the image of visual scene at the back of eye.

Eye performs several functions as an sophisticated moving optical system that follows moving objects and focusses on the several targets in fractions of a second, and also as a neural structure that transform luminous signals to electrical or chemical signals and convey the visual information to the brain through optical nerve, Figure 2.2. The optical nerve has a limited visual information debit. Therefore, the retina has been developed to optimise the visual information that are to be transferred to brain.

The retina is a multi-layer neural membrane in the back of eye that resembles to a part of the brain, but located distant. The outer layer of retina contains the photoreceptor cells, rods and cones, that capture light. The captured visual information is pre-processed in the inner subsequent layers of retina, and then sent to the lateral geniculate nucleus (LT). The LT organises and sends this information to the primary visual cortex (V1), where further processing of the retinal image is carried out. The visual data is then sent to the next levels

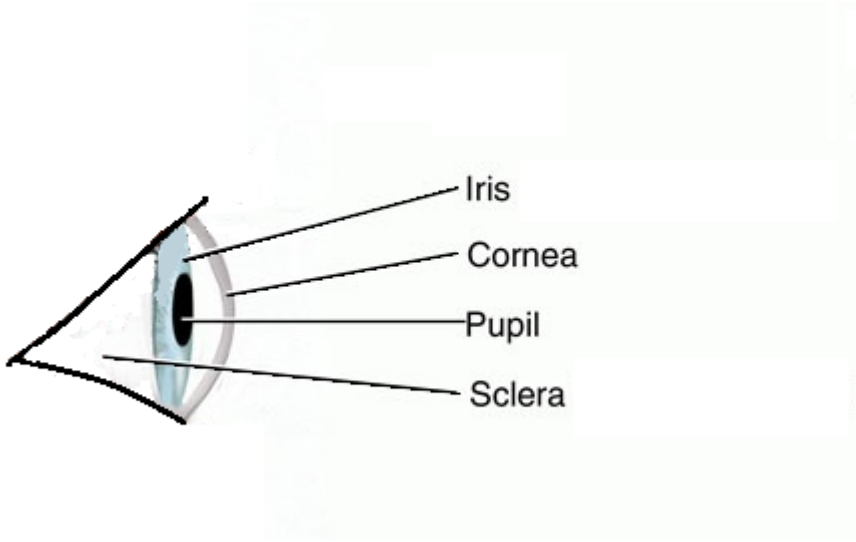


Figure 2.1: Outer parts of human eye

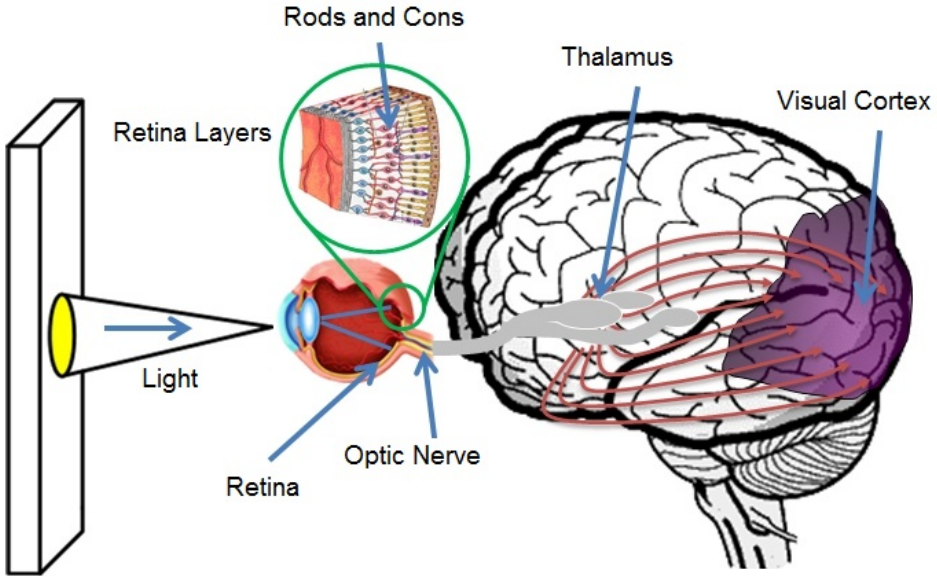


Figure 2.2: A schema of human visual system. Image from *INTECH*.

of HVS, through the ventral and dorsal streams. The ventral stream provides data for inferior temporal cortex (IT) which determines what an object is. The dorsal stream sends information to the visual association cortex, which determines where an object is.

2.2 Eye movements

Human visual receptor, the eyeball, has the advantage to be mobile. Whether eyes are open or close, the eyeball is in constant moving. One type of eye movement that happens voluntary or unconsciously is the tracking movements. The tracking movements could be studied in two categories: saccades and smooth pursuit. Both these movements are elicited to project the target on the fovea.

a) Saccades: A saccade is a rapid eye movement that changes the point of fixation. Saccades are the most rapid movements that human is able to execute ($900°/S$) with a short duration ranging from 20 to 200 ms. Saccades are considered as ballistic movements because the saccade-generating system is not able to modify the trajectory once a saccade is started. Saccades can be executed voluntary or unconsciously and also reflexively when ever the eyes are open. In 1967, Yarbus [Yar67] demonstrated that observing a static stimulus, such as an image, is consisted of a series of saccades and fixations, Figure 2.3. During a fixation the target is projected on the fovea to be processed in the highest spatial resolution.

Figure 2.3: The eye movements of a subject viewing a picture of Queen Nefertiti. The bust at the top is what the subject saw; the diagram on the bottom shows the subjects′ eye movements over a 2–minute viewing period. Image from [Yar67].

b) Smooth pursuit This type of eye movements are slow movements that occur when the eyes jointly fixate the same target while the target is moving relatively to retina. The smooth pursuit is a voluntary movement that surprisingly can not be executed in the absence of the relative movement of observer and target.

There are other eye movements such as vergence movements and vestibule-ocular movements. The first group aligns the fovea of each eye with the targets located at different distance from the observer, and the latter group stabilizes the eyes relative to the external world and compensate for head movements. There are also drift and micro-saccades that are less related to the overt attention [PJ01].

2.3 Visual attention

Although our visual system is intrinsically limited (limited debit of optical nerve), and our environment contains an infinitive quantity of visual information, we are able to perceive automatically the important changes in the visual scene. Our visual system has adopted the strategies to reduce the quantity of the information that must be processed and transported to the superior visual areas of brain [BB82].

First, the biological and physiological design of HVS allows a passive selection of the visual information:

– The photoreceptor cells are only sensitive to the light of visible spectrum.

– They perform a non-uniform sampling of visible light; in the center around fovea the spatial resolution is maximum.

– The visual cells are sensitive to spatial frequencies and the retina and cortical cells reduce the redundancy of visual information, and respond to the contrasts.

Second, a mechanism of attention is developed to deal with the unlimited amount of visual information surrounding us. The visual attention has been described as a "spotlight" that illuminates a limited zone of visual scene to be processed in details [Nei67]. Two stages of attention are covered in this example: a pre-attentive stage that is responsible for selecting the regions of interest, and an attentive stage that is devoted to the further processing of the selected regions [Nei67].

Mechanism of attention enables us to automatically attend to the regions of interest in visual scene and explore it by changing the focus of attention. Moving the focus of attention and the order in which a scene is explored could be performed in two ways: **overt** and **covert**. The overt attention occurs when the focus of attention is moved to a target by eye movements. The covert attention allows us to perceive something in periphery without involving eye movements, for instance when we perceive something from the corner of eye.

The overt attention is correlated to saccades that move the attention from one position to another, and fixations that allow the further processing of selected positions [Riz+87; HS95]. The overt attention might be quantified via eye movement analysis when viewing complex stimuli—static natural scenes [SD04; TV08; Bin10; HPGDG12], as well as dynamic scenes [CI06; DMB10; Mit+11; Cou+12].

Two categories of factors interfere in the mechanism of attention: **Bottom-up** factors and **Top-down** factors.

Bottom-up factors are associated to *"feature driven"* attention also known as bottom-up attention. Bottom-up attention refers to a transitory involuntary attention, also called exogenous, that is driven by the salient visual information.

Top-down factors are associated to *"task-dependent"* attention also known as top-down attention. Top-down attention is a voluntary mechanism of attention, also called endogenous, that is voluntary controlled to attend specific regions of the visual scene, according to the goal and a priori knowledge of the observer [BI11].

Although bottom-up attention is supposed to be a transitory mechanism that occurs at the onset of stimuli, several researches [PLN02; Pet+05], have demonstrated that this mechanism interferes also later in time when top-down mechanism of attention is activated.

2.4 Psychological theories of attention

During years several theories and models of attention were proposed to better understand how the selective attention happens and how the attentional resources are allocated. Here, we present theories and models which have been the basis for computational models of attention. More on psychological theories of attention can be found in the reviews of [Bro58; Bun90; KU00; Sch04; Fri06].

2.4.1 Filter Model

Broadbent's filter model of attention is one of the first theories of attention that describes an early selection of attention [Bro58]. The theory claims that because of the limited processing capacity of human, kind of selective filter limits the information at the very early stages of attention.

The selective filter, depending on the physical properties of the stimuli including colour, loudness, pitch and direction, allows for attended stimuli to pass through the filter for further

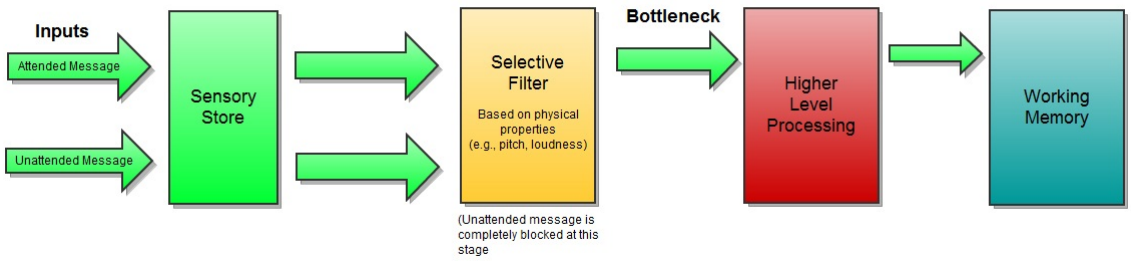


Figure 2.4: Broadbent's Filter Model. Image from *Wikipedia*.

processing, while unattended stimuli will be discarded, Figure 2.4. On the other hand to attend to a stimulus based on the goal or demanded task a voluntary mechanism of attention must interfere.

2.4.2 Feature Integration Theory FIT

FIT is one of the best known and most cited theories of visual attention that was proposed after filter model in 1980 [TG80] and was gradually improved to adopt with current findings [TG80].The theory divides the process of attention into two stages; a pre-attentive and a focused one. According to FIT "different features are registered early, automatically and in parallel across the visual field, while objects are identified separately and only at a later stage which requires the focus of attention" [TG80]. The theory is based on the promise that the process of attention in brain provides several *feature maps* according to the physical attributes of the stimulus. These feature maps are then combined to a *master map* that enhances the important regions of the visual scene. Figure 2.5 shows a schema of FIT.

2.4.3 Guided Search Model

Besides FIT, Guided Search model proposed by Wolf [WCF89], is one of the well known psychological models of attention. This model also has been evolved during years and several versions of its computer simulation are available [Wol94; WG97]. The model, in many aspects, is similar to FIT, but more detailed to be simulated by computer. Figure 2.6 depicts a schema of this model.

The main aspect that differentiate this model from FIT, is that instead of feature types (red, green, etc), the feature dimension (colour, orientation, etc) form each feature map. In addition the model associates a top-down map to each feature map.

2.5 Computational models of attention

Based on the psychological theories and models, several computational models of attention have been developed to improve computer vision systems.

Here we introduce, specifically, the models of attention that process the scenes to determine which regions deploy the attention involuntary in a pre-attentive stage of vision. Based on whether the model is inspired from human visual system or not, there are two main

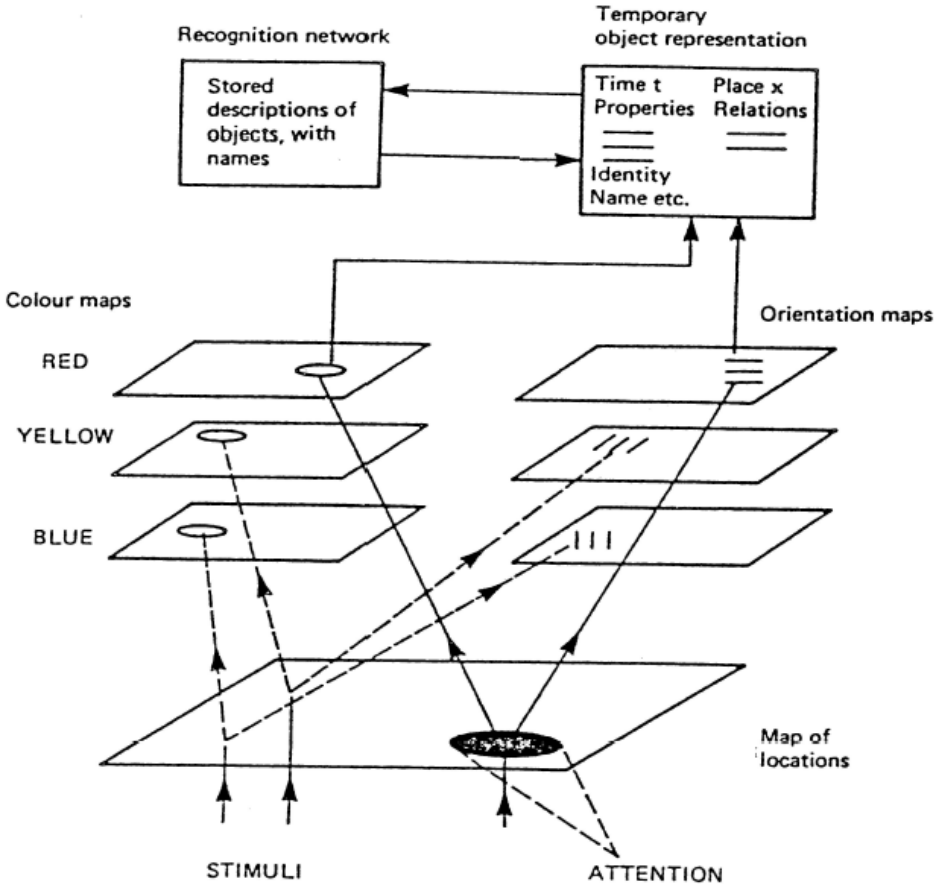


Figure 2.5: A schema of Feature Integration Theory. Image from [TG80].

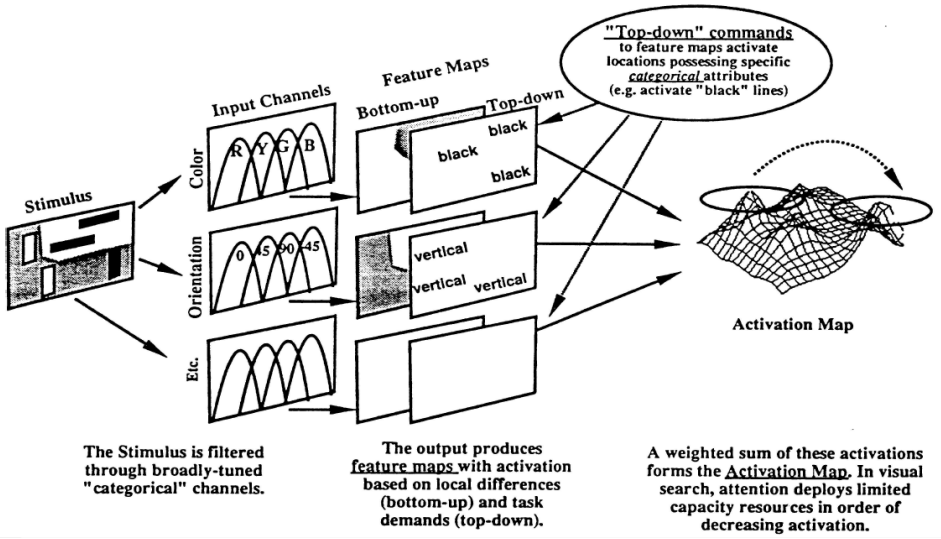


Figure 2.6: A schema of Guided Search model. Image from [WCF89].

categories of computational models of visual attention: biologically plausible and purely computational. In this work we are concerned with biologically plausible models. However, there is an overlap between these two category of models, as most of the well-known models could be considered as biologically plausible models with some pure computational units. We introduce, here, two of the pioneer biologically plausible computational models of attention and name several other models. More literature on computational models of visual attention can be found in following reviews [IK01; Fri05; BI10; Tso11].

2.5.1 Koch and Ullman

The model of selective attention proposed by Koch and Ullman [KU85], is one of the first biologically plausible models of attention. This model has been designed based on FIT [TG80]. Figure 2.7 depicts the scheme of this model. First a set of the elementary features, such as

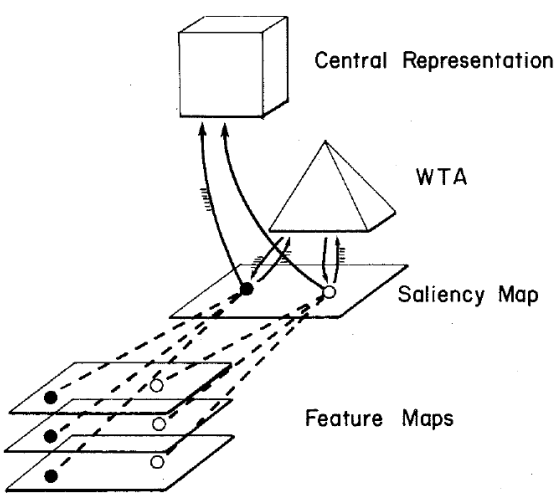


Figure 2.7: A Schema of the model proposed by Koch and Ullman. Image from [KU85].

colour, orientation, direction of motion, etc, are extracted from the input image. Then these features are processed in parallel and form different feature maps, respecting the topology of the input image. The lateral inhibition within feature maps, that simulate the photocells of retina, enhances the regions of the image that are different from their surroundings. The feature maps are fused to a saliency map. The notion of saliency map was first introduced by Koch and Ullman [KU85], according to whom, a saliency map "gives a biased view of the visual environment, emphasizing interesting or conspicuous locations in the visual field". The saliency map represents the salient features in the visual scene, but the order in witch the salient regions are focused is determined by a winner-takes-all (WTA) approach.

2.5.2 Milanese

The model proposed by Milanese [Mil+94] is one of the earliest models of visual attention, based on the Koch and Ullman model [KU85]. It uses filter operations to compute the feature maps. The model considers following features: red-green and blue-yellow colour opponents, 16 orientations, local curvature and intensity when colour information is absent. The local value of the feature maps are compared to their surrounds using center-surround differences. The resulting differences are gathered in conspicuity maps. The term of the

conspicuity was since used to refer to the feature-dependent saliency map. In another version of the model, the top-down information is added to the model [Mil+94] using *distributed associative memories* DMAs recognition algorithm. The DMAs is executed on the regions of interest obtained from the bottom-up model and provides a top-down saliency map that competes with bottom-up conspicuity maps. At the end, the result is a saliency map that include both bottom-up and top-down cues.

2.5.3 Itti

The model of Koch and Ullman have provided the architectural basis for many models that were proposed latter, such as the model proposed by Itti, Koch and Niebur in 1998 [IKN98]. This model has been improved during time [Itt02] and is undoubtedly the best known and most cited computational model of attention. Figure 2.8 shows the general architecture of the model that we present here briefly.

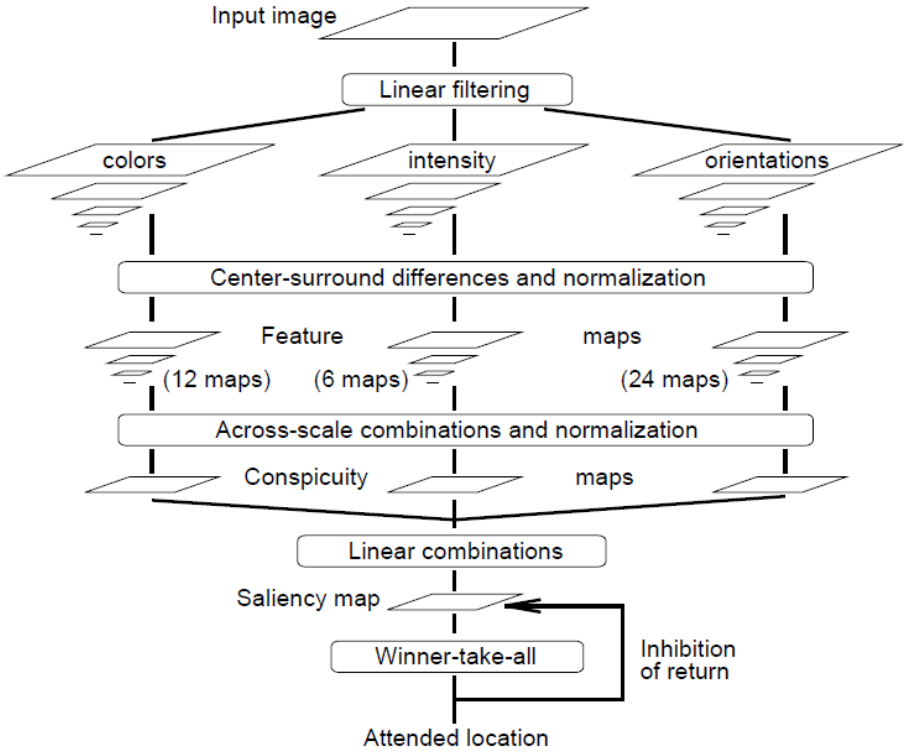


Figure 2.8: A schema of saliency model proposed by Itti and colleagues. Image from [IKN98]

As shown in the figure 2.8, three processing channels are extracted from the input rgb image:

– An intensity channel I where $I = \frac{r+g+b}{3}$. I is used to create a Gaussian pyramid of intensity images $I(\sigma)$, where $\sigma \in \{0...8\}$.

– A colour channel which includes four colour images: $R = r - \frac{g+b}{2}$ for red, $G = g - \frac{r+b}{2}$ for green, $B = b - \frac{r+g}{2}$ for blue and $Y = \frac{r+g}{2} - \frac{|r-g|}{2} - b$ for yellow.

– An orientation channel that is extracted from the intensity images. A pyramid of oriented gabor filters $O(\sigma,\theta)$, where $\theta \in \{0,45,90,135\}$ is used to excerpt four orientation images for each level of pyramid, $\sigma \in \{0...8\}$.

Then the feature maps associated to each feature channel are computed using center-surround differences between a "center" fine scale c and a "surround" coarser scale s as following:

– Intensity: $I(c,s) = |I(c) \ominus I(s)|$

– Colour: $RG(c,s) = |(R(c) - G(c)) \ominus (G(s) - R(s))|$ and $BY(c,s) = |(B(c) - Y(c)) \ominus (Y(s) - B(s))|$

– Orientation: $O(c,s,\theta) = |O(c,\theta) \ominus O(s,\theta)|$

In total 24 feature maps are computed: six for intensity, 12 for colour and 24 for orientation. For each channel the feature maps are normalized and linearly combined to create one unique saliency map for each channel, $N(I), N(C)$ and $N(O)$. The linear fusion of these three maps provide a saliency map that enhances the most attractive regions of the input image. At the end the order in witch the focus of attention (FOA) is moved on the salient regions of the input image is computed through a **WTA** approach. The **WTA** is combined with a mechanism of inhibition-of-return (IOR) to prevent the
acsFOA from returning immediately to an attended salient position. Figure 2.9 shows an example of the image sequences and the saliency map which indicates the location of the focus of attention.

2.6 Other models

There is wide variety of the models in literature. Many are based on similar approaches and differ only in details, for instance different number of features are considered.

The model of Chauvin [Cha03] is another biologically plausible model of visual attention that only processes the luminance information. Likewise the model of Itti [IKN98], this model extracts primary feature channels from 32 Gabor wavelets of 4 frequencies and 8 orientation. But it is different from the model of Itii in several points. Feature channels are normalized through a *divisive inhibition* method. Then different channels are filtered using butterfly filters to enhance aligned and collinear edges. Afterwards, the most important orientation of each frequency band are selected using iterative difference operations. Finally, the saliency map is computed by linear combination of different frequency maps.

The model proposed by Le Meur and colleagues [LMLCB07] is another biologically plausible model that show a high performance. Likewise A. Chauvin, the model uses Gabor wavelets and butterfly filters to extract feature maps. In addition the chrominance information are includes in the model. Moreover the model is extended to temporal dimension to process video stimuli as well.

The model of attention proposed by Courboulay and colleagues [CPDS12] is one of the real time models of attention. The model proposes a visual attention system that adapts its processing according to the saliency of each element of the dynamic scene. However, the model propose an original hierarchical and competitive approach to predict the position of gaze without the need of neither saliency map nor explicit inhibition of return mechanism, which are compute intensive.

In this thesis we study the bio-inspired model proposed by Marat and colleagues [Mar10]. The model is based on luminance information and computes the dynamic and static saliency maps for video stimuli. We present this model in details at section 5.1.

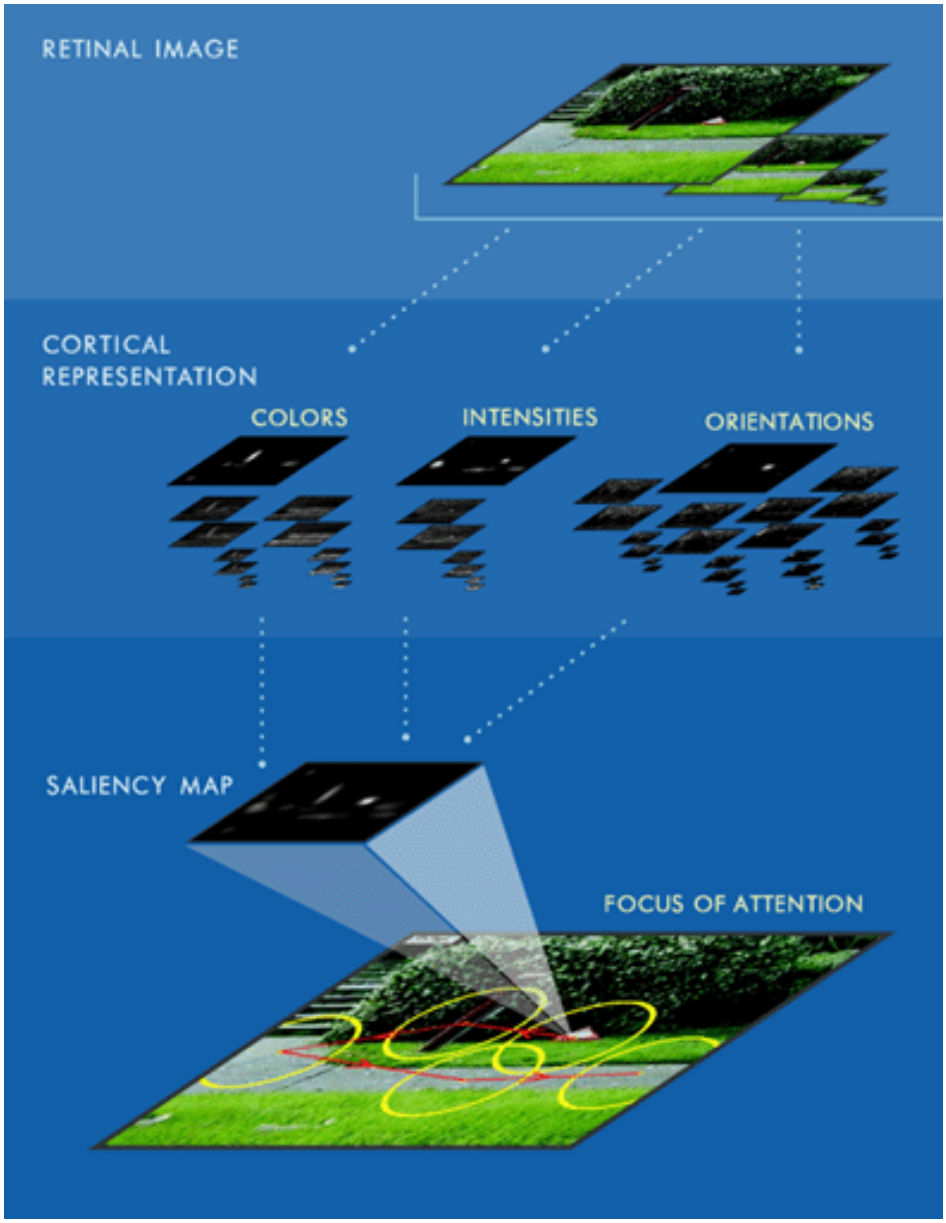


Figure 2.9: An example of the image sequences and the saliency map which indicates the location of the focus of attention. Image from *iLab, USC University of Southern California*.

More literature on computational models of visual attention can be found in [Fri05]. A comparison of the performance of several models is available in review of Borji and colleagues [BI10].

2.7 Conclusion

In this chapter, we presented briefly the human visual system and the mechanism of visual attention. Then we introduced the psychological models of attention before getting into the computational models of visual attention. The modelling of visual attention is a wide field. The current technologies do not allow to create a unique model of the whole aspects of the visual attention. Hence, each of the proposed model emphasizes one aspect of attention. We

discussed several of the bio-inspired computational models of bottom-up attention which join findings from human perception to computer vision systems.

In almost all the bottom-up models of visual attention, low level features like intensity, colour, spatial frequency are considered to determine the visual saliency of regions in static images, whereas motion and flicker are also considered in the case of dynamic scenes. Whereas, classical models of attention use colour as an elementary feature, some studies suggested that colour has little effect on location of regard [BT06a], [HPGDG12], [FHK08], which brings to question the necessity of the inclusion of colour features in bottom-up models of attention [DMB10].

In this thesis, we focus on the role of colour in guiding visual attention as well as in the computational model of attention. To achieve this goal, it seems essential to understand how colour is perceived as well as how colour is represented. Therefore, in the following chapter we describe a trajectory of colour from perception to the digital representation.

3

Colour from psycho-physical phenomena to numerical representation

Our work is focused on quantifying the contribution of colour information in human visual attention. We study from one side the influence of colour information on the eye movements and from the other side we introduce colour information in a computational model of attention. Before getting to the core of our work, it seems essential to understand how colour is formed, perceived and represented. In this chapter we describe, briefly, a trajectory of colour from psycho-physical phenomena to digital representation. First we introduce the notion of colour as a psycho-physical phenomena. Then, we address the vast topic of colour representation systems in a non-exhaustive way. Afterwards, the objective methods of colour measurements are presented followed by a numerical example of colour measurements for an LCD display. At the end we discuss colour to grayscale conversion methods.

3.1 Colour

Colour is a complex aspect of the appearance of an object that is related to divers fields such as optical physics, physiology and psychology. The human perception of colour is the interpretation of the colour signals transmitted from retina, by visual cortex, Figure 3.1. The colour of an object results from the interaction of three elements:

- Light
- Object/material
- Observer

3.1.1 Light

In 1666 Sir Isaac Newton realized a series of light decomposition experiments using a prism. He observed that light is decomposed to a multicolour band similar to a rainbow. This decomposition showed that white light is resulted from the mixture of a high number

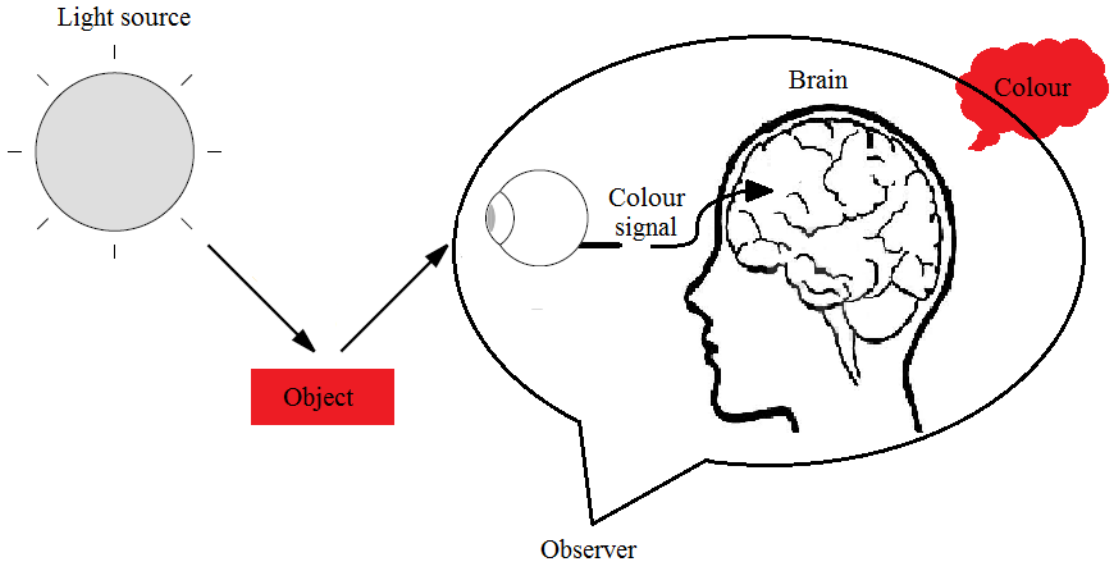


Figure 3.1: Three elements that form the perception of colour of an object by human observer. Image from [Van00].

of coloured radiations and ended the previous believes considering white light as a non-decomposable element.

Colour can be defined as a group of electromagnetic waves resulted from the propagation of the photons. The electromagnetic waves are identified by their wavelength, λ, measured in m. Human eye perceives only a small band of electromagnetic vibrations spectrum with wavelengths from 380 nm to 780 nm. This narrow band is called the *visible spectrum*, Figure 3.2. *Visible spectrum* includes colours from violet, 380 – 450 nm approximately, to red, 630 – 780 nm.

Light source For colorimetric purposes the light is produced by warming a material to its incandescent temperature or by exciting the atoms and molecules using an electrical spike or discharge. In order to identify colour of a light, it is compared to the incident light of a *black body*. A black body is an idealized source that absorb all incident electromagnetic radiations. The colour of a light source can be expressed as a temperature in Kelvin. The *temperature of the colour* is the equivalent temperature of the *black body* that has the most similar visual aspect to the light source.

The principal source of light is daylight. We observe the natural colour of the objects under daylight condition. Daylight is constituted of sunlight and the light diffused from the atmosphere. The solar spectre is extended from 200 to 4000 nm that is the equivalent of a black body in 5800 Kelvin. According to the influencing factors such as the latitude, the season, the weather conditions and time, daylight might produce colour temperatures from 4000 Kelvin to 6000 Kelvin. Hence, the daylight source had to be normalized to be reproducible and consistent. There exist several normalized daylight sources, called *illuminants*. An international organization called *Commission International de L'Eclairage, (CIE)* is the responsible for establishing the characteristics of standard illuminants as well as other normalisations and recommendations required in the science of colour measurements.

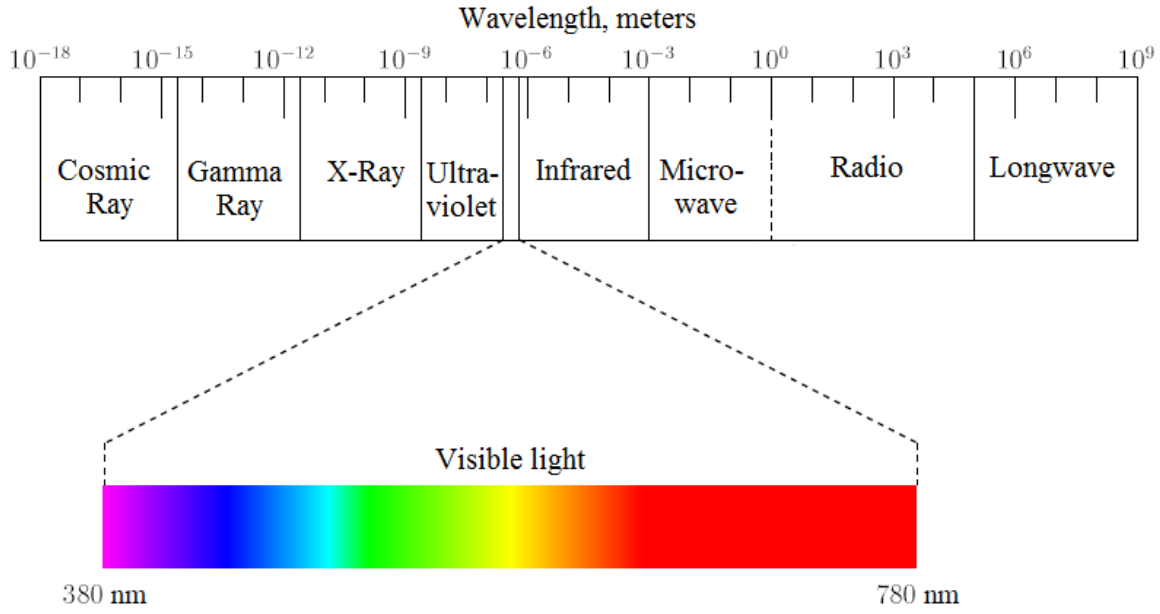


Figure 3.2: The visible spectrum. Image from [Van00].

An illuminant is characterized by its relative spectral energy $S(\lambda)$. Most of the illuminants are normalized to 1 or 100 at $\lambda = 560$ nm. Table 3.1 represents the characteristics of some well-known standard illuminants.

Table 3.1: Example of some standard illuminants that are employed in different applications

Illuminant	Equivalent	Temperatur in Kelvin
A	Tingesten lamp	2856
B	Sun	4870
C	averaged daylight	6770
D	different daylights D50 D55 D65	
D65	daylight at 6500 kelvin	6500
E	Non physical equi-energetic source	
F	from F1 to F12 for different sources	
F2	florescent lamp	

3.1.2 Object

Eye perceives the objects and materials according to the way that light is modified and reflected from the surface of the object. We can see objects that either reflect the radio-magnetic waves with wavelengths situated in the *visible spectrum*, called reflective objects, or emit light, called self-luminous objects. A reflective object absorbs part of incident light and reflects other part. The reflected part might be captured by eye to form an image of the object on the retina. Colour of the object is associated to the electromagnetic waves and their spectral distribution. The selective absorption of certain wavelength by an object determines its colour. For example a red surface absorbs blue, green and yellow lights and reflects the

red light. The colour of a reflective object might be represented by its spectral reflectance curve, which expresses the reflected part of the incident energy. Figure 3.3 shows the spectral reflectance of different colour objects.

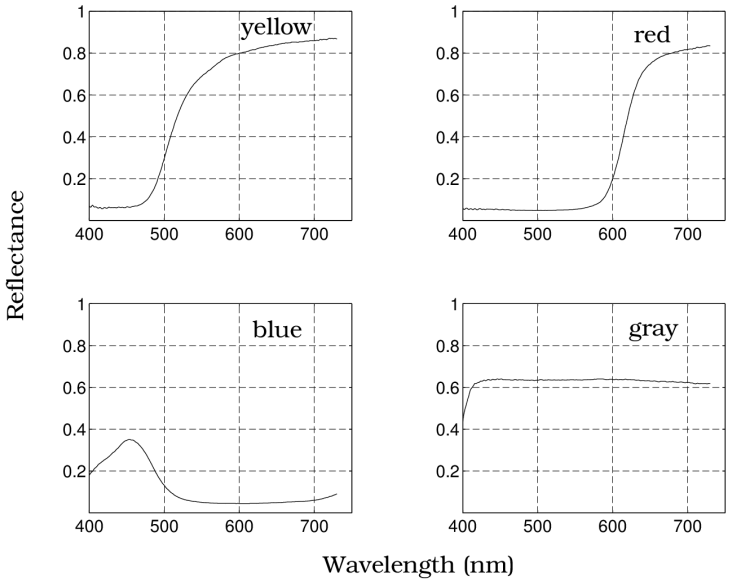


Figure 3.3: Spectral reflectance for yellow, red, blue and gray objects. Image from *Stanford university, Foundations of Vision, Brian A. Wandell*

3.1.3 Observer

The visual perception of an object is the result of the brain interpretation of the signals transmitted from retina. The retina contains photoreceptor cells that capture light and transform it to the signals which are interpretable by visual cortex. There are two categories of photoreceptors, *cones* and *rods*. The cone contribute in photoptic vision under well-lit conditions and are sensitive to colour. The rods are responsible for scotopic vision under low-light conditions and are sensitive to lightness variations, but they are not sensitive to colour. There are approximately 8 million cones and 120 million rods in retina. The cones are placed randomly near to fovea, while the rods are grouped and concentrated in the outer edges of retina resulting a high sensibility for lateral vision, Figure 3.4.

The cones, according to their peak respond to different wavelengths, are divided to three types. *Red* cones or long-wavelength*(L-cones)* respond the most to the light of long wavelength. They have a maximum sensibility to reddish light. *Green* cones or middle-wavelength*M-cones* are the photoreceptors that respond the most to the green light of middle wavelength, and the *blue* cones or short-wavelength *S-cones* are the ones with the peak response to the bluish lights of short wavelength, Figure 3.5.

Information about wavelength and intensity is confounded at the output of each individual cone. In the retinal ganglion cells, the output signals from different cones are compared conveyed to the brain through three channels: a luminance channel and two colour-opponent channels. In luminance channel the signals from L- and M-cones are added *(L+M channel)* to reproduce the intensity of an object. In red-green colour-opponent channel the signals from L- and M-cones are subtracted from each other *(L-M channel)* to obtain the red-green component of an object. And in yellow-blue colour-opponent channel the signal

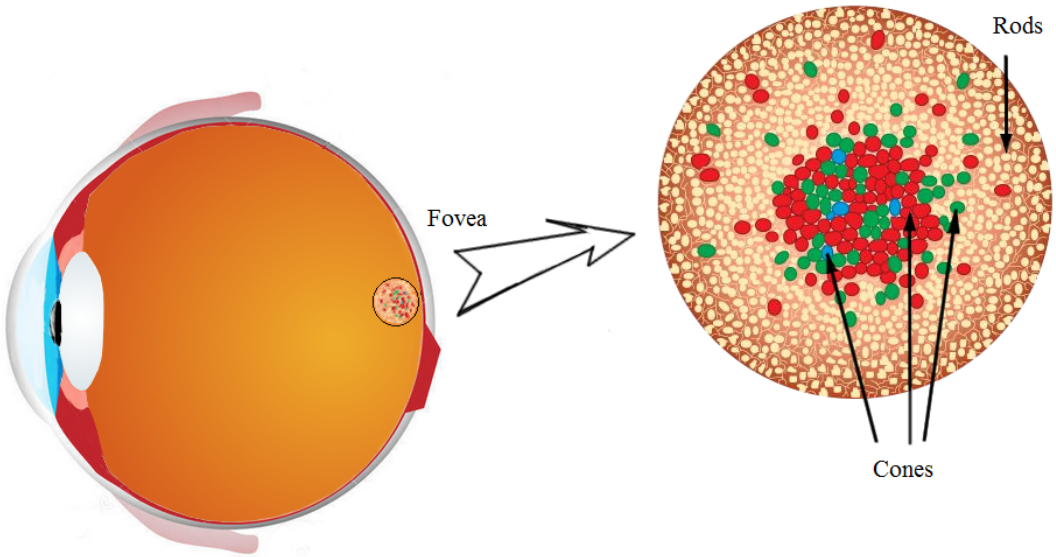


Figure 3.4: There are four types of photoreceptor cells in the human retina. Short-wavelength cones (blue), Medium-wavelength cones (green), Long-wavelength cones (red) and rods.

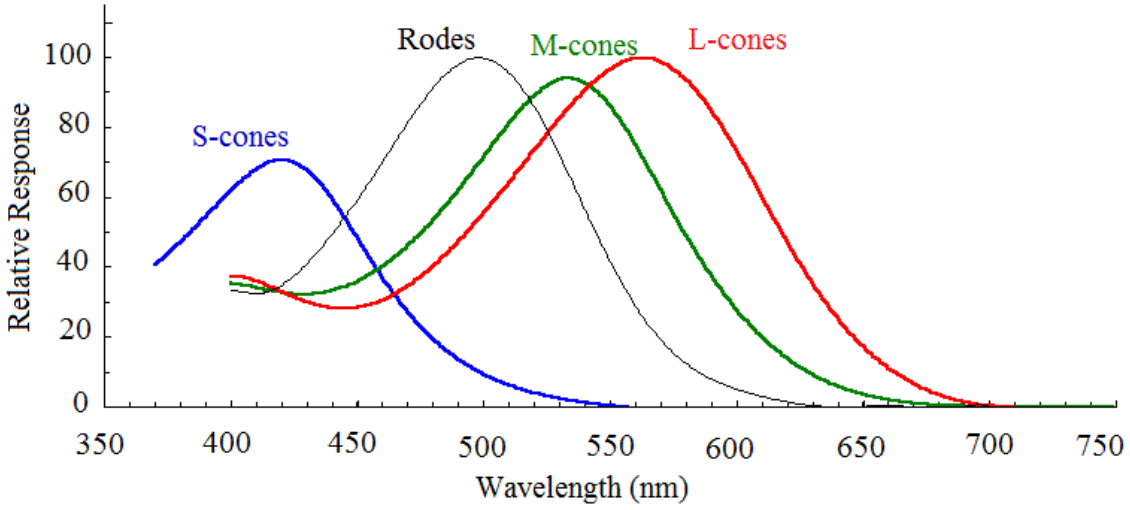


Figure 3.5: Spectral sensitivity of rods and cones. Note that rod curve is not to scale.

from sum of the signals from L- and M-cones is subtracted from S-cones *(L+M-S channel)* to compute the yellow-blue component of an object. These three channels are independent and are transmitted in physiologically distinct pathways; luminance information stimulates cells in the magno-cellular layers of *LGN*, red-green information stimulates cells in parvo-cellular layers and yellow-blue information stimulates cells in koniocellular layers. Although the functionality of primary stages of the perception of colour has been studied in details, cortical

stage of colour perception are less well studied. The V4 area of brain is considered as colour centre of brain. However, like most of the visual attributes our perception of colour depends on the activities of several cortical areas [Geg03].

3.2 Colour measurement

Colour is an interpretation of signals from retina by cortex. This interpretation is different from one to another, resulting a subjective definition of colour. But, many applications need an objective measure of colour of a stimulus. The science of colour measurements, colorimetry, have been developed to meet these needs.

3.2.1 Trichromacy of colour mixtures

According to the trichromacy of colour mixtures or *additive colour synthesis* any given colour can be reproduced by mixing various proportions of the three primary monochrome *red*, *green and blue* lights. The *additive colour synthesis* is used in colour displays like television receivers and computer-controlled monitors to reproduce colours. But, in some instruments such as printers, *subtractive colour synthesis* method is employed. The *subtractive colour synthesis* is based on the absorbing characteristics of materials as the colour of an object is the result of the parts of the visible spectrum of the light that are not absorbed. In *subtractive synthesis* the primary colours are *Magenta*, *Cyan*, and *Yellow*, Figure 3.6.

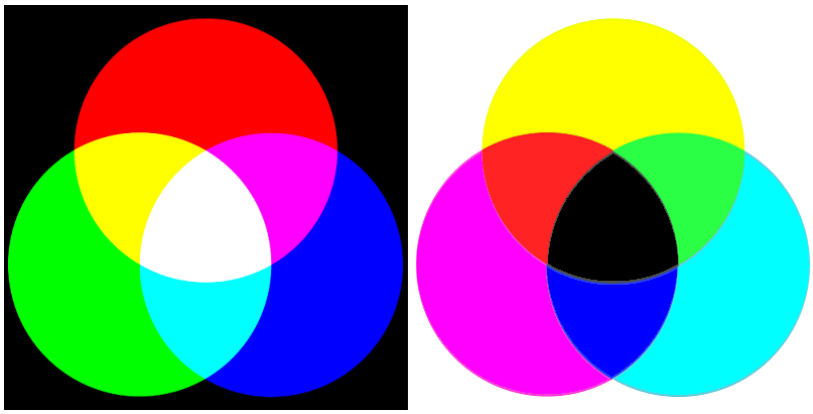


Figure 3.6: Additive (left) and subtractive (right) colour synthesis.

Colour mixtures were studied by Maxwell. In 1855 he presented the first three colour projections, a description of which appears in [Mac70]. In 1857 he developed the theory of additive colour mixtures for the three colour primaries from Young-Helmholtz theory[1]. He specified 13 different colours in form of a triangular diagram in witch the three primary colours (red, green, blue) are located in the three angles of an equilateral triangle and their mixture in the center produces the white point, Figure 3.7. Maxwell's studies built the basis for the colour measurement methods.

[1] Young-Helmholtz theory has been developed in 1807. Helmholtz established the theory of trichromacy of colour vision due to three types of photoreceptors in retina. Later in 1860s, he detailed the characteristics of these photoreceptors, today known as cones.

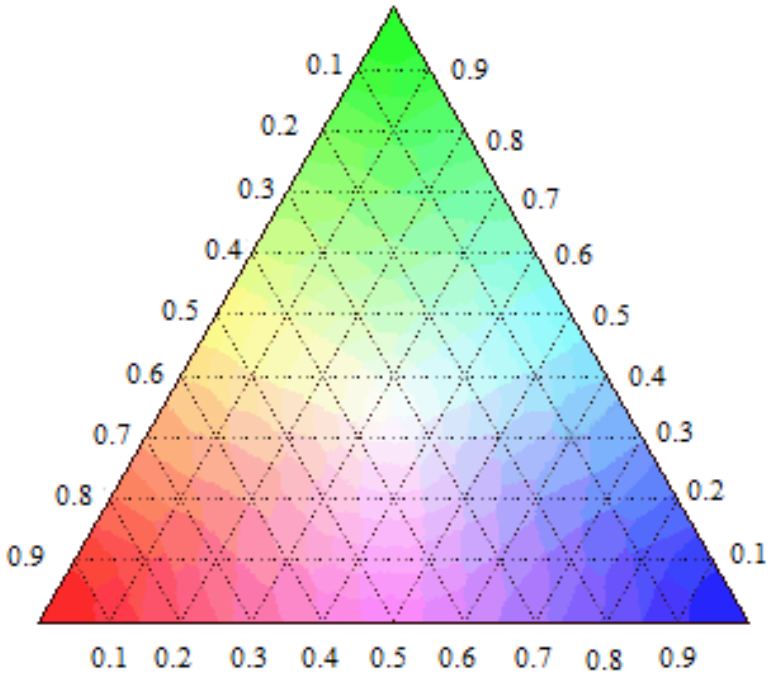


Figure 3.7: Maxwell's triangle. 13 different colours are specified on this triangular diagram. The three primary colours (red, green, blue) are located in the three angles and their mixture in the center produces the white point

3.2.2 Grassman's law

In colorimetry it is supposed that there is a relation between the three primary colour stimuli and the perceived sensation that arises from them. The characteristics of this relation is based on Grassman's law that allow treatment of colour mixtures as a linear system. Two of the more important laws can be explained as follow:

- Any colour stimulus C can be reproduced by a linear combination of three primary colours P_1, P_2 and P_3.

$$C \equiv p_1\mathbf{P_1} + p_2\mathbf{P_2} + p_3\mathbf{P_3} \tag{3.2.1}$$

  Here the $\equiv$ sign indicates a colour matching. The colour in one side appears the same as the colour on the other side.

- The luminance of a colour mixture is equal to the sum of the luminance of each colour

$$L = L_1 + L_2 + L_3 \tag{3.2.2}$$

In a more general way, the properties of the relation between colour stimuli A, B, C and D respects equivalence properties as follows [TFMB04]:

– Reflexive: $A \equiv A$

- Symmetric: if $A \equiv B$ then $B \equiv A$
- Transitive: $A \equiv B$ and $B \equiv C$ then $A \equiv C$
- Additive:
  - $A \equiv B$ then $A + C \equiv B + C$
  - $A \equiv B$ then $\equiv$
- Multiplicative: $A \equiv B$ then $k.A \equiv k.B$
- Simplification: $A + C \equiv B + C$ then $A \equiv B$

These rules have been experimentally tested through *colour matching experiments* and have been confirmed for a wide photoptic range, but they are not valid for the weak luminosities close to mesopic and the very high luminosities [TFMB04].

3.2.3 Colour matching experiments

At the foundation of trichromacy of vision and colour mixtures lies a series of idealized colour matching experiments. These experiments were set to determine what mixture of three primary colours would appear like a given spectral colour. Subjects were asked, first, to match a white point that has equal energy in all parts of visible spectrum. Then they adjusted the strengths of the three primaries to match each spectral colour, Figure 3.8.

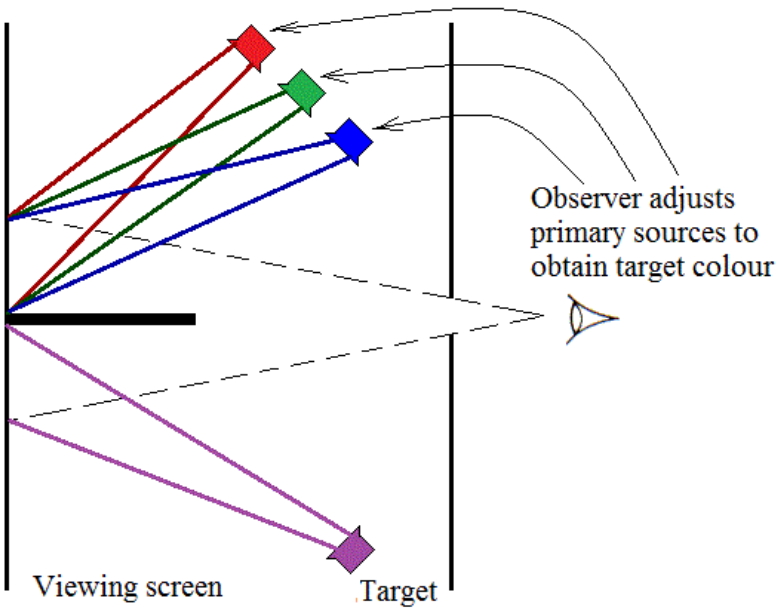


Figure 3.8: Colour matching experiment. Each frequency of target light is projected on the viewing screen on one side of the black partition. The observer adjusts each of the red, green and blue lights to reproduce the same colour as the target.

In 1931 CIE defined the colour matching data for a *standard observer*, based on the data obtained from two separate colour matching experiments: Wright in 1928 and Guild in 1931. Figure 3.9 shows the chromaticity coordinates versus wavelength of the spectral colours for

seven observers in Guild experiment and figure 3.10 shows the *standard observer* chromaticity functions of CIE 1931.

In Guild experiment the visual field was rectangular giving 2 degrees of visual angle in diameter. Guild had obtained data from 7 subjects with spectral primaries at 630 nm, 543 nm, 460 nm and *National Physical Laboratory (NPL)* reference white , which was most similar to the new CIE Standard Illuminant B in colour temperature and spectral power distribution [Sha03]. In Wright experiment the visual field was also rectangular giving 2 degrees of visual angle. Wright had obtained data on 10 observers and the monochrome primaries were at 650 nm, 530 nm, 460 nm and were normalized at a different white point. In 1965 CIE introduced chromaticity functions obtained for a 10 degree of visual angle diameter.

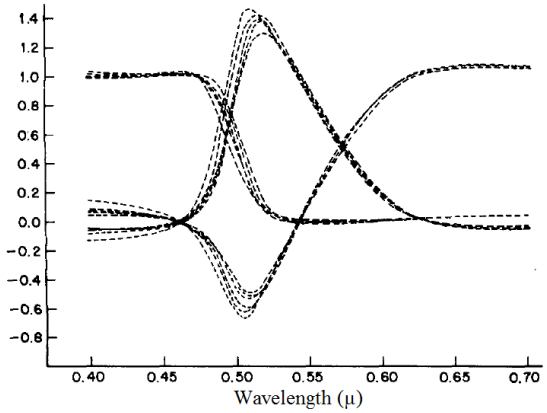


Figure 3.9: The chromaticity coordinates versus wavelength of the spectral colours for seven observers using Guild's trichromatic colorimeter primaries, and the NPL reference white. Image from [LRT77].

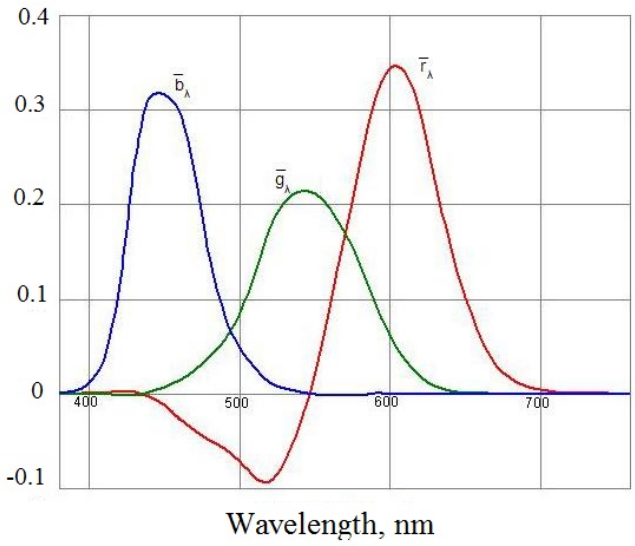


Figure 3.10: Colour matching functions for standard observer.

A standard relative luminous efficiency function (luminosity function, $V(\lambda)$) had also been adopted by CIE for photopic (normal day-light vision) and scotopic (night-light vision)

conditions, Figure 3.11. The luminosity function represents the results derived from several different photometric methods of brightness matching for spectral energy sources and describes the average spectral sensitivity of human visual perception of brightness.

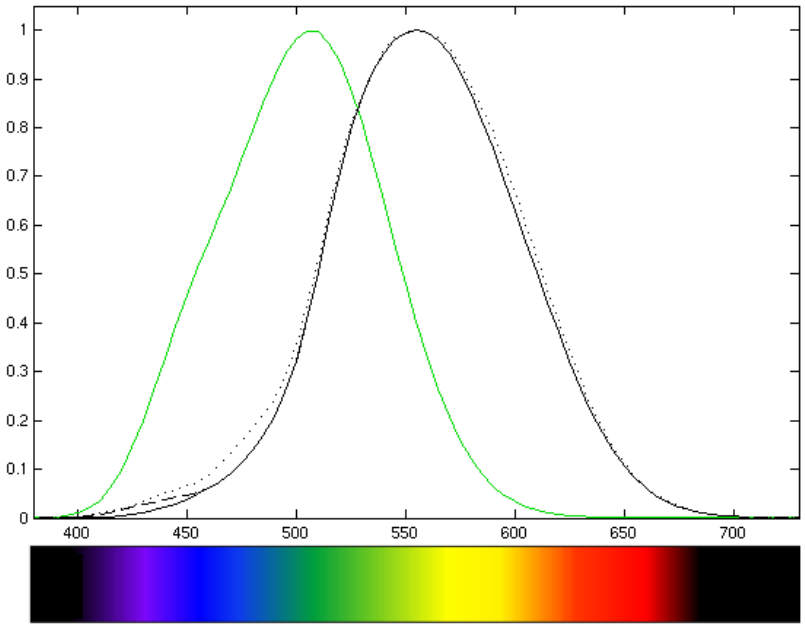


Figure 3.11: Photopic (black) and scotopic (green) luminosity functions Image from *Photometry (optics), Wikipedia*.

3.2.4 Colour transformations

As we saw for Write and Guild colour mixture experiments, the choice of primaries is not unique. However, a transformation could be found between any two arbitrary sets of primaries. This problem has been solved by a number of researchers for certain special cases. Wintringham [Win51] has treated the problem in a very general form in which the two reference whites are not identical. The result can be expressed in terms of a 3×3 matrix transformation, Equations (3.2.3) to (3.2.5).

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = P \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.2.3}$$

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} \tag{3.2.4}$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = P^{-1} \times \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \tag{3.2.5}$$

On the other words any set of primaries can be obtained from the linear mixture of another set of primaries.

3.3 Colour representation systems

There are numerous models established to represent the colour information and quantify the colour sensation. These models vary basically according to how sensation of colour is considered: a physical phenomena, a physiological, a psychological phenomena or a combination of these phenomena which might provide a more precise model.

Colour matching experiments show that any colour could be reproduced using a linear combination of three primaries using a colour matching function. Based on this approach several colour representation systems have been proposed that introduce a three-dimensional space for colour representation with three primary colours as the three unit vectors. Theoretically as much as there are primary systems there are colour representation systems. Any such system can be transformed to another using a transformation matrix **P** as it was explained in 3.2.4. Many colour representation systems were also established without introducing new primaries.

Because of the diversity colour representation systems and their applications, discussing them in details is beyond of the scope of this section. Therefore, we present briefly the different categories of colour representation systems and we introduce the primary-based systems that are essential for proceeding the colorimetric analysis.

3.3.1 Different categories of colour representation systems

The colour representation systems could be classified, according to their characteristics, into four categories [Van00],[TFMB04].

a) Primary-based system The *colour matching experiment* leads to one of the most common categories of colour representation systems. Any colour of visible spectra can be deduced from linear combination of a set of primary sources , R, G and B. Because of the needs of colour measurements in this thesis, we discuss this category of colour representation systems in details at section 3.3.2.

b) Luminance-chrominance systems This category of colour representation systems have been developed to dissociate the luminance and the chrominance information. In most of such systems there is a luminance component and two chrominance components such as in YC_bC_r. The family of YC_bC_r systems have been developed to ensure a compatibility between colour and black and white TV receivers. The YC_bC_r components could be computed from R G B coordinates using a linear transformation. Different television standards uses different coefficients for the transformation which result to several systems of YC_bC_r type. For example NTSC standard has employed Y I Q system, while PAL standard used Y U V system.

Another type of systems that could be considered in Luminance-chrominance category is antagonist systems. This family of colour representation systems have been inspired from the theory of Young [You02], according to which the perceived colour information by human visual system are transmitted to brain as three signals, one corresponding to achromatic information and two signals for two colour opponents red-green and blue-yellow. AC_1C_2 is one the systems of this type, proposed by Faugeras [Fau79]. $AC_{r1}C_{r2}$ is another such system proposed by Krauskopf and colleagues [KWH82].

c) Perceptual systems Human observers usually describe the sensation of colour by the hue or tonality, the saturation or degree of the purity and the luminosity or brightness. Several colour representation system have been developed to quantify these features of colour. One of the main attributes of these colour systems is to be able to describe the perceptible colour differences of similar colours. The Hue-Saturation-Value ($H\ S\ V$) and $L^*a^*b^*$ are two example of the systems of this category.

d) Independent axis systems In most of colour representation systems the three components are less or more correlated. This correlation prevent to process one component independent from the other components. To deal with this problem several colour systems have been proposed in which the components are independent. This type of systems is usually called $X_1X_2X_3$. One of the pioneer methods that allow to decorrelate the components is the model based on Karhunen-Loeve transformation [HW71].

A representation system might belong to more than one of the presented categories. For example *XYZ CIE* system is a primary-based system that also dissociate luminance information, component Y, and chrominance information, component X and Z. A very complete presentation of colour representation systems could be found in [Van00] and [TFMB04].

3.3.2 Primary-based systems

As we saw in *colour matching experiments*, section 3.2.3, any colour C can be reproduced from a weighted mixture of three primaries R, G and B. Colour C is, therefore, presented by a point in a three-dimensional space with origin O and three primaries **R**, **G** and **B** as unit directing vectors. In this space the colour equivalence properties are described by vectorial equalities. Such colour system respects Grassman colour mixture laws presented in section 3.2.2.

As much as there are R, G and B primaries and the white point there are R G B primary systems. Any primary system can be transformed to another using transformation matrix P, Equation (3.2.4). Several systems of primaries has been defined regarding the chosen primaries and white point and are commonly used in different applications [TFMB04]. Two of the most common primary systems are: The RGB system of CIE with equal energy white point (illuminant E), NTSC system with illuminant C. Another primary-based system, which is the CIE colorimetric reference, is XYZ.

a)RGB system of CIE The RGB system of CIE was first defined in 1931 based on Wright and Guild colour mixture experiments. The chosen reference white point was equal-energy white point, illuminant E. In this system the three primaries R, G and B are associated to three directing normal vectors **R**, **G** and **B** and form a vectorial space, Figure 3.12. Each colour C, in this space, is presented by a point C and a colour vector **OC** where O is the origin point of this vectorial space. According to *additive colour synthesis* the colour C is obtained from equation (3.3.1).

$$C = R.\mathbf{R} + G.\mathbf{G} + B.\mathbf{B} \tag{3.3.1}$$

The coordinates of vector **OC** are the tristimulus values R, G and B. The negative coordinates represent the colour stimuli that could not be reproduced by an additive synthesis. The colour stimuli with positive tristimulus values form the colour cube, Figure 3.12. The origin point O represents the black point (R, G and B = 0), while the reference white is

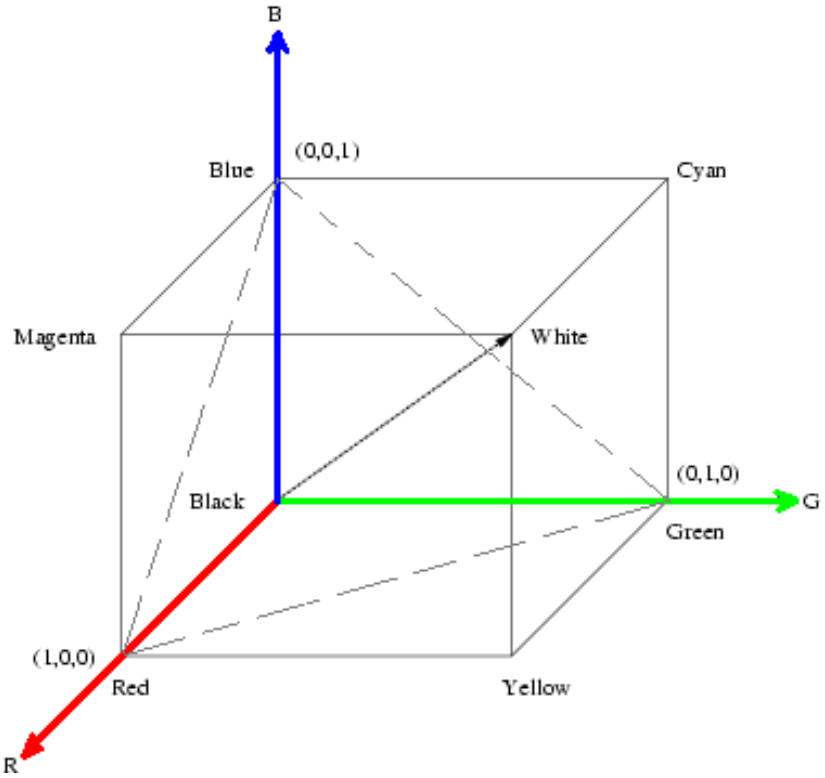


Figure 3.12: R G B cube, The origin point O represents the black point (R, G and B = 0), while the reference white is the point with all coordinates equal to one (R, G and B = 1).

the point with all coordinates equal to one (R, G and B = 1). The diagonal axis **OW** is the achromatic axis, and the points on this axis represent the shades of gray.

The tristimulus values R, G and B of a colour stimulus depend on its luminosity, therefore two different colour stimuli might have same tristimulus values . To remove luminosity information, colorimetrists use the normalized components called *chromaticity coordinates* expressed in equation (3.3.2).

$$\begin{aligned} r &= \frac{R}{R+G+B} \\ g &= \frac{G}{R+G+B} \\ b &= \frac{B}{R+G+B} \end{aligned} \tag{3.3.2}$$

Equation (3.3.2) can be considered as a projection of the point C to the plan $R+G+B=1$ with achromatic axis as normal vector. The intersections of this plan and colour cube form *Maxwell's triangle* also known as *colour triangle*, with R, G and B points as the vertices. In figure 3.12 Maxwell's triangle is specified with dashed gray lines.

In spectral domain the normalization process of equation (3.3.2) provides the *spectral*

chromaticity coordinates of CIE, $r(\lambda)$, $g(\lambda)$ and $b(\lambda)$, obtained from equation (3.3.3).

$$\begin{aligned} r(\lambda) &= \frac{R(\lambda)}{R(\lambda)+G(\lambda)+B(\lambda)} \\ g(\lambda) &= \frac{G(\lambda)}{R(\lambda)+G(\lambda)+B(\lambda)} \\ b(\lambda) &= \frac{B(\lambda)}{R(\lambda)+G(\lambda)+B(\lambda)} \end{aligned} \tag{3.3.3}$$

The chromaticity coordinates of the spectral colours for the Standard Observer are shown in figure 3.13. The colour space formed by chromaticity coordinates is called *normalized*

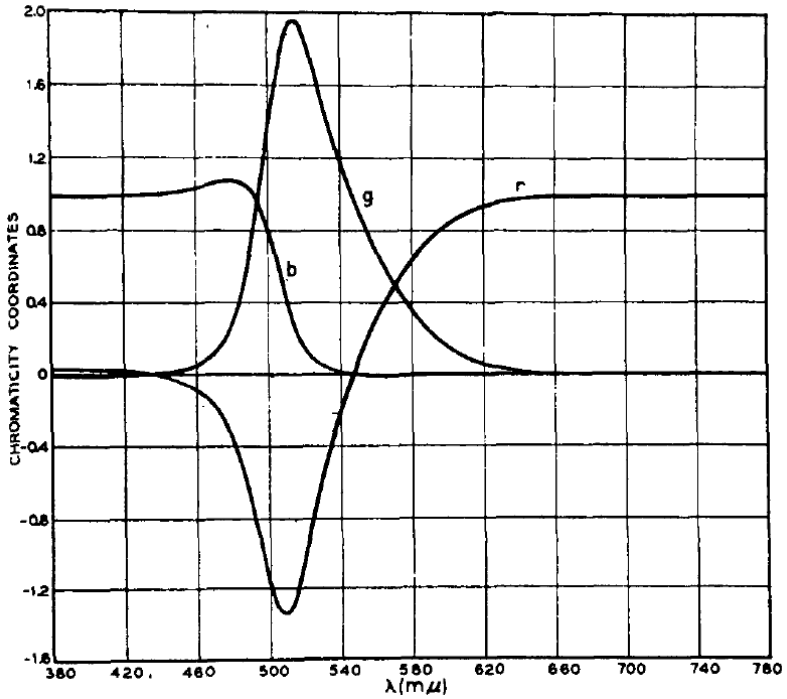


Figure 3.13: Chromaticity coordinates of spectrum colours for the Standard Observer. Primaries: 700.0 nm, 546.1 nm and 435.8 nm (NPL primaries). Reference white: equal-energy white. Image from [LRT77].

RGB space or *rgb space*, where, $r + b + g = 1$. Only two components are necessary to represent a given colour stimuli. A graphical representation of this information is obtained from the chromaticity diagram of r and g chromaticity coordinates, Figure 3.14.

The spectral colour plot on the elongated horseshoe shapes a curve called the spectral locus. The straight line connecting the two extremities of the spectral locus is called the line of purples. The spectral locus extends outside *Maxwell's triangle*. In section 3.2.3, we presented the colour matching functions for standard observer, $\overline{r}(\lambda)$, $\overline{g}(\lambda)$ and $\overline{b}(\lambda)$. As shown in figure 3.10, $\overline{r}(\lambda)$ is negative in part of visible spectrum. Consequently one of the primaries must be added to some of the spectral colours to carry out additive synthesis, which is the equivalent to moving one of the terms from the right side of equation (3.3.1) to the left. In addition in *R G B* the luminance is not presented as an independent component, and is obtained from the sum of luminosity coefficient of each chromaticity coordinate. To cope

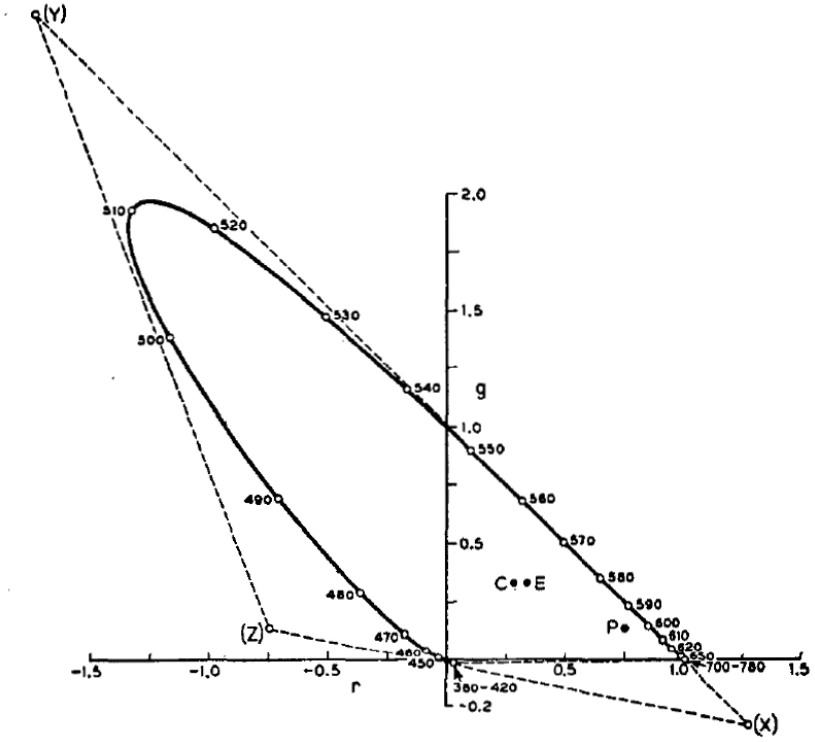


Figure 3.14: rg chromaticity diagram. Image from [LRT77].

with these drawbacks of *R G B* systems, among the others a new set of non-physical primaries called *X Y Z* were proposed.

b) Other *RGB* systems We discussed the *R G B* system of CIE which is based on equal-energy illuminant E. Several another R G B systems are adopted in different applications. For example colour televisions were one of the first commercialized products in which the trichromacy of colour mixtures was exploited. For instance, in Cathode Ray Tube (CRT) receivers the three lights could be considered as the three primaries. In the United States, in 1953, the Federal Communications Commission (FCC) adopted the National Television Standards Committee (NTSC) recommendations for use as a standard in colour televisions. In Europe, the German standard, (PAL) defined by European Broadcast Union (*BEU*), and the French standard (SECAM) are employed. The NTSC considers illuminant C as the white reference while the *EBU* has preferred the illuminant D_{65}.

In computer-controlled displays different primary systems are adopted by constructors, which are different from the CIE standards for CRT televisions.

No matter which illuminant is employed, any primary system can be transformed to another using a transformation matrix P. As an example to transform the RGB system of system to the RGB system of NTSC the equation (3.3.4) is used:

$$\begin{bmatrix} R_{NTSC} \\ G_{NTSC} \\ B_{NTSC} \end{bmatrix} = \begin{bmatrix} 0.6752 & 0.1252 & 0.0727 \\ -0.1035 & 1.3237 & -0.2004 \\ 0.0060 & -0.0691 & 0.8971 \end{bmatrix} \begin{bmatrix} R_{CIE} \\ G_{CIE} \\ B_{CIE} \end{bmatrix} \tag{3.3.4}$$

c) XYZ system In 1931, the works of Judd [Jud30] led the CIE to propose a *colorimetric reference system* with non-physical primaries, *X Y Z*. The system is obtained from a primary

transformation matrix **P**, Equation (3.3.5) and the equal energy white of CIE (illuminant E) is considered as reference.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = P \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.3.5}$$

$$P = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5942 \end{bmatrix} \tag{3.3.6}$$

In general, the transformation matrix that converts any RGB system to XYZ is a linear transformation that considers chromaticity coordinates of both primaries and white points. That means: for any RGB system there is a transformation matrix P converting the RGB system to XYZ system, equation (3.3.6), and there is an inverse transformation matrix P^{-1}, equation (3.3.8), to perform the inverse operation.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = P^{-1} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \tag{3.3.7}$$

$$P^{-1} = \begin{bmatrix} 2.7690 & 1.7518 & 1.1300 \\ 1.0000 & 4.5907 & 0.0601 \\ 0.0000 & 0.0565 & 5.5942 \end{bmatrix} \tag{3.3.8}$$

X, Y and Z primaries were chosen in the way that the triangle made by them contains the spectral locus 3.14, implying that all spectral colours can be matched in an additive synthesis with positive weights of primaries. In spectral domain colour matching functions, $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$, do not take negative values in any part of visible spectrum, Figure 3.15. The $\overline{y}(\lambda)$ is identical to the luminosity function, $V(\lambda)$, for the standard observer presented in figure 3.11. The chromaticity diagram of xy is presented in figure 3.16.

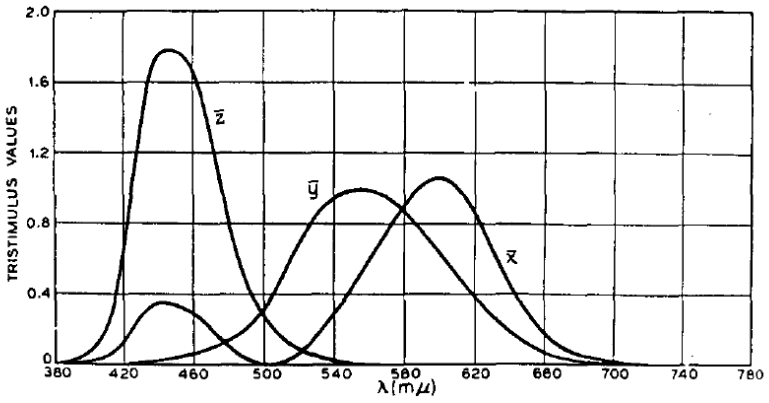


Figure 3.15: Tristimulus values, $\overline{x}, \overline{y}, \overline{z}$. Image from [LRT77].

An additive colour synthesis can be easily performed using the xy diagram. All the colour mixture properties of the rg diagram are also valid for the xy diagram. The chromaticity of a colour mixture is the linear combination of the chromaticity of the mixed colours and the luminosity of the mixture is equal to the sum of the luminosity of each components.

Suppose that we wish to compute the primaries X, Y and Z, that are required to identify subjectively the colour of a stimulus. The colour of a stimulus is the result of the interaction

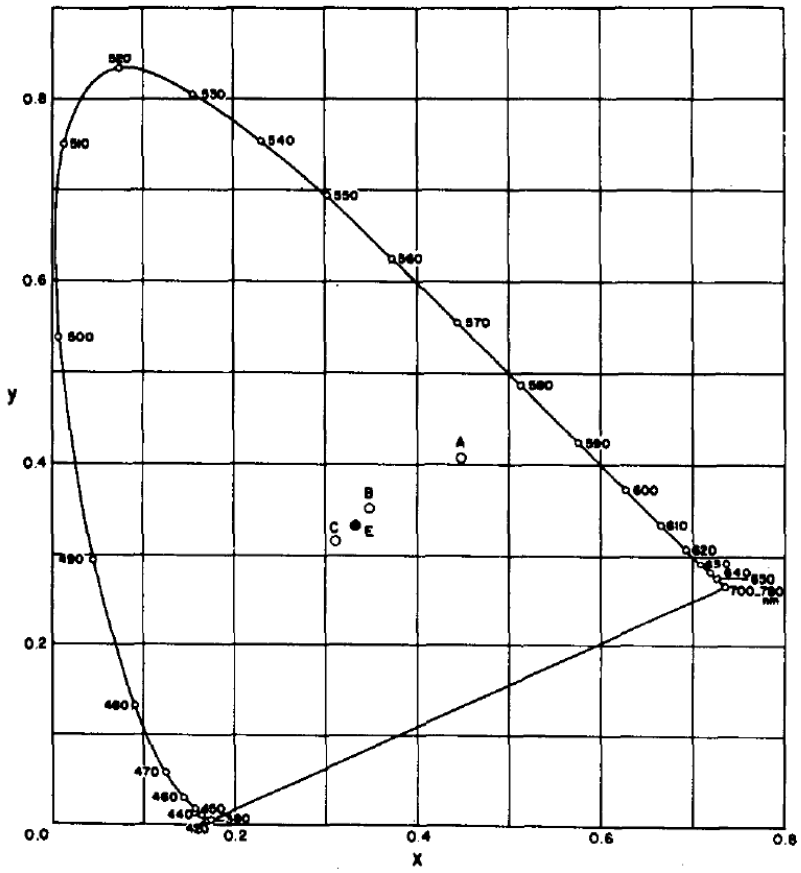


Figure 3.16: CIE xy diagram. Image from [LRT77].

between the spectral illuminant, $I(\lambda)$, which is composed of an infinite series of spectral colours, and the spectral reflectance of the stimulus, $R(\lambda)$. X, Y and Z primaries are expressed by relation (3.3.9).

$$\begin{aligned} X &= k \sum_{\lambda_{min}}^{\lambda_{max}} S(\lambda).\overline{x}(\lambda).\Delta\lambda \\ Y &= k \sum_{\lambda_{min}}^{\lambda_{max}} S(\lambda).\overline{y}(\lambda).\Delta\lambda \\ Z &= k \sum_{\lambda_{min}}^{\lambda_{max}} S(\lambda).\overline{z}(\lambda).\Delta\lambda \end{aligned} \tag{3.3.9}$$

where

– k is a normalisation coefficient. In absolute colorimetry k is a constant expressed in terms of maximum efficiency of radiant power, equal to 683.002 $lum.W^{-1}$. In relative colorimetry, k is chosen to result an Y value equal to 100 for a chosen reference white point: $k = \frac{100}{\sum_{\lambda_{min}}^{\lambda_{max}} S(\lambda).\overline{y}(\lambda).\Delta\lambda}$.

– $S(\lambda)$ is the spectral radiance. For an illuminant the spectral radiance $S(\lambda)$ is the indecent illumination $I(\lambda)$ and for a reflective stimulus is the product of the incident illumination,

$I(\lambda)$ and spectral reflectance $R(\lambda)$: $S(\lambda) = I(\lambda).R(\lambda)$.
– $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$, are the tristimulus values of standard observer

In other words each of the tristimulus values at each step of wavelength, $\Delta\lambda$, are obtained from the product of their colour matching functions, $\overline{x}(\lambda)$, $\overline{y}(\lambda)$ and $\overline{z}(\lambda)$, and the spectral radiance $S(\lambda)$. The tristimulus values X, Y and Z are equal to the sum of the products. The luminosity information is obtained from tristimulus value of Y, while the luminosity of the other two components is zero.

3.4 Colorimetry of a computer-controlled colour display, an LCD display particularly

In this section we express the measurements necessary to obtain the characteristics of a display. The characteristics are the rgb curves of the display which is considered as an illuminant, and the relation between digital value and the gain of R,G and B channels.

3.4.1 Instrument

The colorimetric instruments are similar to those used for spectrophotometry purposes. In general two types of instruments are used in colorimetric measurements: spectral colour measurement instruments such as spectrodiameter/spectrophotometer and filter-based instruments like tristimulus colorimeter [Sha03].

– Spectrodiameter: A spectrodiameter analyses the spectral distribution of an optical radiation as a function of the wavelength. In the colorimetric applications only the visible spectrum is examined. Spectrometers can measure the photometric characteristics of both self luminous and reflective stimuli. The measurements are performed by sampling the spectral distribution of the stimulus. The spectral distribution of unknown samples is compared to the spectral distribution of known illuminants, such as D65, and also to the chromaticity functions of the standard observer. These comparative measurements allow computing the tristimulus values of the unknown samples.

– tristimulus colorimeter: A tristimulus colorimeter measures only colour tristimuli. They provide the numerical data that represent the difference of colour between a reference sample and the unknown samples. The principal advantages of the tristimulus colorimeter, to be widely used in applications such quality control, are the simplicity, low measurement time as well as low cost. However, the new technologies used in spectrodiameters and their price drop made them more competitive.

3.4.2 A numerical example

A numerical example of the colorimetric measurements, that we carried out to determine the characteristics of an LCD display, is presented in this section.

First the light emitted from the computer-controlled display was measured for a series of monochrome step wedges, R, G and B. Step wedges were displayed in video format to have exactly the same configuration as the stimuli of the experiment to be carried out. Each series of step wedges was consisted of 18 levels of digital values from 0 to 255, Table 3.2. We measured the light from our computer-controlled LCD display using a *Photo Research PR650* colorimeter.

Table 3.2: Measured gain for R, G and B channels for 18 levels of digital values

		red		green		blue	
dv	dv_255	gain	luminance	gain	luminance	gain	luminance
0.00	0	0.00	0.20	0.00	0.20	0.00	0.20
0.06	15	0.00	0.29	0.00	0.46	0.00	0.23
0.12	30	0.01	0.54	0.01	1.20	0.01	0.30
0.18	45	0.02	1.04	0.02	2.70	0.02	0.42
0.24	60	0.04	2.00	0.04	5.18	0.04	0.63
0.29	75	0.07	3.22	0.07	8.81	0.07	0.93
0.35	90	0.11	4.75	0.11	13.01	0.11	1.30
0.41	105	0.16	6.60	0.15	17.73	0.16	1.77
0.47	120	0.20	8.56	0.20	23.07	0.20	2.28
0.53	135	0.26	10.86	0.25	29.35	0.26	2.81
0.59	150	0.32	13.37	0.32	36.73	0.31	3.43
0.65	165	0.40	16.53	0.39	45.01	0.39	4.24
0.71	180	0.48	19.93	0.47	54.72	0.47	5.11
0.76	195	0.58	23.89	0.57	65.98	0.56	6.13
0.82	210	0.69	28.37	0.65	76.19	0.68	7.44
0.88	225	0.78	32.17	0.76	88.53	0.78	8.58
0.94	240	0.88	36.12	0.87	101.70	0.87	9.65
1.00	255	1.00	41.04	1.00	116.70	1.00	11.20

We obtained the data of the light emitted for each channel for the visible spectrum from 350 to 750 nm and a sampling step of $\Delta\lambda = 4$. Figure 3.17 shows the radiance as a function of the wavelength for each channel and for the 18 series of step wedges.

For each one of the monochrome stimulus we computed the stimulus value Y, Equation (3.4.1).

$$\begin{aligned} Y_R &= k \sum_{\lambda_{min}}^{\lambda_{max}} S_R(\lambda).V(\lambda).\Delta\lambda \\ Y_G &= k \sum_{\lambda_{min}}^{\lambda_{max}} S_G(\lambda).V(\lambda).\Delta\lambda \\ Y_B &= k \sum_{\lambda_{min}}^{\lambda_{max}} S_B(\lambda).V(\lambda).\Delta\lambda \end{aligned} \tag{3.4.1}$$

Where $S(\lambda)$ is equal to the emitted radiance $I(\lambda)$ for each channel.

In LCD displays a non-linear function relates input digital values to output luminance. For the calibration purposes this relation must be determined. Figure **??** shows the measured output gain as a function of digital value for each channel for the LCD display of our experiments.

We used a very common model-based data fitting approach to determine the relation between digital input values and output luminance. The model parameters were fitted to the measurements via a linear least square regressions. Figure 3.18 shows the resulting curves.

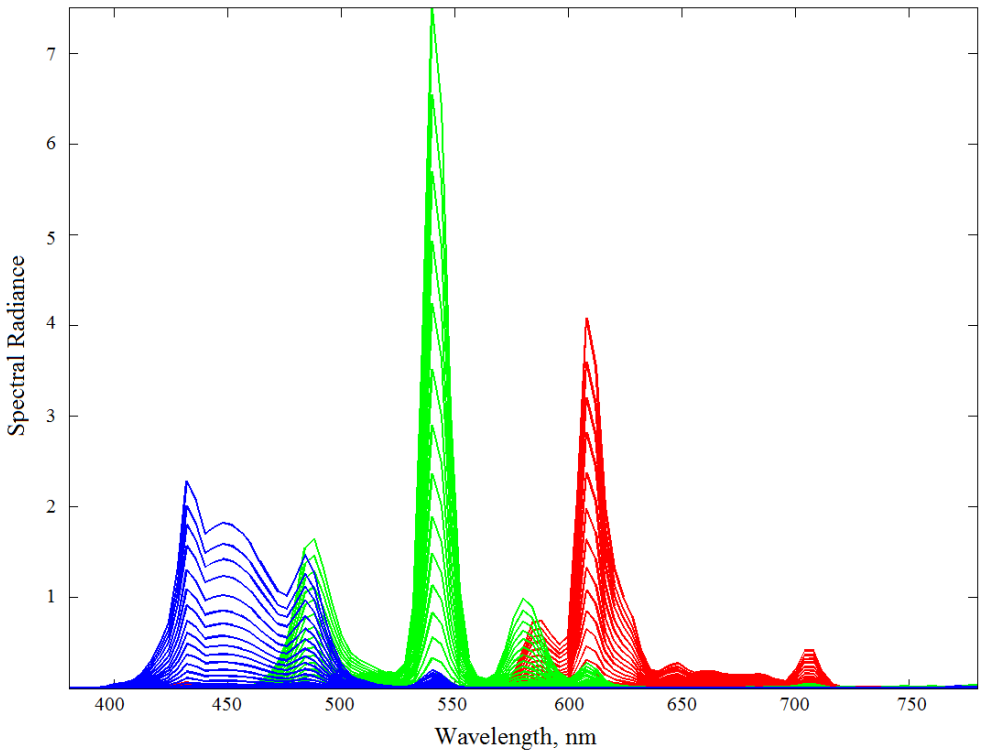


Figure 3.17: Spectral radiance for 18 levels of digital values for R, G and B channels. The data were obtained from the colorimetric measurements that we carried out on the LCD monitor which is used in our eye-tracking experiments.

3.5 Colour to grayscale conversion

Colour to grayscale conversion of videos and images is necessary in certain applications. A common example is rendering colour videos to a monochrome device. Another example is the colour documents printed in grayscale. In such applications the perceptual properties are needed to be preserved. Grayscale conversion might also be a pre-processing step in the context of vision algorithms, for example in stero-matching algorithms.

Colour to grayscale conversion is considered as a *dimensionality reduction* problem from 3-D information to 1-D representation. The conversion is, therefore, lossy. Different grayscale conversion methods have been proposed to reduce the loss of information according to the needs of applications.

According to [Ben+10] grayscale conversions could be divided in two groups: *functional* and *optimizing*.

Functional methods are pixel-wise methods that process image locally and compute for each pixel a grayscale value from the chromatic values using a given function. A simple functional method is to take the value of one of the channels as the grayscale representation of image and omit the other channels. For instance, the *value* HSV method uses the V value of HSV representation of image as its grayscale representation. On the other words the grayscale value for each pixel is the maximum of its colour values. This method is highly

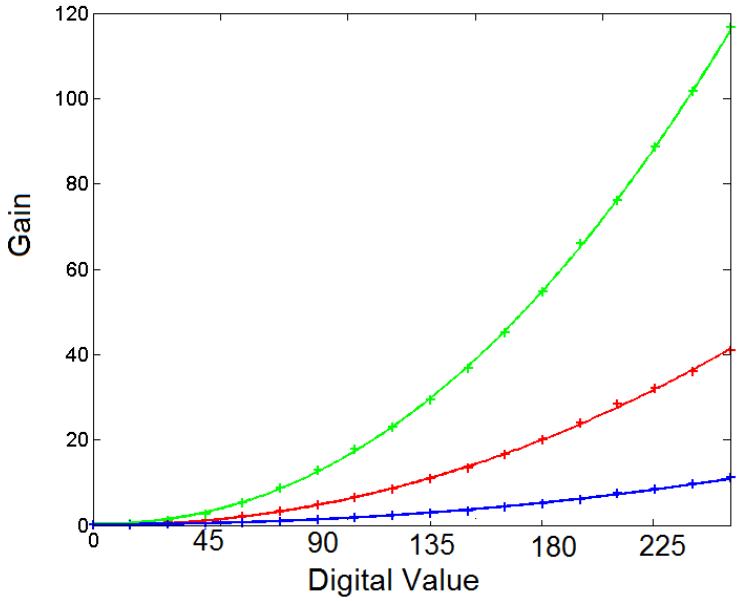


Figure 3.18: Output gain as a function of input digital value. The data were obtained from the colorimetric measurements that we carried out on the LCD monitor which is used in our eye-tracking experiments.

lossy. To consider values of the three channels, the grayscale value could be obtained from the mean of the three colour channels, which is called *naive mean* method. This method can be simply improved by using the weighted sum of colour channels according to the power distribution of channels as well as the perceptual characteristics of human observers. The (*ITU*)[2] recommends to compute the luminance information of a video signal by a relation that takes into account the luminosity function of standard observer, Figure 3.11, as well as the spectral distribution of the primaries of the display. *ITU* in recommendation 601 recommendation 601 proposes the weighted sum of relation (3.5.1) to compute the luminance of non-linear colours i.e gamma corrected colours such as in CRT displays.

$$Y = 0.299 \times R + 0.587 \times V + 0.114 \times B \tag{3.5.1}$$

This relation was adopted by NTSC[3].

In recommendation 709 for real and natural colours (HD receivers) the luminance is computed from relation (3.5.2).

$$Y = 0.2126 \times R + 0.7152 \times V + 0.0722 \times B \tag{3.5.2}$$

Optimizing methods are more advanced models that consider the whole image properties and global characteristics to compute the grayscale image that preserve the most the features

[2] *ITU*), International Telecommunication Union (*ITU*) is a specialized agency of the *United Nations (UN)* that is responsible for issues that concern information and communication technologies.

[3] *NTSC*, *National Television System Committee* was first developed in 1941 for grayscale broadcasting. In 1953 this standard was adopted to meet needs of colour television broadcasting. In this standard, instead of three colour signals, one luminance and two colour opponents signals are used. The considered luminance was compatible with the existing black and white receivers

of original image. For instance Bala and colleagues [BE04] uses a spatially adaptive algorithm to locally preserve distinction between adjacent colours. Gooch and colleagues [Goo05] have proposed a salience-preserving method. In this method the source image is processed in a perceptually uniform *CIE L*a*b** colour space. The chrominance and luminance differences of nearby pixels are used to obtain a grayscale representation of the original image. Then, the grayscale representation is modulated as function of the chroma information of source image.

At section 4.1.1.2 we present a display-dependent grayscale conversion method to find the best fit of the three colour channels to the luminosity function of standard observer. The goal of this approach is to ensure the luminosity matching between colour and grayscale video stimuli that we needed to display to perform the eye-tracking experiments related to this thesis.

3.6 Conclusion

In this chapter we introduced the mechanism of colour vision. We presented colour as a physical phenomena that is interpreted by human brain. Then we introduced, briefly, colour representation systems. Then, the colorimetric measurements which are necessary to quantify the characteristics of an LCD display, were expressed in detail. At the end we presented the grayscale conversion methods which are lossy operation.

4

How colour information influences eye movements in videos

Guiding faculty of colour features when exploring natural scenes is being debated. Several studies have been conducted to determine the contribution of different features to the deployment of attention. Wolfe and Horowitz [WH04] classified the visual attributes when performing a visual search from undoubtedly guiding attributes—colour, motion and orientation—to otherwise non-guiding attributes, such as intersection and light sources. According to the latter study, colour is one of the most guiding attributes.

Some eye-tracking studies suggest that colour has very little effect [BT06a] or no effect on eye position, but an effect on fixation duration with shorter fixations for colour images [HPGDG12]. Another study shows that the effect depends on the category of images [FHK08]. They investigated the saliency of different colour features (saturation, red–green and yellow–blue contrasts) within seven semantic categories of images: face, flower and animal, forest, fractal, landscape, man-made, and rainforest. They report that the contribution of colour features to visual attention depends on the category of the images. colour information increases the congruency of fixation position between participants in *rainforest*, while in *fractal* colour decreases the congruency.

All these studies consider static scenes, whereas natural scenes are mostly dynamic. In fact, motion is found to be one of the most crucial features in guiding eye movements [IB09; Mit+11; Mar+09]. Therefore, the present study aims at evaluating the contribution of colour to guiding eye movements for dynamic scenes.

The purpose of the current chapter is to evaluate the influence of colour information on the eye movements during free-viewing of videos. One of the main concerns in the set-up of eye tracking experiments is the content of the video stimuli and its variety. We studied two different dataset of video stimuli through experiments **A** and **B**. In experiment **A** to be presented at section 4.1, a dataset of videos with various contents from man-made indoor scenes to landscapes is studied, while in experiment **B** to be presented at section 4.2 we focus on the video stimuli containing human faces to study the influence of colour on eye movements in presence of the faces, because faces are considered as particularly salient regions of a visual scene [MMP05; CFK09].

For both experiments, we analyse the evolution of eye movements across time. We hypothesized that presence of colour information in the visual scene influences the eye movements' attributes as well as the location of gaze. The study, to be reported, examines the possible influence of colour in global, across the time and as a function of category of stimulus. We test the hypothesis using several evaluation measures and eye movement attributes, such as, eye positions' dispersion among participants, regions of regard, fixation duration and saccade amplitude. The findings from this work could provide the influence factors of colour information in a saliency model, and possible tuning weights of colour feature maps.

4.1 Eyetracking experiment A, General stimuli

Experiment **A** aims to record eye movements of participants when looking freely at video stimuli with various contents in two conditions: colour and grayscale. The recorded data would be compared to identify whether colour influence eye-movements.

4.1.1 Stimuli

Two points were took into consideration when creating the video stimuli of experiment **A**: first the variety of the contents of the video stimuli and second the conversion from colour to grayscale of original colour videos.

4.1.1.1 Content

Our dataset consisted of 20 video clips, each for about 20 seconds. These clips were created by concatenating 134 short video extracts of from one to three seconds, called *video snippets*. We concatenated the snippets to increase the heterogeneity of the visual stimuli and to reduce possible top–down processes [CI06; Mar+09]. The snippets were extracted from various colour video sources, including professional videos, such as films, TV series, and documentaries, and also amateur videos of urban roads. The stimuli had a spatial resolution of 640×480 pixels (25×19 degrees of visual angle) and a temporal resolution of 25 frames per second.

The chosen snippets were classified according to their contents into the following categories: *daylight outdoor* scenes (42 snippets), *night light outdoor* scenes (26 snippets), *indoor* scenes (37 snippets) and *urban road* scenes (29 snippets). The main difference between *urban road* and *daylight outdoor* categories was the presence of traffic signs in the former. Because traffic signs are considered as particularly salient objects in a scene [Itt05] and their significance is related to their colour, the videos including them were considered as a separate category. Figure 4.1 shows some frames from each category.

Initially, the videos were in different compressed formats. We converted all videos to no compressed AVI format.

4.1.1.2 Grayscale conversion

To compare colour and grayscale stimuli, the original colour videos must be converted to grayscale. In chapter 3.5, we presented common grayscale conversion methods. We pointed out that colour to grayscale conversion is obviously a lossy operation. But, when comparing a colour video stimulus with its grayscale version, no matter the grayscale conversion method,

Figure 4.1: Example frames in colour. The columns from left to right correspond to the categories daylight outdoor, night light outdoor, indoor, and urban road.

in addition to chrominance features some other features like intensity are modified. For instance, in NTSC grayscale conversion the weights of channels do not correspond to the LCD displays' characteristics and red channel is weighted too high. Therefore, using this method, the red pixels are brighter in grayscale image than colour image, Figure 4.2. To avoid such systematic errors and to be able to study the influence of presence of colour features on the eye-movements, the grayscale conversion must at least preserve the brightness features of the original stimuli (i.e the luminosity of grayscale pixels must be identical to the initial colour video). We use a weighted sum of colour channels to compute the grayscale version of the stimuli. But, instead of using the conventional weights of the colour channels, we measure the characteristics of our experimental display and we compute the weights of R, G and B channels for our experiment design.

Figure 4.2: Example frame for different method of grayscale conversion. (a) Original image, (b) grayscale image computed using naive mean of three colour channels, (c) grayscale image computed using NTSC grayscale conversion method, (d) grayscale image computed by the weighted sum of equation (4.1.3).

The colour to grayscale conversion that we propose is a weighted sum of R, G and B channels, equation (4.1.1), in which the weights are computed according to the spectral distribution of the R, G and B channels of the LCD display used for eye-tracking experiments. To ensure the luminance matching between colour and grayscale stimuli the right side of relation (4.1.1) must be fitted to the standard observer luminosity function.

$$Y = w_r \times R + w_g \times G + w_b \times B \tag{4.1.1}$$

We compute the w_r, w_g, w_b according to the radiance of colour channels, such that the weighted sum in equation (4.1.1) be equal to the luminosity function of the standard observer. Figure 4.3 shows the relative power of each channel maximum output measured for the LCD display as well as the luminosity function, $V(\lambda)$, of standard observer.

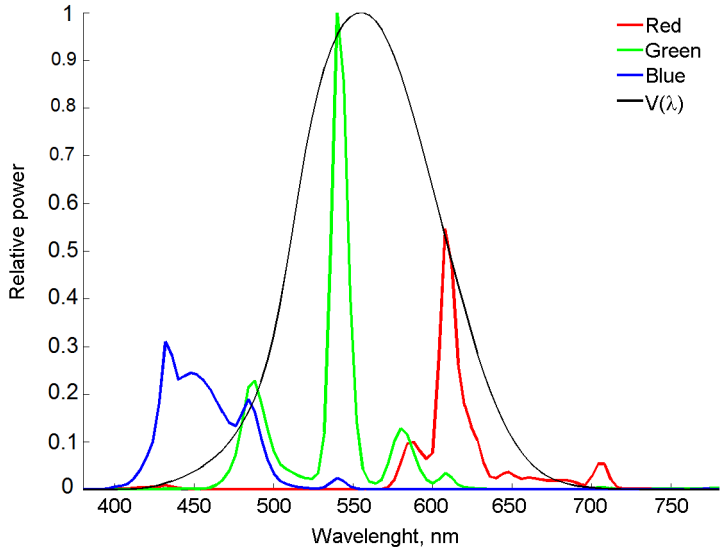


Figure 4.3: The relative power of each channel maximum output measured for the LCD display as well as the luminosity function, $V(\lambda)$, of standard observer.

A matrix operation is performed to compute the weights, Equation (4.1.2).

$$\begin{bmatrix} w_r & w_g & w_b \end{bmatrix} = \frac{Y_{m\times 1}}{\sum Y_{m\times 1}} \times \begin{bmatrix} R_{m\times 1} \\ G_{m\times 1} \\ B_{m\times 1} \end{bmatrix}^{-1} \tag{4.1.2}$$

By solving the equation for the measured data, presented at section 3.4.2, we obtain the following grayscale conversion, Equation (4.1.3):

$$Y = 0.5010 \times R + 0.4911 \times G + 0.0079 \times B \tag{4.1.3}$$

The weights are normalized to sum to result an $Y = 1$ when R, G and B are equal to 1. Figure 4.4 shows the grayscale version of some example frames.

4.1.2 Participants

Thirty-seven volunteers, (17 women and 20 men, aged from 18 to 47 years, mean = 29 ± 5.5) took part in the experiment. All reported normal or corrected to normal visual acuity,

Figure 4.4: Example frames in colour (first and third rows) and grayscale (second and fourth rows). The columns from left to right correspond to the categories daylight outdoor, night light outdoor, indoor, and urban road.

while their normal colour vision was tested using Ishihira colour plates, presented on the experimental display. All participants gave their consent to take part in the experiments.

4.1.3 Apparatus

The LCD colour monitor of 21 inches, for which the weights of equation (4.1.3) were computed, at a refresh rate of 85 Hz was used to display the video clips. The participants were at a distance of 57 cm from the display, resulting in a visual stimulus over 25×19 degrees of visual angle. The eye movements were recorded with an SR research Eyelink 1000 eye tracker. The eye tracker was used in a pupil-tracking mode at a frequency of 1000 Hz. The stimulus presentation, synchronization, and recording were carried out by software developed in our laboratory [IGGD09]. Only the dominant eye of each participant was tracked.

4.1.4 Experimental design

Each experiment session was divided into two parts. During the first part, the participants watched one-half of the video clips in one stimulus condition (colour/grayscale), while during the second part, the participants watched the other half of the videos in the other condition (grayscale/colour). Each part started with a 9-point eye-tracker calibration. Moreover, each video clip started with a drift correction. A new calibration was run if the drift error was

above 0.5 degrees. Each video was followed by a gray background displayed for 2 s. Both parts took place on the same day in a darkened room in the presence of the experimenter. The participants were asked to carefully watch the video clips while keeping their head immobile on a chin rest.

4.1.5 Data

During the experiment, the eye movements of the participants were recorded. The eyelink software reported, in a data file at each millisecond, the raw eye positions and some detected events, such as saccades, fixations, and blinks. We extracted the eye positions of the participants on the video frames, the durations of the fixations, and the amplitudes of the saccades for each participant.

Eye Positions. For each participant, 40 raw eye positions per frame were recorded. These 40 positions were summarized into a *median position* with median x and median y coordinates, referred to as the eye position of one participant per frame. To simplify the notation, the eye positions recorded under the colour stimulus condition are called *colour positions (C)*, whereas eye positions under the grayscale stimulus condition are called *grayscale positions (GS)*.

Saccade Amplitudes and Fixation Durations. The EyeLink 1000 tracker parser detects saccades according to three thresholds: motion (degrees), velocity (degrees/sec), and acceleration (degrees/s^2). Here, the acceleration, velocity, and motion thresholds were set to 30 degrees/s, 8000 degrees/s^2 and 0.15 degrees, respectively. We analysed both the amplitude of the saccades and the durations of the fixations.

4.1.6 Method

In this study we test a dataset of eye positions to identify whether colour information influences the eye movements when freely viewing video stimuli. In this section first we present the metrics that allow us to analyse the eye positions.

4.1.7 Metrics to study the position of regard

Dispersion. To evaluate the variability of the eye positions between the participants, we used a metric called the *dispersion* [Mar+09; SG00]. This metric was computed using the *leave one out* method [Tor+06]. First, the Euclidean distances between the eye position of one participant and the eye positions of the other participants were calculated. Then the final dispersion for each frame was obtained by averaging the dispersion over all participants,

$$D = \sqrt{\frac{1}{N^2} \sum_{i,j<i} {d_{i,j}}^2} \tag{4.1.4}$$

where N is the number of eye positions for a frame and $d_{i,j}$ is the Euclidean distance between the eye positions of participants i and j.

The dispersion was calculated for each frame separately, for C positions of each frame (D_C) and GS positions (D_{GS}). It measures the variability between the eye positions of the participants for each stimulus condition. Lower values of the dispersion are observed when the eye positions are located in similar positions: this is interpreted as a high level of inter-participant consistency.

Clustering. The salient objects of a visual scene correspond to the regions of interest of a scene fixated by a group of participants at the same time. These regions can be estimated for each frame by clustering the recorded eye positions. Here, we clustered the eye positions to compare the number of regions of interest between the colour and grayscale conditions.

Clustering methods use distance metrics between the eye positions to find the regions of interest. K-means is one of the clustering methods previously used to cluster eye positions [FLMB11; PS00; Lat88]. This method has one main drawback: the number of clusters must be determined a priori. Another clustering method, which leads to consistent results, is the mean-shift method. Santella and DeCarlo [SD04] employed this method on eye fixations to quantify visual areas of interest. The mean-shift algorithm is a non-parametric clustering technique which does not require prior knowledge of the number of clusters, and does not constrain the shape of the clusters. In this study, we employed this method to cluster the eye positions per frame. In this clustering method, a distance parameter is required. Since all video clips have the same size, we set empirically this distance to 100 pixels, equal to approximately four degrees of visual angle.

4.1.8 Results

The aim of this eye-tracking is to determine how colour influences eye movements during free viewing of videos. The main question is whether colour influences the location of the gaze. The design of the experiment provide the needed data to compare the eye positions recorded while viewing colour and grayscale stimuli. The influence of colour on the variability between the eye positions of the different participants is evaluated using the dispersion metric. The number of regions of interest under colour and grayscale conditions is studied using the mean-shift clustering method. These two metrics were computed for each frame. Moreover, we compared the duration of the fixations and the amplitudes of the saccades under both conditions. Finally, we compared the eye positions under the two stimulus conditions to the computational saliency maps.

Here, we analyse the effect of the stimulus category (daylight outdoors, night light outdoor, indoor, or urban roads) and the effect of the stimulus condition (colour or grayscale) on the different metrics obtained from the eye-tracking experiment: Dispersion, number of clusters, duration of fixations, amplitude of saccades, and NSS.

We also studied the temporal evolution of these metrics frame-by-frame. We limited the temporal analysis to the first 65 frames of each snippet, because most of the snippets have at least 65 frames and the influence of a top–down procedure of attention on the participants would be minimal this way. We defined three periods of observation: early (frames 1 to 15, 600 ms), middle (frames 16 to 40, one second) and late (frames 41 to 65, one second). The terminology is similar to that used by Follet et al. [FLMB11] for static images. These metrics were computed frame-by-frame and were averaged over all frames for each video snippet.

To measure the influence of colour on the eye positions we compute an ANOVA test for all 134 snippets. This is done for all evaluation metrics. Likewise, to determine the temporal evolution of metrics over 65 frames. To measure the influence of colour on the eye positions according to stimulus category we run **Bonferoni** multiple comparison.

4.1.9 Dispersion of eye positions

First the dispersion was analysed on average over the whole snippet. Figure 4.5 shows the mean dispersion under colour and grayscale stimuli according to the stimulus category.

Repeated measures ANOVA were run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (colour, grayscale) as a within-item factor.

We observed a principal effect for the *Stimulus Category* ($F(3,130) = 4.09,\ p < 0.01$). But, no effect of the *Stimulus Condition* ($F(1,130) = 2.06,\ p = 0.15$), or interaction of the *Stimulus Condition* × *Stimulus Category* ($F(1,130) = 1.28,\ p = 0.29$) was observed.

We ran Bonferroni multiple comparison tests to compare the mean dispersions obtained for the different categories. The mean dispersion for *night light outdoor* category was significantly lower than those for the categories *daylight outdoor* and *indoor* ($p < 0.01$). This was expected, because in this category only a limited region has been illuminated.

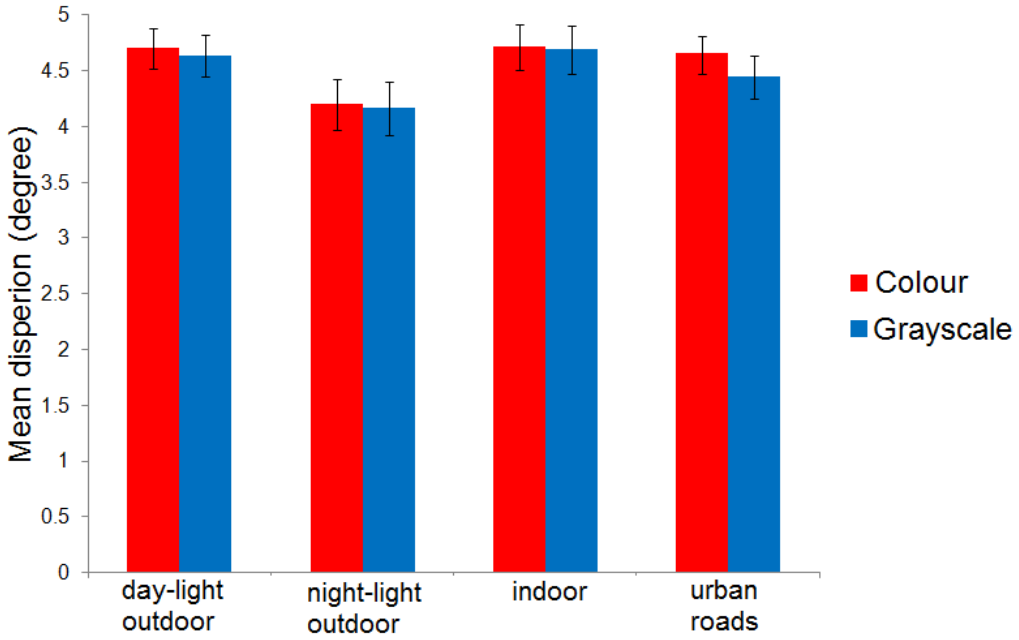


Figure 4.5: Mean dispersion according to stimulus category for colour stimuli (red columns) and for grayscale stimuli (blue columns)

We also studied the temporal evolution of the dispersion. Figure 4.6 shows the evolution of the mean dispersion for the colour and grayscale stimuli as a function of the viewing time (frame rank). The two curves followed the same pattern for both stimulus conditions. In the early period of observation, the mean dispersion reached its minimum value (*colour*, 3.2 , *grayscale*, 3.1) and increased during the middle and the late periods. Because we did not observe any principal effect of the *Stimulus Condition* for the global analysis we did not further analyse the effect of the *Period of Observation*.

4.1.10 Number of clusters in eye positions.

Clustering the eye positions determined the principal fixated regions of a scene. Figure 4.7 shows the mean number of clusters for colour and grayscale stimuli according to stimulus category. As for dispersion, a repeated measures ANOVA was run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (colour, grayscale) as a within-item factor. A principal effect was observed for the *Stimulus Category*, $F(3,130) = 4,4; p < 0.005$), as well as for the *Stimulus Condition*, $F(1,130) = 4,9; p < 0.03$). However, no effect of the interaction of *Stimulus Condition* × *Stimulus Category* was observed, ($F(3,130) = 0,374; ns$).

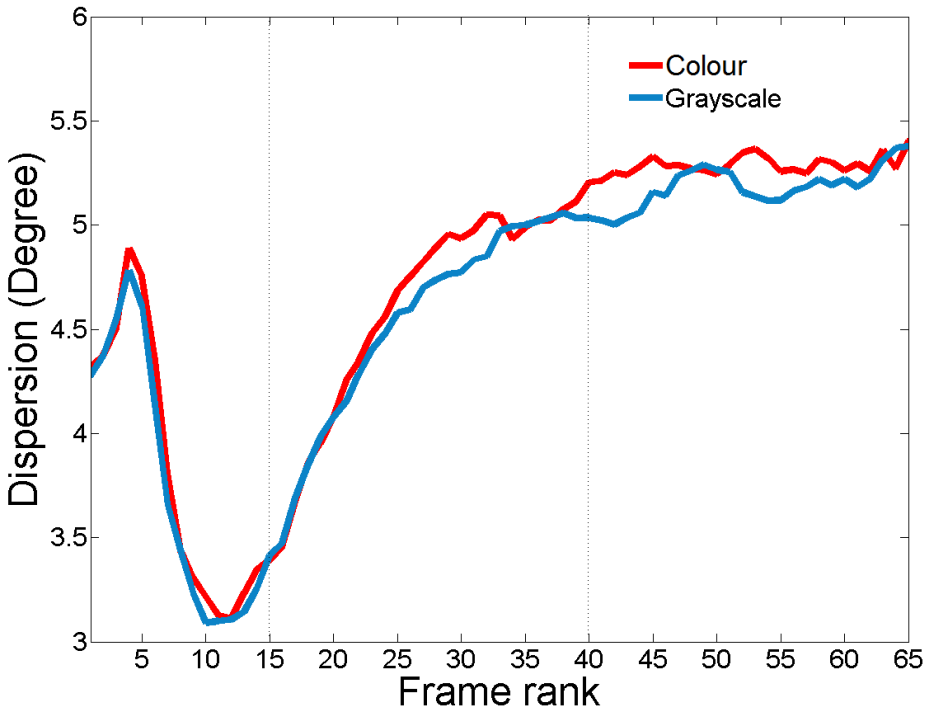


Figure 4.6: Mean dispersion according to the stimulus condition (colour and grayscale) in degrees of visual angle across time (frame rank)

Bonferroni multiple comparison tests showed that the mean number of clusters for the *night light outdoor* category was significantly lower than that for the *daylight outdoor* and *indoor* categories ($p < 0.01$).

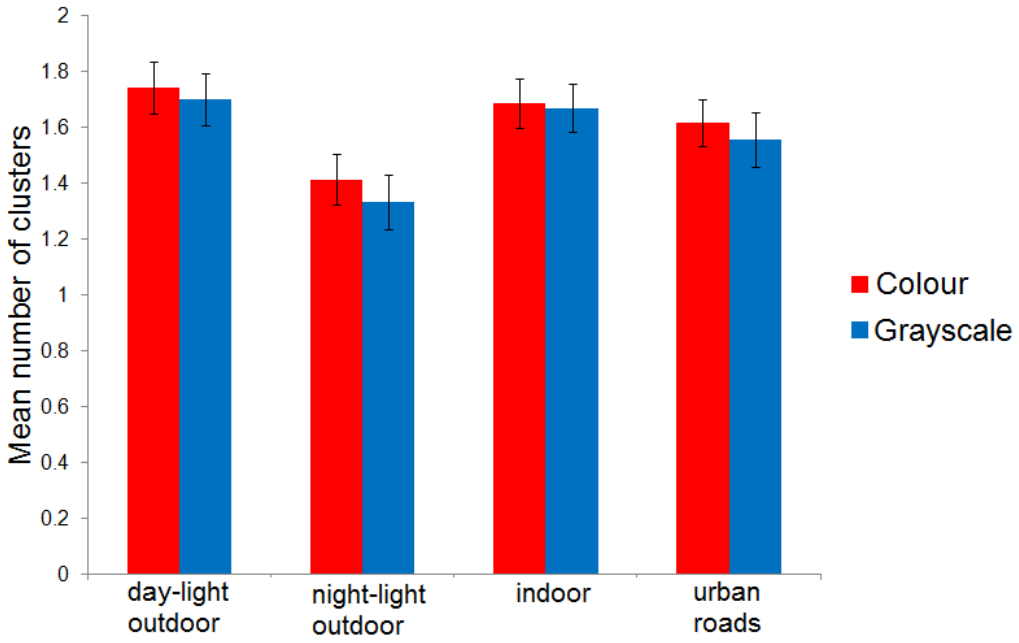


Figure 4.7: Mean number of clusters according to stimulus category

The mean number of clusters for colour stimuli was significantly higher than for grayscale (1.62 versus 1.58). Even the effect of colour was small regarding the mean number of clusters: this is an interesting result, which reveals that colour increases the number of fixated regions and hence the number of salient regions. Figure 4.8 shows an example frame with the regions of interest for colour (red ellipses) and those for grayscale (green ellipses).

Figure 4.8: An example scene depicting the different clusters. Red ellipses represent the clusters extracted from C positions and green ellipses represent the clusters extracted from GS positions.

Finally, we analysed the temporal evolution of the mean number of clusters, Figure 4.9. We ran repeated measures ANOVA with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (colour, grayscale) and *Period of Observation* (early, middle and late) as within-item factors. We observed a principal effect of the *Stimulus Condition* ($F(1,112) = 9.7; p < 0.001$), a principal effect of the *Period of Observation* ($F(2,224) = 2.46; p < 0.001$), and a principal effect of the *Stimulus Category* ($F(3,112) = 2.9; p < 0.05$). A significant effect of the interaction of the *Stimulus Condition* × *Period of Observation* was also observed, ($F(2,224) = 14.5; p < 0.0001$). Finally no effect of the triple interactions was observed. These results, showing the interesting effect of the interaction of the *Stimulus Condition* × *Period of Observation*, are interesting. As shown in Figure 4.9, in the early period of observation there is no significant difference between the mean number of clusters for colour and grayscale stimuli. But, in the middle period of observation, the mean number of clusters for colour stimuli is higher than that for grayscale, and this effect decreases in the late period of observation.

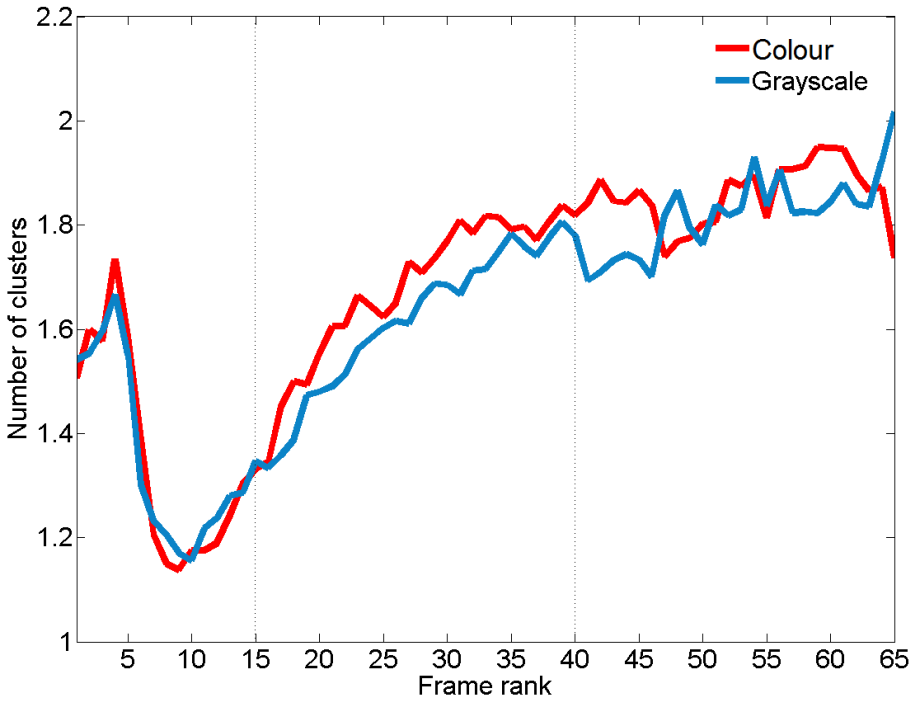


Figure 4.9: Mean number of clusters according to the stimulus condition over time (frame rank)

4.1.11 Duration of fixations and amplitude of saccades

To assess the influence of colour information on eye movements, we also studied the durations of the fixations and the amplitudes of the saccades. Two separate repeated measures ANOVA were run with the *Stimulus Category* as a between-item factor and the *Stimulus Condition* (colour, grayscale) as a within-item factor.

For the mean duration of the fixations, a principal effect of *Stimulus Category* ($F(3,130) = 11.71,\ p < 0.001$) was observed. But, we observed no effect of *Stimulus Condition* (colour, 318 ms versus grayscale, 324 ms, $F(1,130) = 0.36,\ p = 0.55$), or of the interaction of *Stimulus Condition* × *Stimulus Category* ($F(1,130) = 0.52,\ p = 0.68$). Bonferroni multiple comparisons were run to determine which categories were different from the other categories. The mean duration of the fixations for the *night light outdoor* category was significantly higher than for the other three categories (night light outdoor: 373 ms versus daylight outdoor: 307, indoor: 290 and urban roads: 314 ms, $p < 0.01$).

We also observed a principal effect of *Stimulus Category* on the amplitudes of the saccades, ($F(3,130) = 11.71,\ p < 0.001$). But, no effect of *Stimulus Condition* (colour, 4.35 degrees versus grayscale, 4.41 degrees $F(1,130) = 0.36,\ p = 0.55$), or of the interaction of *Stimulus Condition* × *Stimulus Category* ($F(1,130) = 0.52,\ p = 0.68$) was observed.

Bonferroni multiple comparisons determined that the mean amplitude of the saccades for *night light outdoor* category is significantly higher than for *daylight outdoor* (night light outdoor: 3.9 degrees, daylight outdoor: 4.52 degrees, $p < 0.05$).

Eye movements and position of gaze is highly influenced by the presence of faces in visual stimuli, such that there are evidences of a center of face perception in human cortex [KY06]. A number of studies have shown an impact of faces on eye movements when viewing static

images [Tor+06; MLH02], and also using dynamic stimuli [RPH14]. In experiment **B** we try to study weather the impact of colour information on the eye movements is different in person-present stimuli.

4.2 Eye-tracking experiment B, Face stimuli

Experiment **B** aims to record eye movements of participants when looking freely at video stimuli containing faces in two conditions: colour and grayscale. The recorded data would be compared to identify whether presence of faces in visual stimuli can modify the impact of colour on eye-movements.

4.2.1 Participants

45 volunteers (25 women and 20 men, aged from 25 to 39 years old, M= 26, SD=4,9) took part in the experiment. All reported normal or corrected to normal visual acuity, while their normal colour vision was tested using the Ishihira test on the experimental display.

4.2.2 Stimuli

Our dataset consists of 65 short video extracts of 5 to 7 seconds, called video snippets. The snippets are extracted from various open source colour videos. Stimuli had a spatial resolution of 640×480 pixels, subtending a visual angle of 25×19 degrees, and a temporal of 25 frames per second. Chosen snippets can be classified into following categories: One person (19 snippets), two persons (15 snippets), more than two persons (11 snippets) and person – absent (19 snippets). Initially the videos were in mp4 compressed format. We converted all videos to no compressed AVI format. Figure 4.10 shows example frames from each category in colour and grayscale.

Figure 4.10: Example frames in colour and grayscale. The five columns from left to right correspond to categories: One person, Two persons, More than two persons and person-absent.

4.2.3 Method

In experiment **B**, we follow the same methodology as experiment **A** which was presented at section 4.1.6. The major differences between experiments A and B are the content of video

	Content	Display form: Clip/snippet	Duration	Onset
Experiment **A**	Various	Clips of 10 snippets	Clips of 20 seconds long each contains about 5 snippets of 2 to 3 seconds	Fixation cross at the beginning of each clip
Experiment **B**	Faces	Snippets	5 to 7 seconds	Fixation cross at the beginning of each snippet

Table 4.1: The major differences between Experiments **A** and **B**.

stimuli , the duration of video snippets, the onset point. In experiments **A**, video snippets were concatenated and formed 20 video clips that each clip was displayed after a fixation cross at the middle of display. While in experiment **B**, video snippets are not concatenated; each snippets is display after a fixation cross at the middle of display. Table 4.1 illustrates major differences between experiments **A** and **B**.

4.2.4 Results

Dispersion. As shown in Figure 4.11, the dispersion of colour eye positions is significantly higher than grayscale ($t(63) = 2,5804, p < 0.01$). This raw result shows that there is more variability between the eye positions of observers when viewing colour videos. Yet, a large dispersion might be observed in two different situations: (i) when all observers look at different areas, or (ii) when there are several distant clusters of eye positions. We later measured the number of clusters in eye position data.

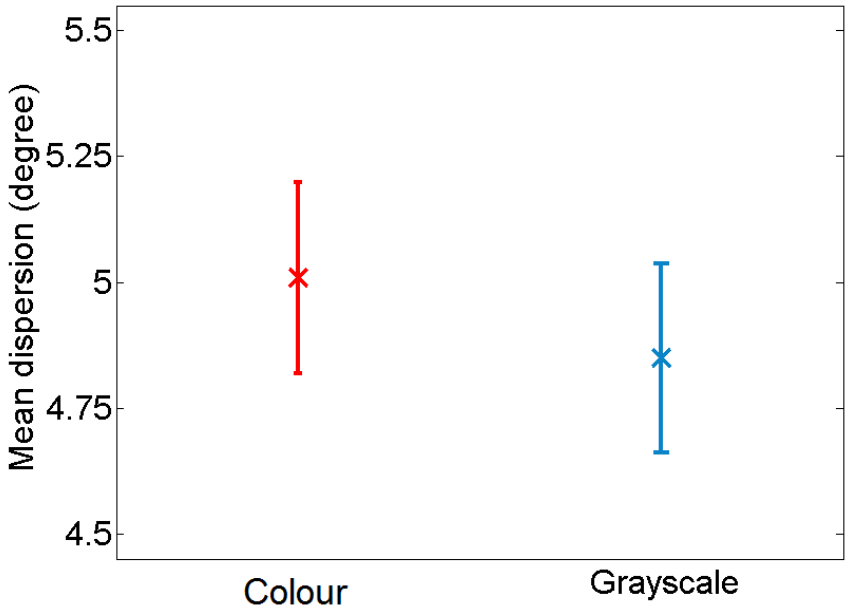


Figure 4.11: Mean dispersion in degree of visual angle over all video snippets according to the stimulus condition with standard errors.

We analysed eye positions as a function of time. Since the snippets were at least 3 seconds long, we limited the temporal analysis to the first 3 seconds—76 frames— of each snippet.

We defined three periods of viewing time: early (frames 1 to 25, one second), middle (frames 26 to 51, one second) and late (frames 52 to 76, one second). The terminology is similar to the one used by Follet and colleagues [FLMB11] for static images.

Figure 4.12 shows the evolution of the dispersion metric for C and GS eye positions, D_C and D_{GS} as a function of viewing time (frame rank). The dispersion curves follow the same pattern in both stimulus conditions. In early period of viewing, the dispersion is lower than middle and late periods and increases with time ($Early: D_C = 2.7,\ D_{GS} = 2.6,\ middle: D_C = 5.2,\ D_{GS} = 5.1,\ Late: D_C = 5.9,\ D_{GS} = 5.6,$). The difference between D_C and D_{GS} is not significant in the early period of vision, but it increases over time such that the difference between mean of D_C and D_{GS} is more prominent in the late period of observation than in the middle period ($middle: t(63) = 1,\ p < 0.2, late: t(63) = 2.37,\ p < 0.01$).

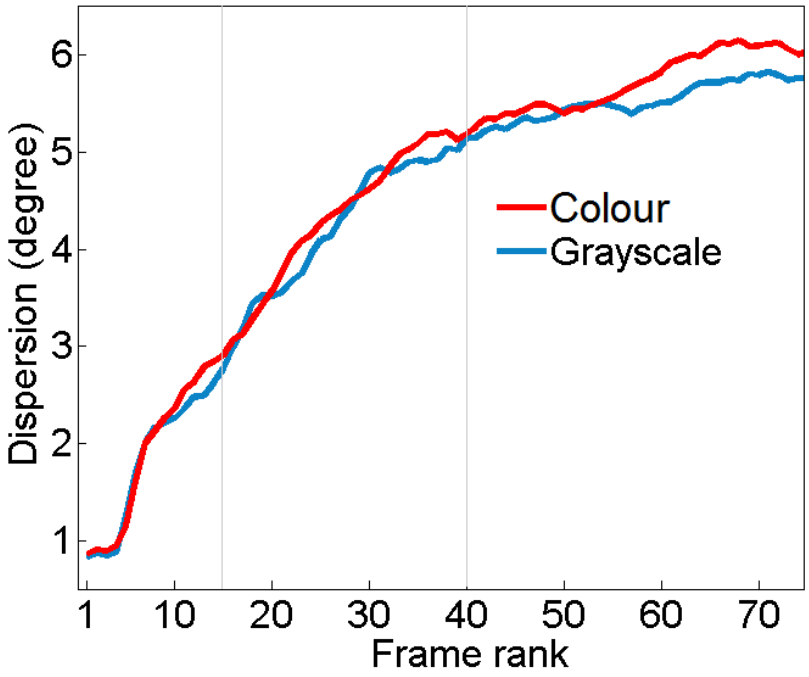
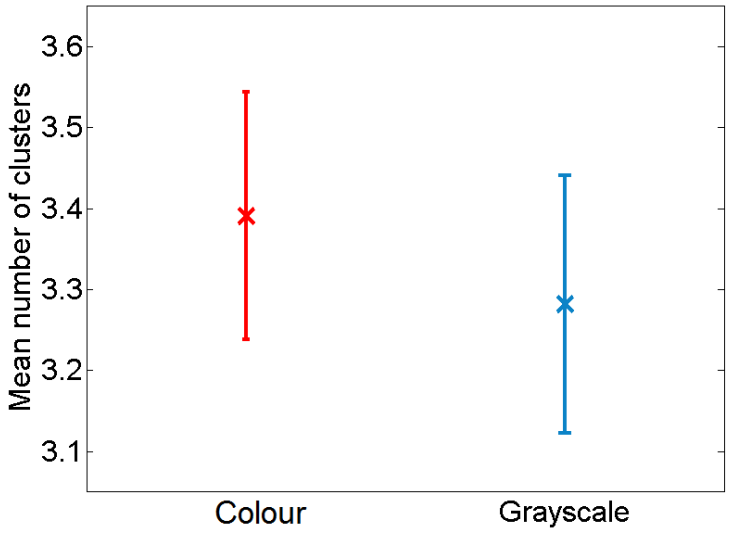


Figure 4.12: (a) Mean dispersion according to the stimulus condition in degree of visual angle across time (frame rank).

Clusters. We clustered the eye positions using mean-shift algorithm on each frame. As shown in Figure 4.14b, the mean number of clusters on colour snippets was significantly higher than grayscale ($t(63) = 2.6, p < 0.01$). The result is consistent with high dispersion values for C positions. In addition the higher number of clusters for C positions indicates that the high dispersion value of C positions is not due to high variability of the eye positions, but it is related to the higher number of regions of interest in colour stimuli.

We further analysed the number of clusters across time by accumulating the eye positions over frames of each period of viewing time. The difference between the mean number of clusters between colour and grayscale stimulus is not significant in the early and middle period of vision (early: $t(63) = 0.47, p < 0.4$, middle: $t(63) = 0.7, p < 0.3$), but in late period of observation the number of clusters in colour is significantly higher than grayscale (late: $t(63) = 1.9, p < 0.03$), Figure 4.14b. As illustrated in the qualitative example of Figure 4.15, at

(a)

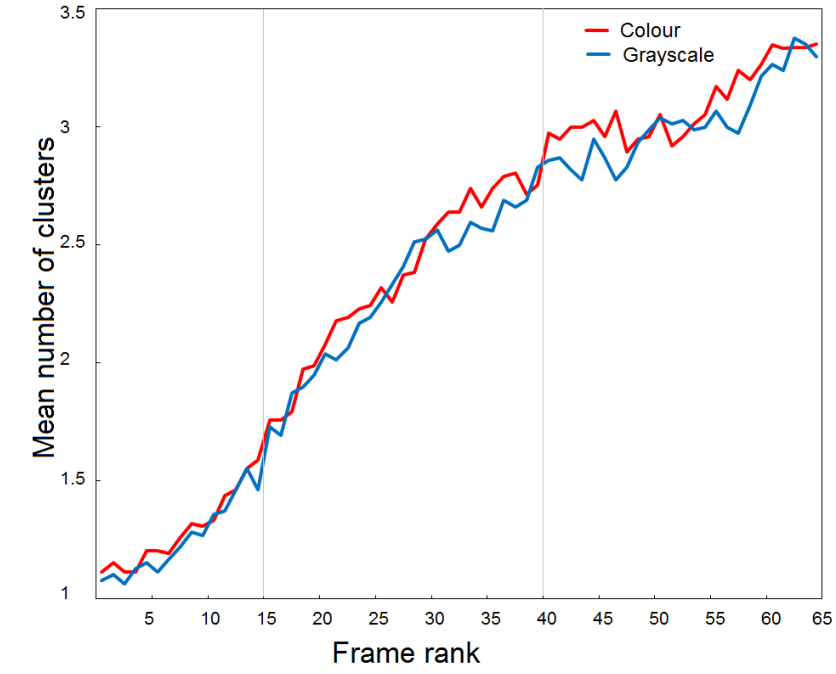


(b)

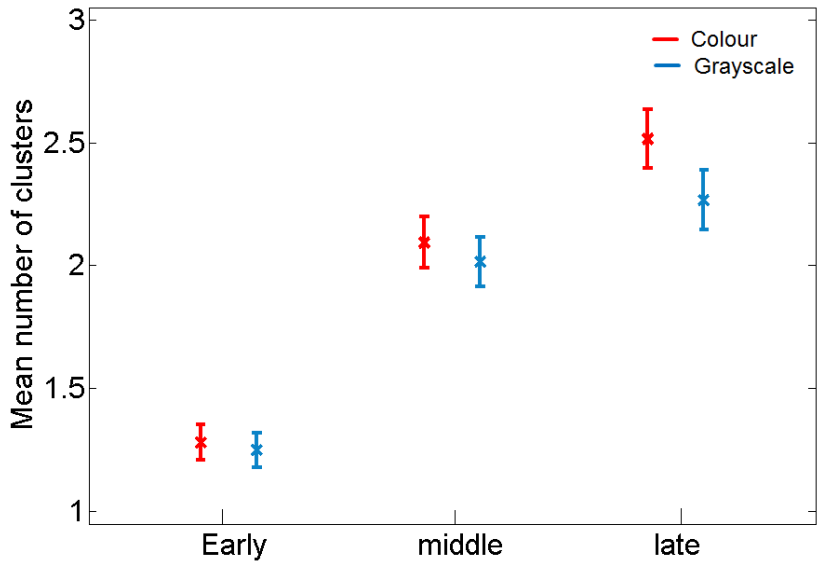
Figure 4.13: (a) An example scene depicting eye positions' clusters, red ellipses represent the clusters extracted from C positions and green ellipses represent the clusters extracted from GS positions, (b) averaged number of clusters according to the stimulus condition with standard errors.

the beginning of viewing, the main regions of interest are similar in both conditions—colour and grayscale. However, more regions of interest appear later in colour snippets.

Gaze position regarding face location To study the influence of colour on the exploration of scenes including human faces, first we studied the eye positions located on the faces (on-face eye positions). Table 4.2 shows the percentage of the eye positions that were located on the face for different categories according to the stimuli condition. The On-face eye positions are more frequent when viewing grayscale stimuli which is in accordance with the low dispersion of GS positions.



(a)



(b)

Figure 4.14: (a) Mean number of clusters per period, (b) Clusters per period for eye positions accumulated over each period.

Duration of fixation and amplitude of saccades We studied *saccade amplitudes* and *fixation durations* of the subjects while viewing videos in the two stimulus conditions (C and GS). We calculated the mean value of saccade amplitudes, as well as the mean value of fixation durations for each of the 39 subjects and ran a paired t-test. The mean value of saccade amplitude per subject in colour is significantly higher than grayscale ($t(44) = 1.7, p < 0.03$)

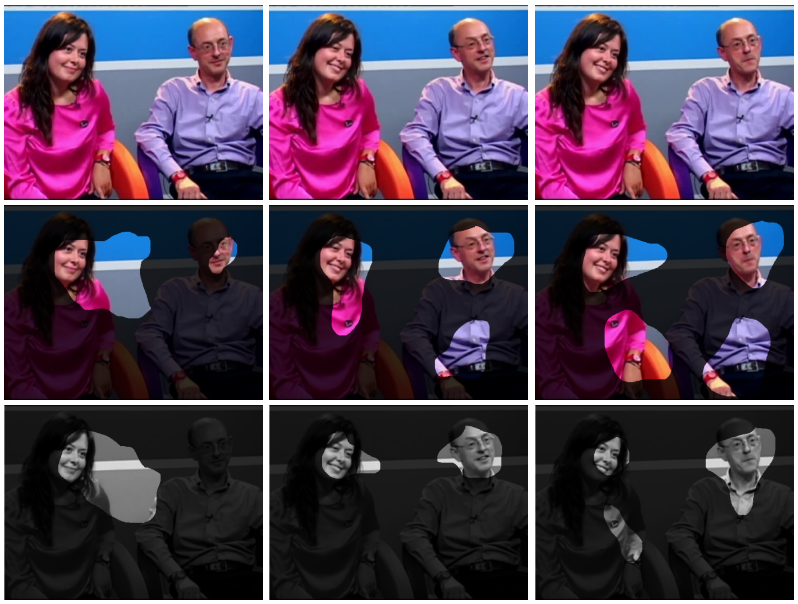


Figure 4.15: Example of the clusters obtained by accumulating the eye positions over three periods of viewing time for one snippet. The first row shows example frames of each period, the second and third rows show the clusters for colour and grayscale conditions, respectively. First column: the early period, second column the middle period, and third column the late period.

	One Face	Two Faces	More Faces
C positions	87	73	47
GS positions	90	79	47

Table 4.2: The percentage of the eye positions that were located on the face for different categories according to the stimuli condition.

while the mean value of fixation durations is slightly lower ($t(44) = 0.3, p < 0.5$), as shown in Figure 4.16.

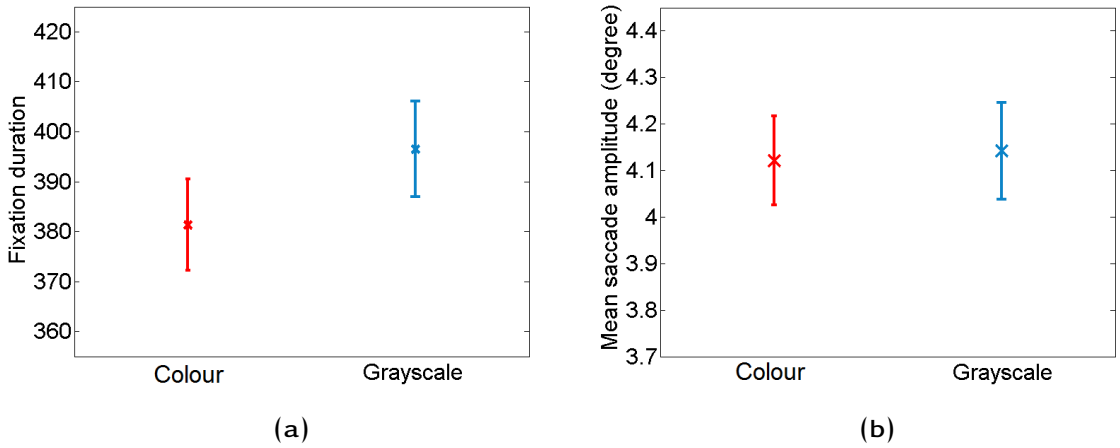


Figure 4.16: (a) Fixation duration as a function of fixation rank in ms and (b) Saccade amplitude as a function of fixation rank in degree of visual angle, according to the stimulus condition.

4.3 Discussion

In experiment A, we measured the influence of colour information on the eye movements recorded during the free exploration of videos. We compared the eye positions for colour and grayscale stimuli. We used a display-dependent grayscale conversion method to ensure the luminance matching between colour and grayscale stimuli. The grayscale version of stimuli were obtained from the weighted sum of colour channels to fit $V(\lambda)$. However, this conversion method is still lossy and the $V(\lambda)$ corresponds to the average standard observer while the response of photo-cells varies from one observer to another and the random cone mosaic of human eye might affect equiluminance thresholds [AM12].

colour and grayscale eye positions were compared using various metrics: the dispersion and the mean number of clusters to directly compare the eye positions, the mean amplitude of the saccades, the mean duration of the fixations, and finally, the similarity of the eye positions to the predictions of a saliency model. All the comparisons were also done taking into account the semantic category of the dynamic scene. We studied different categories : *daylight outdoor*, *night light outdoor*, *indoor*, and *urban roads*. Evidences from research of Frey and colleagues [FHK08] show that the influence of colour on eye positions depends on the semantic category of the image. The latter study introduced two extreme categories of static images: *fractal* and *rainforest*. In *fractal*, colour information renders the participants' fixation patterns more dissimilar, whereas in the *rainforest* category, colour increases the participants' consistency significantly. Based on the conclusions of that study, we had anticipated that the influence of colour on eye positions would be related to the category of the video snippet. Here, we instead found that the influence of colour remains insignificant across different categories of videos. Concerning the influence of category, independent from the stimulus condition, we found that for videos belonging to the *night light outdoor* category eye movements are different from the ones for the other categories.

Concerning the effect of stimulus condition, we found that colour does not influence the dispersion metric, i.e., the variability of the eye positions among participants. Yet, the number of clusters of the eye positions showed that there are slightly more clusters for colour eye positions than for grayscale eye positions. These results might suggest that colour information increases to certain extend the number of salient regions in the dynamic scenes. Moreover this effect was not constant across the periods of viewing time being larger in the middle period (frame 16 to 40).

The temporal analysis of eye positions showed a typical shape for the evolution of the mean dispersion and the mean number of clusters according to the frame rank. Note that this evolution is independent of the stimulus condition. In the early period of observation, eye positions are influenced by the central bias [Tat07; Bin10; Mar+13]. This could be observed on the two curves of figures 4.6 and 4.9. Due to this bias, a high consistency of the eye positions of participants is observed about 400 ms (the 10th frame) after the onset of a stimulus, which is in accordance with the low dispersion, as well as the small number of clusters for colour and grayscale eye positions. Then both metrics increase to reach a plateau.

In addition, for dynamic scenes, we found that colour information does not influence the duration of fixations neither the amplitude of saccades; this result differs from a previous study on static images [HPGDG12]. This difference between static and dynamic scenes, concerning the influence of colour on eye movements, could be due to the temporal changes and dynamic nature of the video stimuli. Moreover, the viewing time in the present experiment is shorter than those for the mentioned experiments with static images (Ho-Phuoc 5 sec, Frey 6 sec, present study 2 to 3 sec depending on the duration of the stimulus).

4.4 Conclusion

In this study, we compared the eye movements of different participants when viewing colour videos and the same videos in grayscale, to determine whether colour information influences eye movements. Because differences were found in static images as a function of their semantic category [FHK08], we chose videos with various contents and videos that can be classified into different categories, where colour might be more or less important. We examined the effect of colour, both globally and as a function of the category, on different parameters extracted from recorded eye movements: the eye position, the duration of the fixations, and the amplitude of the saccades. The comparison was made both on average over the whole video and frame-by-frame taking into account the course over time of the video. Such a methodology was already used in a previous study analysing the influence of sound on eye movements [Cou+12].

Whereas eye-tracking experiments do not reveal a significant influence of colour information on the eye movements when exploring videos, we showed an effect of colour on the mean number of clusters (i.e. gazed locations); with a significant effect in the middle period of viewing time. These result suggest that in some cases, colour information should be taken into account in a model of visual saliency.

5

A colour-wise saliency model

Eye movements, when exploring a visual scene, are not made randomly. They are guided by several factors such as, visual features of the stimuli, a priori knowledge of the observer, the given task to the observer, etc. Numerous models of visual attention try to predict eye movements by simulating the mechanism of visual attention. This mechanism allows selecting the relevant parts of a visual scene at the very beginning of exploration. The selection is driven by the properties of visual stimuli through bottom-up processes, as well as by the goal of observer through top-down processes [CEY04; Itt05; BK09; BI11; MK11]. Visual attention models tend to predict the parts of the scene that are likely to deploy the attention [IKN98; Fri06; LMLCB07; Mar+09; Kan+09]. Most of the models are bottom-up models based on the Feature Integration and Guided Search theories [TG80], [WCF89], which were presented at section 2.4. These theories stipulate that some elementary salient visual features such as intensity, colour, depth and motion, are processed in parallel at a pre-attentive stage, subsequently combined to drive the focus of attention. This approach is in accordance with the physiology of the visual system. Hence, in almost all the models of visual attention, low level features like intensity, colour and spatial frequency are took into consideration to determine the visual saliency of regions in static images, whereas motion and flicker are also considered in the case of dynamic scenes [IKN98], [LMLCB07], [Mar+09]. More recently, the contribution of different features like colour in guiding eye movements when viewing natural scenes has been debated. Some studies suggested that colour has little effect on fixation locations [BT06a], [HPGDG12], [FHK08]. In chapter 4, we studied the influence of colour information on the eye movements. We observed a moderate contribution of colour information in guiding eye movements for different categories of video stimuli, which brings to question the necessity of the incorporation of colour features in the saliency models [DMB10]. Here, we tend to study the contribution of colour information in the saliency models.

In this chapter, we propose a colour saliency model, which is based on the saliency model of Ho-phuoc and colleagues [HPGGD10] for static images. We incorporate, the proposed colour saliency model, into the luminance-based saliency model of Marat and colleagues [Mar+09]. First, we introduce different steps of the luminance-based model. Then, we describe the proposed colour saliency model. Afterwards, we evaluate the contribution of colour information in predictive power of the model. Finally, we compare the results to the

model of Itti and colleagues [IKN98], which is one of the reference models of the visual attention.

5.1 Luminance-based saliency model

The saliency model of Marat and colleagues [Mar+09], [Rah13] is a biologically plausible model that imitates the human visual system from retina to cortex. The initial version of the model [Mar+09] is consisted of two pathways, static and dynamic. The model has been improved latter by adding a face pathway [Rah13]. Figure 5.1 illustrates different steps of this model. Both initial and improved versions of the model are only based on the luminance-visual information.

M_{ls} and M_{cs} are luminance-static and chrominance-static maps respectively

As shown in figure 5.1, in a pre–processing step, the luminance-visual information is elaborated by retina-like filters, and then is decomposed using cortical-like filters. Static pathway processes the luminance-static information, that provides luminance-static saliency map (M_{ls}), and dynamic pathway processes motion information that provides luminance-dynamic saliency map (M_{ld}). Different processing steps of the model are presented bellow.

5.1.1 Retina-like filters

Retina model, which has been described in detail in [Mar+09], roughly simulates the functioning of photoreceptors, horizontal, bipolar, and ganglion cells without taking into account the spatially variant resolution of the retinal photoreceptors.

Photoreceptors are modelled as a low-pass Gaussian filter with a high cut-off frequency, that removes the high frequency noises from the initial grayscale image (5.1.1).

$$y = \frac{255 + x_0}{x + x_0}.x \tag{5.1.1}$$

$$where,\ x_0 = 0.1 + \frac{410g}{g+105}$$

The output image of photoreceptor P is then processed by horizontal cells that are also modelled as a low-pass Gaussian filter with a lower cut-off frequency than photoreceptor. The response from these cells is twice than the previous retinal low-pass filter.

Next in order are the bipolar cells modelled as a band-pass filter, which simply calculates the difference between outputs from photoreceptor cells P and horizontal cells, H. The bipolar cells model might be in ON or OFF position as the bipolar cells (5.1.2). In 'ON' position only the positive part of the difference is kept otherwise, in 'OFF' position the absolute value.

$$Y = ON - OFF \tag{5.1.2}$$

$$where,\ ON = |P - H|\ and\ OFF = |H - P|$$

The parvo-cellular output is the difference between the responses of ON and OFF bipolar cells. As for the magno-cellular output a low-pass Gaussian filter is added down the line, which models the amacrine cells. The magno-cellular output is equivalent to a band-pass filter to the photoreceptors' output, which keeps less high frequencies than parvo-cellular output.

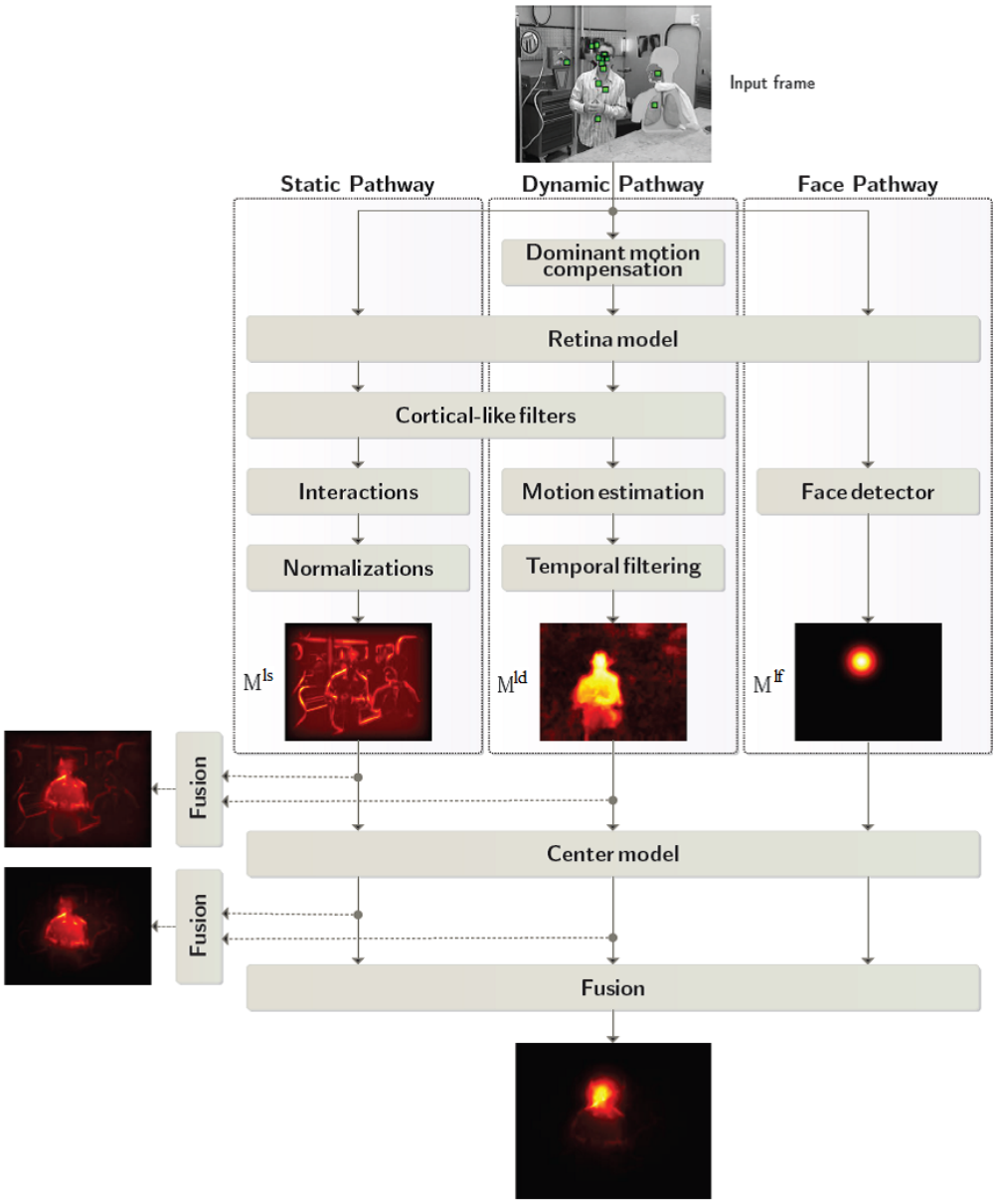


Figure 5.1: The luminance-based spatio-temporal saliency model. Image from *GIPSA-lab, AGPIG team, Perception project* .

In summary, retina model (retina-like filters) decomposes the input image into two main outputs: a parvocellular-like output that enforces the high spatial frequencies to enhance the contrasts, and a magnocellular-like output that conveys lower spatial frequencies. The first output is used to compute the luminance-static saliency maps and the latter to compute the luminance-dynamic saliency maps.

5.1.2 Cortical-like filters

The frequency and orientation selectivity of visual cortex is modelled by a bank of Gabor filters. Gabor filters are oriented band-pass filters. These filters are characterized by their

frequency selectivity and orientation. Each Gabor filter G_{ij}, (5.1.3), is defined by its central radial frequency f_j and its standard deviations σ_{ij}^{θ} and σ_{ij}^{f} in orientation θ_j and its orthogonal orientation, respectively, $i = 1,...,N_\theta$, $j = 1,...,N_f$, $\frac{f_j}{f_{j-1}} = 2$ and $f_{N_f} = 0.25$.

$$G_{ij}(u,v) = exp\left\{-\left(\frac{(u'-f_j)^2}{2(\sigma_{ij}^{f})^2} + \frac{v'}{2(\sigma_{ij}^{\theta})^2}\right)\right\} \tag{5.1.3}$$

$$where,\ u' = u\ cos\theta + v\ sin\theta\ and\ v' = v\ cos\theta - u\ sin\theta.$$

For luminance information the initial model, proposed by Marat and colleagues [Mar+09], uses six orientations and four frequencies. Hence, $N_\theta = 6$ and $N_f = 4$, Figure 5.2.

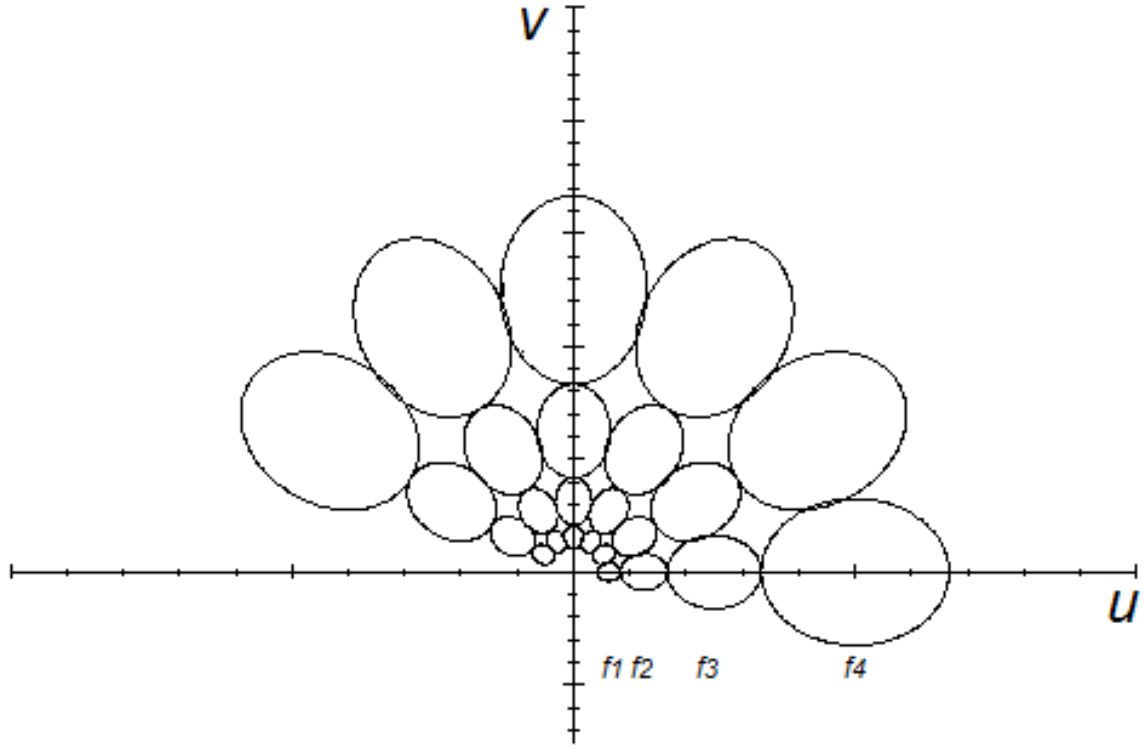


Figure 5.2: Half-value plot of the Gabor filters in the frequency plane tuned to 4 frequencies and 6 orientations (f4 = 0.25, f3 = 0.125, f2 = 0.0625, f1 = 0.0313).

5.1.3 Luminance-static saliency map

Two operations are carried out to create one luminance-static saliency map from the output of cortical-like filters, $M_{u,v}$ intermediate maps: the interactions and the normalization. The interactions between neighbouring pixels of the intermediate maps, models the lateral neural connections of visual cortex. They are modelled as linear combination of neighbouring pixels. The interactions, depending on the orientation or the frequency, may be excitatory when in the same direction, or inhibitory otherwise.

$$M_{u,v} = M_{u,v}.w \tag{5.1.4}$$

where,

$$w = \begin{bmatrix} 0 & -0.5 & 0 \\ 0.5 & 1 & 0.5 \\ 0.0 & -0.5 & 0 \end{bmatrix} \tag{5.1.5}$$

Then the intermediate maps are normalized using the method proposed by Itti et al [IKN98]. First, each intermediate map is normalized to [0 1], then it is multiplied by $(max(M_{u,v}) - \overline{M_{u,v}})^2$. Then all values lower than 0.2 are set to zero. The normalization enforces the saliency of the regions that are different from their surrounding, by unifying the dynamic range of the intermediate maps. Then a luminance-static saliency map, M_{ls} is obtained by summing up all the normalized maps, Figure **??**.

5.1.4 Luminance-dynamic saliency map

Dynamic saliency is related to the moving objects of the scene. The magno-cellular output is used to detect the objects that are moving against to the background. A differential approach is used for motion estimation by solving a system of optical flow equations [BP02]. For every frame a motion vector is defined per pixel. Only the modulus of the vector is used to define the dynamic saliency of a region, assuming that the motion saliency map of a region is proportional to its speed against the background. Then a temporal median filter is applied to five successive frames to remove the possible noisy detected motions. The output of temporal filtering is considered as luminance-dynamic saliency map, M_{ld}, Figure **??**.

5.2 Chrominance-based saliency map

In this section we present different processing steps for chrominance information. Colour information are processed in two streams: one red-green colour opponent stream and one yellow-blue stream. There are several colour spaces proposing different combination of cone responses to define the principal components of luminance and opponent colours, red-green (RG) and blue-yellow (BY) [TFMB04]. The colour space proposed by Krauskopf et al. [KWH82] is one of the validated representations to encode visual information where the orthogonal directions, A, $Cr1$ and $Cr2$, represent luminance, chromatic opponent red-green and chromatic opponent yellow-blue, respectively. Equation (5.2.1) is used to compute A, $Cr1$ and $Cr2$. In our model we use $Cr1$ and $Cr2$, as input images, to compute the chrominance-static saliency maps.

$$\begin{pmatrix} A \\ Cr1 \\ Cr2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ -0.5 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} L \\ M \\ S \end{pmatrix} \tag{5.2.1}$$

L, M and S values in equation (5.2.1) correspond to the response of the three types of cones of the human eye; their name was chosen because of their maximum sensitivity at long, medium and short wavelengths of the light. Here, L, M and S values are calculated from tristimulus values of 1931 $CIE\ XYZ$ colour space as follows:

$$\begin{pmatrix} L \\ M \\ S \end{pmatrix} = \begin{pmatrix} 0.4002 & 0.7076 & -0.0808 \\ -0.2263 & 1.1653 & 0.0457 \\ 0 & 0 & 0.9182 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \tag{5.2.2}$$

Figure 5.3 illustrates a given frame and its corresponding luminance (A) and chrominance components $Cr1$ and $Cr2$.

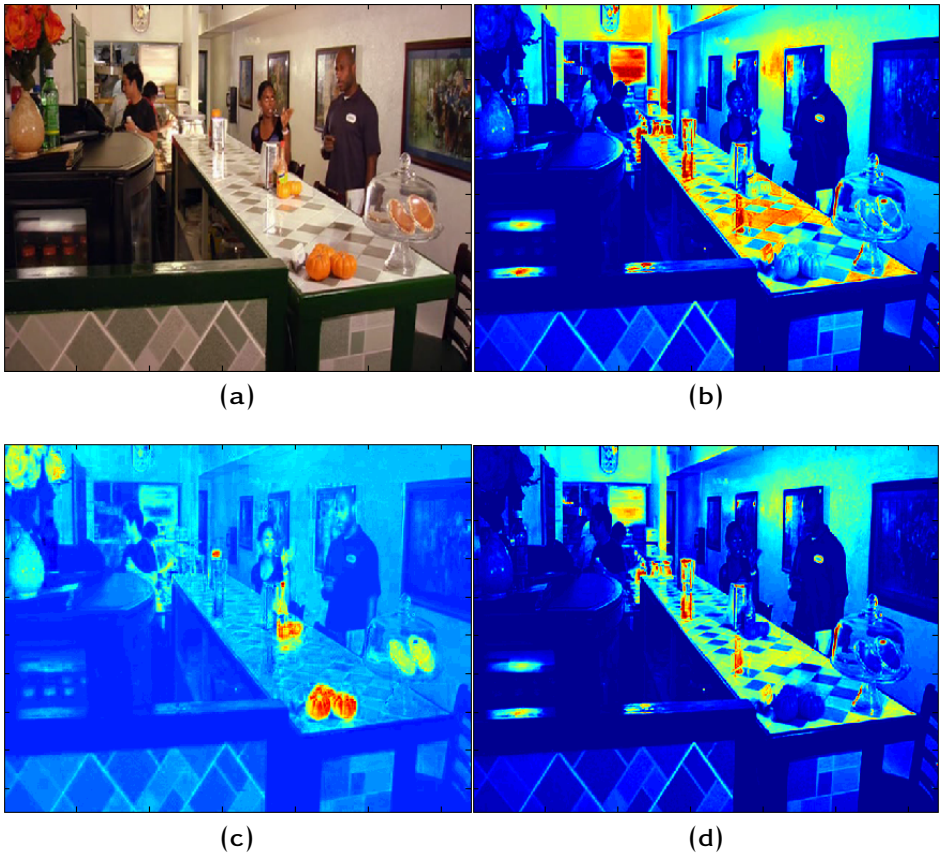


Figure 5.3: An example frame, (a) original coloured image, (b) luminance component A, (c) red-green chrominance component $Cr1$ and (d) yellow-blue chrominance component $Cr2$. Note that luminance component A and yellow-blue component $Cr2$ are highly correlated.

The different steps of the saliency model for colour opponent red-green image, $Cr1$, and colour opponent yellow-blue image,$Cr2$, are presented as follows.

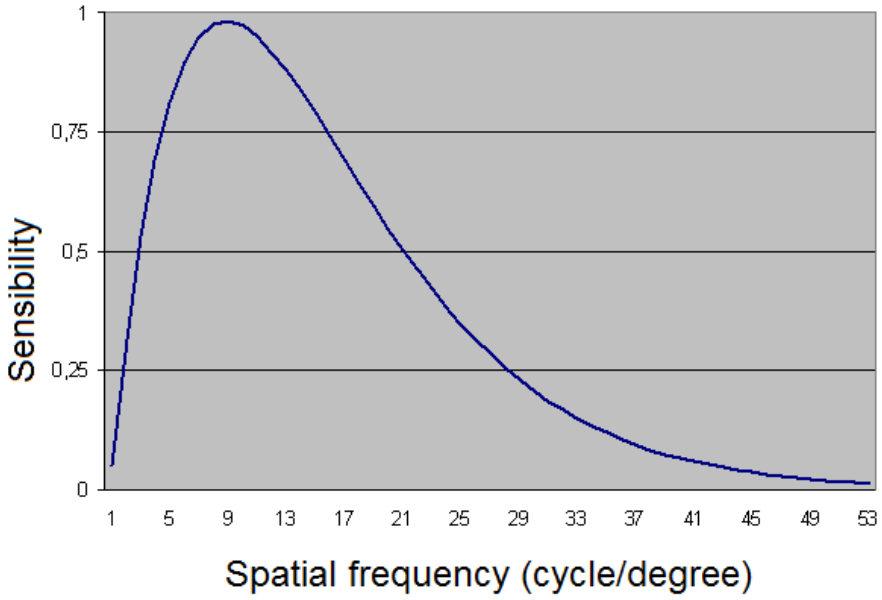
5.2.1 Retina-like filters

The function of retina, especially cone photoreceptors, is modelled using the contrast sensitivity function (CSF) of standard observer. Figure 5.4 shows the (*CSF*'s), for chrominance information (cone cells) and for luminance information (rods cells) . Note that the *CSF* for chrominance information is different from the CSF for luminance information.

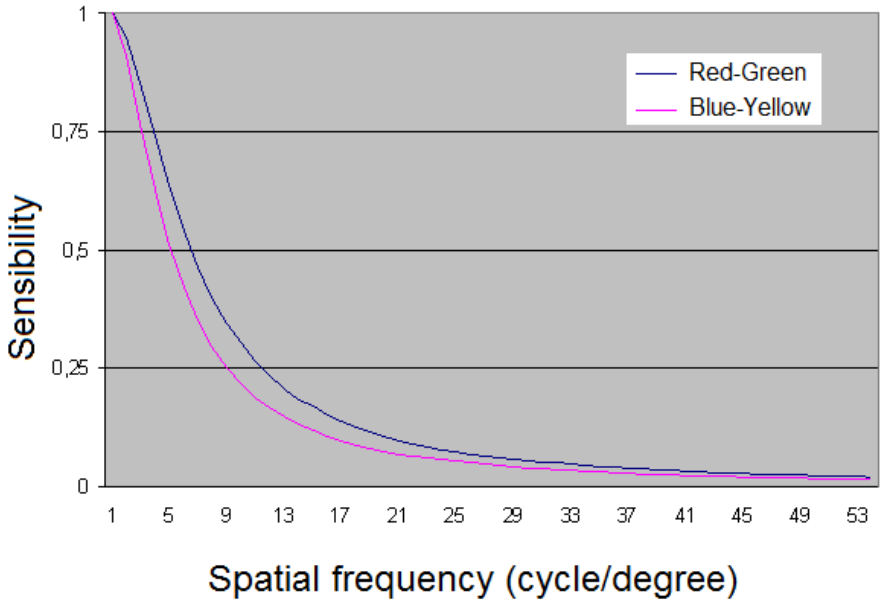
The retina models for $Cr1$ and $Cr2$ are simple low-pass filters that their transfer function reproduces the *CSF* curves depicted in figure 5.4. Cut-off frequency for red-green ($Cr1$) image is slightly higher than for yellow-blue ($Cr2$) image, 5.1 and 4.1 cycle per degree, respectively. Figure 5.5 shows the retina output images for two colour opponent images $Cr1$ and $Cr2$.

5.2.2 Cortical-like filters

Like luminance information, the cortical processing of chrominance information is modelled by a bank of Gabor filters. But, since the amplitude spectra of the two colour-opponent $Cr1$ and $Cr2$ images do not have as many specific orientations as the amplitude spectra of the luminance image [BM05], for both $Cr1$ and $Cr2$ images, only Gabor filters with four



(a)



(b)

Figure 5.4: The normalized contrast sensibility functions, (a) for luminance component and (b) for colour components red-green and blue-yellow. Image from [LM05].

orientations are used (0, 45, 90 and 135 degrees). In addition, because human visual system is less sensitive to the high spatial frequencies of chrominance information [Geg03], only two lowest frequencies are used (0.0313 and 0.0625 cycle per degree). Figure 5.6 shows intermediate cortical images for red-green $Cr1$ component.

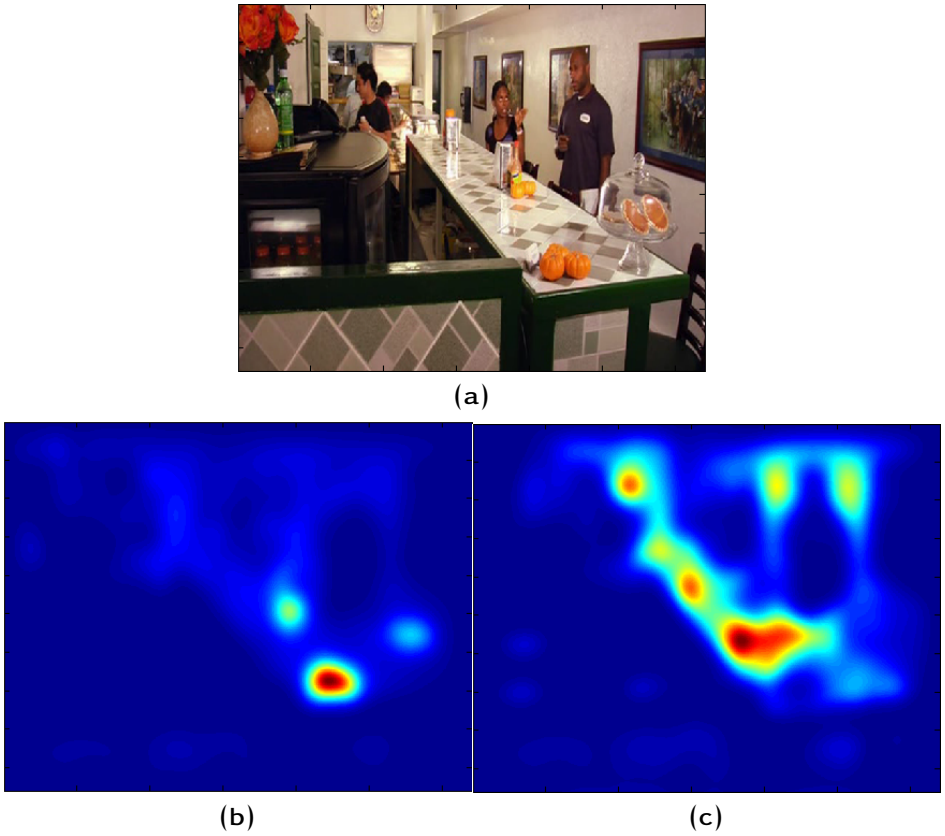


Figure 5.5: An example of retina-like filters output for (a) the original input image, (b) red-green chrominance component, $Cr1$, input image and (c) yellow-blue chrominance component, $Cr2$, input image.

5.2.3 Chrominance-based static saliency map

Likewise luminance-static saliency map, the intermediate maps of the output of cortical like filters are normalized. To compute the chrominance-static saliency map, first the red-green and blue-yellow intermediate maps are normalized to [0 1], then are summed up to obtain a chrominance-saliency M_{cs}, Figure 5.7.

5.3 Fusion

Chrominance-static saliency map M_{cs}, luminance-static saliency map M_{ls} and luminance-dynamic saliency map, M_{ld}, after normalizing to [0 1], are combined, according to equation (5.3.1), to obtain a master spatio-temporal saliency map per video frame. This map predicts the salient regions i.e. the regions that stand out in a visual scene.

$$Saliency\ \ map = \alpha M_{ls} + \beta M_{ld} + M_{cs} + \alpha\beta(M_{ls} \cdot M_{ld}) \tag{5.3.1}$$

where, α and β are the max of M_{ls} and skewness of M_{ld} respectively, and $M_{ls} \cdot M_{ld}$ is a pixel to pixel multiplication.

The weights of maps in equation (5.3.1) were found to result a good fusion regarding the fact that the saliency maps from both the static and dynamic pathways exhibit different characteristics i.e. static saliency map has larger salient regions based on textures, whereas

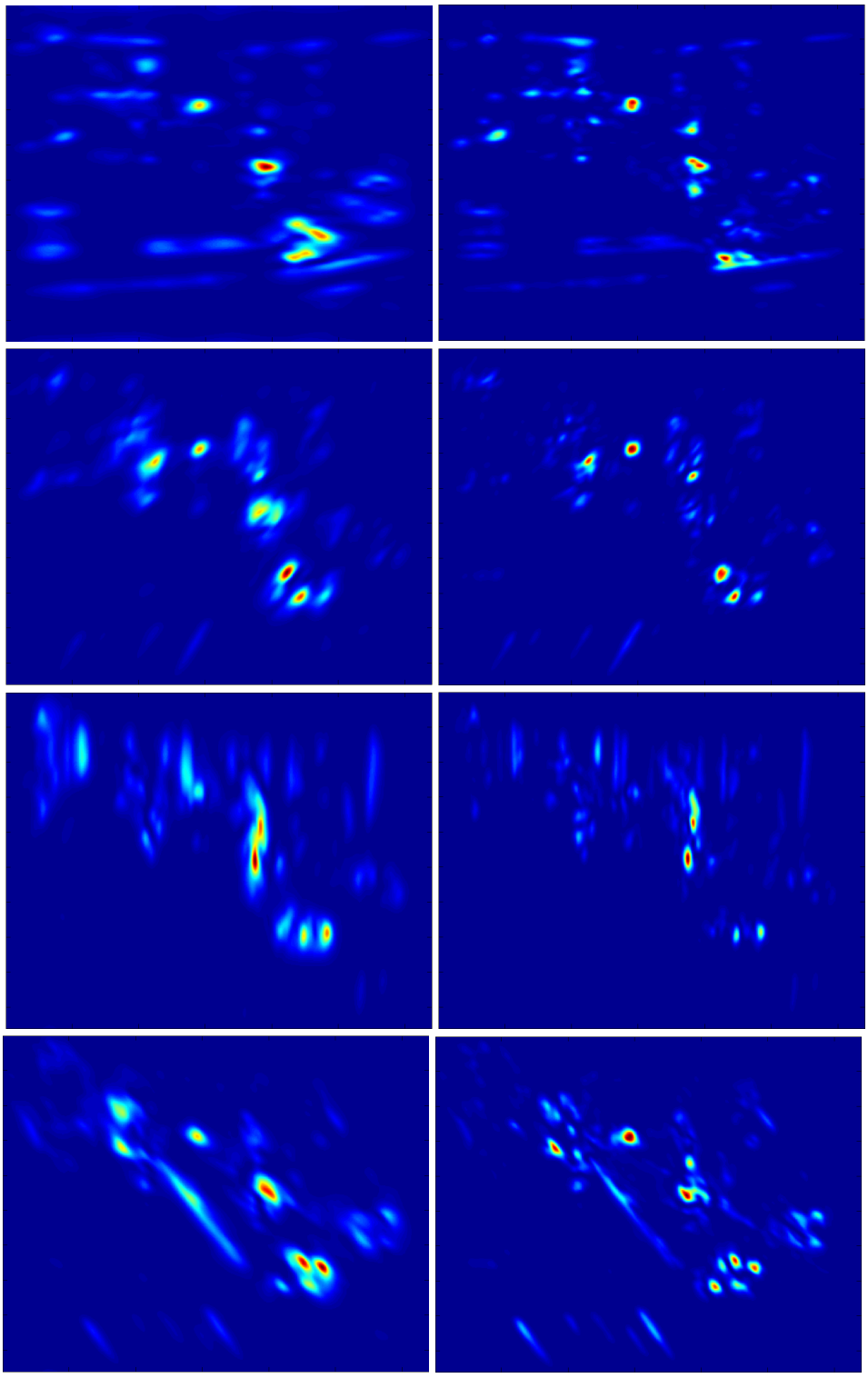


Figure 5.6: An example of output images of cortical-like filters for $Cr1$ input image of figure 5.5a, from left to right $f_1 = 0.0313$ and $f_2 = 0.0625$, from top to down $\theta_1 = 0, \theta_2 = 45, \theta_3 = 90, \theta_4 = 135$.

dynamic saliency map has smaller salient regions depending on the moving objects [Mar+09; Rah13]. Static and dynamic maps are modulated using maximum and skewness respectively. The reinforcement parameter $\alpha\beta$ is used to include the regions that have low motion, but

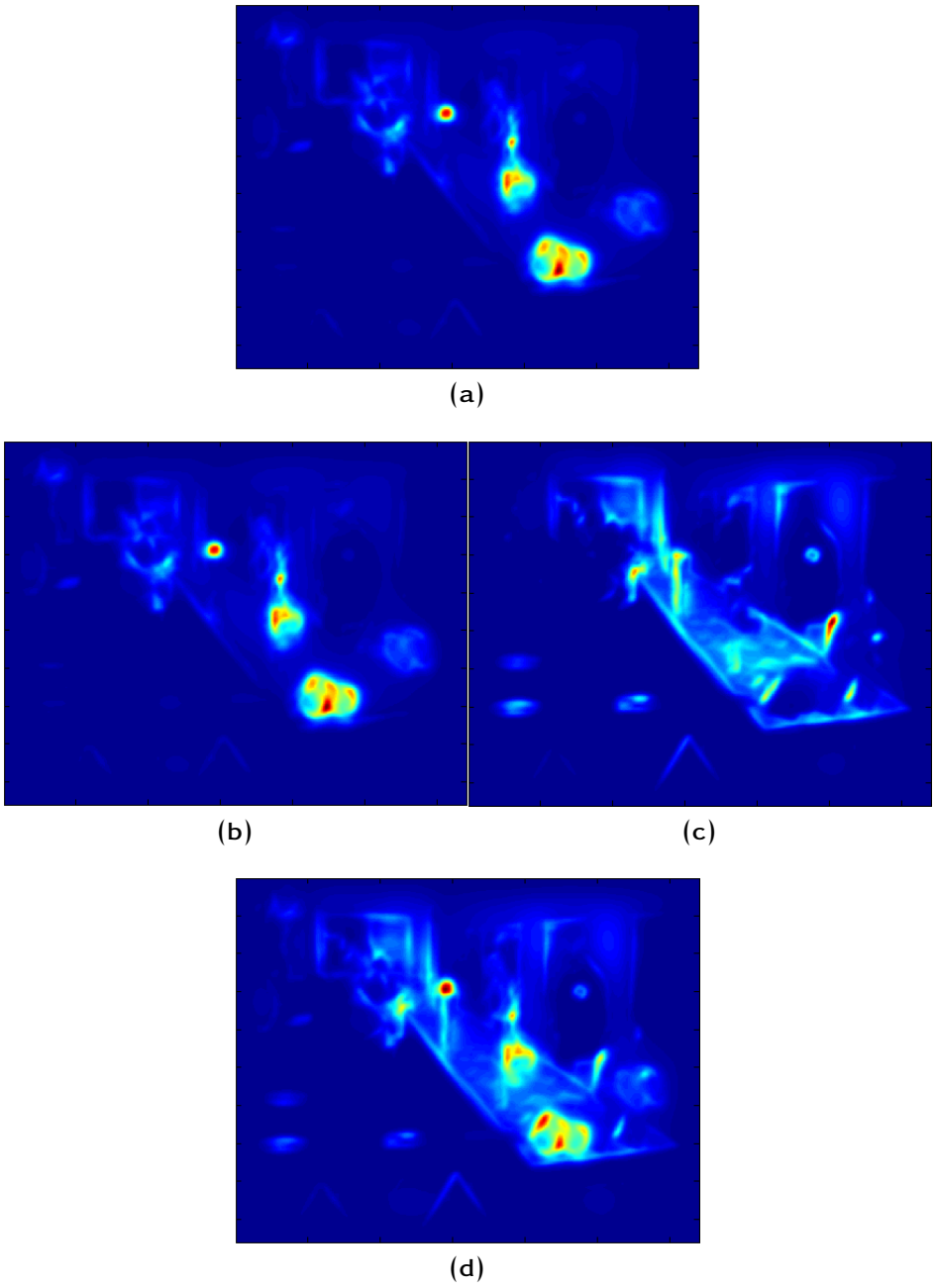


Figure 5.7: (a) Input frame in colour (b) saliency map for red-green chrominance component $Cr1$, (c) saliency map for yellow-blue chrominance component $Cr2$, and (d) final chrominance-based saliency map M_{cs}.

contain large salient regions in static saliency map. Figure 5.8 depicts the schema of our colour-wise saliency model.

5.4 GPU implementation

The saliency model presented above with static (luminance-based), dynamic (luminance-based) and chrominance pathways is compute-intensive. Rahman and colleagues [RHP11] have proposed a parallel adaptation of luminance-based pathways onto GIPSA-lab. They

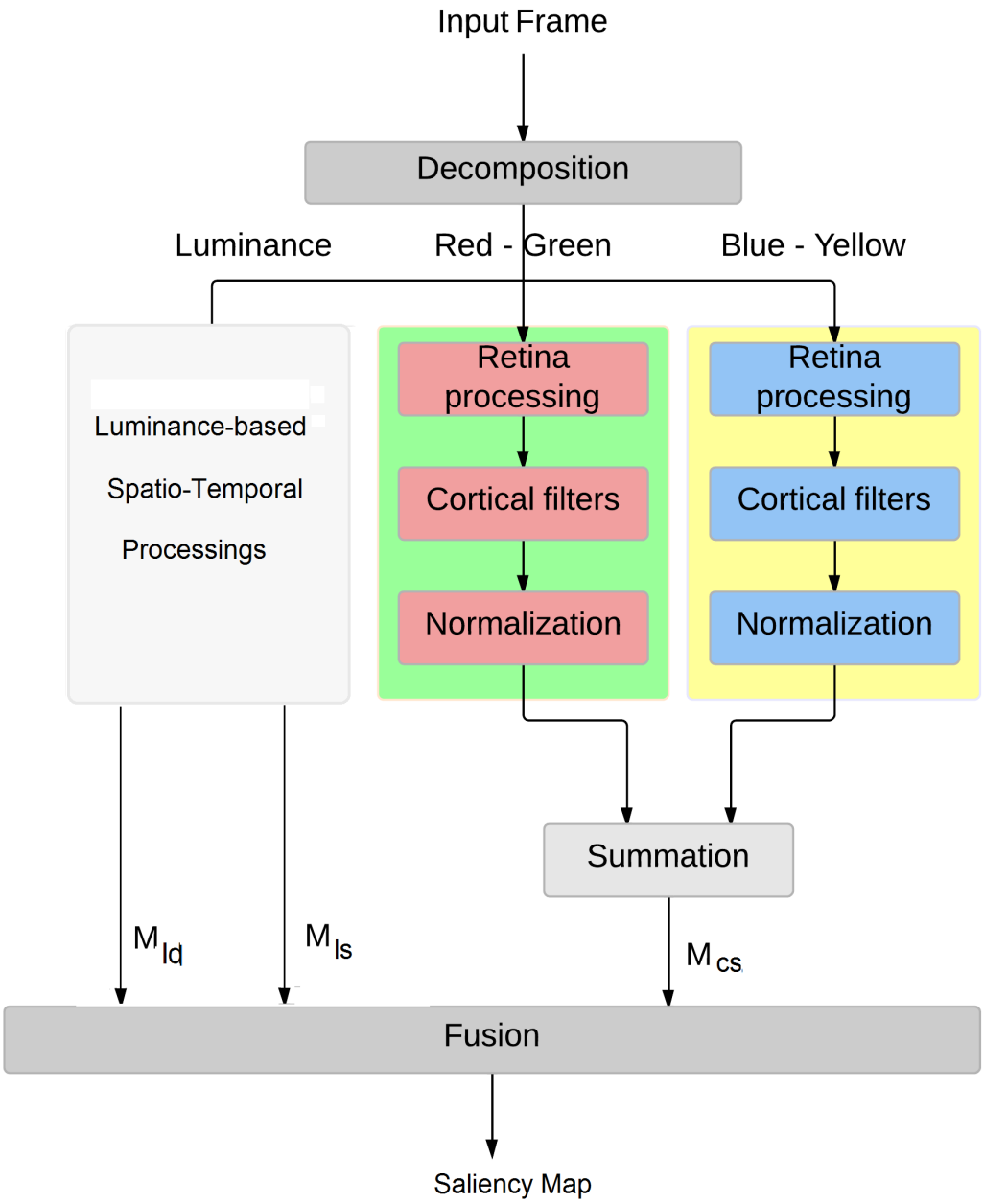


Figure 5.8: The spatio-temporal saliency model. M_{ld} is luminance-dynamic map, M_{ls} and M_{cs} are luminance-static and chrominance-static maps respectively.

applied several optimizations subtending to a real-time solution on multi-GIPSA-lab. We include the parallel adaptation of chrominance pathway to this GIPSA-lab implementation maintaining the real time solution.

The NVIDIA CUDA fast Fourier transform library (cuFFT) is used to perform the complex Fourier transformations. The reductions are carried out using *Thrust* library, an interface to many GIPSA-lab algorithms and data structures. Such as the implementation of luminance-static and luminance-dynamic pathways, chrominance pathway is tested on a 2.67 *GHz quad-core* system with 10 GB of main memory, and *Windows 7* running on it. CUDA v3.0

programming environment on *NVIDIA Geforce GTX 480* is used. The model is available at *GIPSA-lab, real-time visual perception project*.

5.5 Evaluation

A model of visual attention provides saliency maps of a visual scene. These maps enhance the regions of the scene that are more likely to be gazed. To evaluate the performance of a saliency model, the regions enhanced by model must be compared to the zones fixated by observers. Several metrics have been proposed based on this method. These metrics measure the correspondence between salient regions of a stimulus identified by a saliency model and regions fixated by observers when exploring the stimulus. This comparison could be made directly between a set of eye positions and saliency map or between fixation map, created from eye positions [Vel+96; BT06b], and a saliency map. For instance, TC (*Percentage of correct predictions*) [Tor+06; PI08], **Roc!** (*Receiver Operation Characteristic* [TBG05], NSS (*Normalized Scanpath Saliency* [IKN98],), KL divergence, and AUC (*area under the curve*) are some of the best known metrics. More readings on the metrics for evaluating the performance of a saliency map might be found in reviews of [**LeMeur2012**; BI10].

The NSS is one of the most used metrics that we introduce here. We employ this metric to evaluate the contribution of colour information in the proposed model of saliency. We use NSS metric to compare the performance of colour-wise saliency model to the luminance-based saliency model for two conditions of stimuli: colour and grayscale.

5.5.1 NSS metric

A common metric to compare experimental data to computational saliency maps is the Normalized Scanpath Saliency (NSS) [Itt05]. We use this metric to compare colour (C) and grayscale (GS) eye positions to their equivalent saliency maps. To compute NSS, first the saliency maps were normalized to zero mean and unit standard deviation. The NSS value of frame k corresponds to averaged saliency values at the locations of eye positions on the normalized saliency map M as shown in equation (5.5.1):

$$NSS(k) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\sigma_k} (M(X_i) - \mu_k) \tag{5.5.1}$$

where N is the number of the eye positions, $M(X_i)$ is the saliency value of the eye position (X_i), μ_k and σ_k are the mean and standard deviation of the initial saliency map of frame k. A high positive value of NSS indicates that the eye positions are located on the salient regions of the computational saliency map. A NSS value close to zero represents no relation between eye position and computational saliency map, while a high negative value of NSS means that eye positions were not located on the salient regions of computational saliency map.

We use data of eye positions from experiments **A** and **B** to evaluate the performance of the saliency model and to compare luminance-based and luminance-chrominance saliency models. The performance of the model is also compared with one of the reference saliency models, Itti and Koch saliency model [Kla], [IKN98].

Table 5.1: NSS results for Marat et al. model and Itti and Koch saliency model with and without colour features.

		Marat		Itti	
		luminance	luminance +chrominance	luminance	luminance + chrominance
NSS	**C positions**	0.59	1.18	0.91	0.95
	GS positions	0.60	1.17	0.93	0.97

Table 5.2: Timings of sequential (C and MATLAB) and parallel (GIPSA-lab) implementations in ms.

	M_{ls}	M_{cs}	M_{dl}
MATLAB	34.01	22.67	237.03
C	10.73	7.15	31.24
CUDA	0.04	0.03	0.12

5.6 Results

In chapter 4, we presented two eye-tracking experiments, **A** and **B**, that were run to evaluate the influence of colour information in guiding eye movements when exploring video stimuli of different categories. Here, we use the eye position data obtained from both experiments **A** and **B** to evaluate the contribution of colour information in a saliency model.

Evaluation of the model using data of Experiment A First, we studied whether *Marat's luminance-based saliency model* [Mar+09] predicts the eye positions for the two stimulus conditions conditions with equal efficiency. NSS score for colour eye positions is lower than the one for grayscale eye positions(0.88 *vs.* 0.91, $t_{(147)} = 1.5, p = 0.07$). We also compared the C and GS positions to the luminance-chrominance saliency model. In global the model performance in predicting C positions is slightly improved (0.90 *vs.* 0.88, $t_{(147)} = 1.5$, $p = 0.076$). This slight improvement was expected due to the slight differences that were observed when comparing the two datasets of eye positions.

	luminance–based model	colour–wise model
C	0.88	0.90
GS	0.91	0.89

Evaluation of the model using data of Experiment B Second, we studied whether *luminance-based saliency model* [Mar+09] predicts the eye positions in both conditions with equal efficiency. Then we performed NSS analysis, but using the model of saliency with chrominance. As shown in table 5.1 colour information improves significantly the performance of presented model for both C and GS positions ($GS : t(63) = 4.5, p < 0.01, C : t(63) = 4.86, p < 0.01$), while it improves slightly the performance of the model of Itti and Koch [IKN98].

In addition , as presented in table 5.2, GPU implementation of chrominance-static pathway results in a significant speed-up over MATLAB and C implementations.

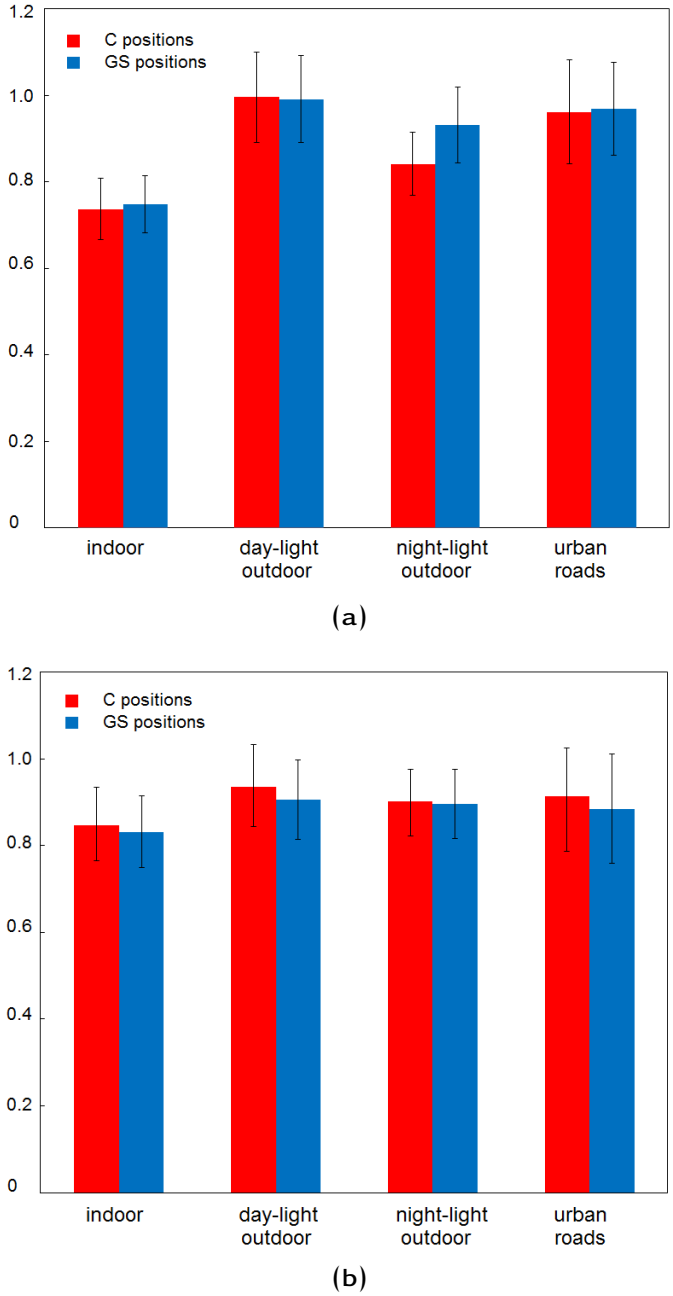
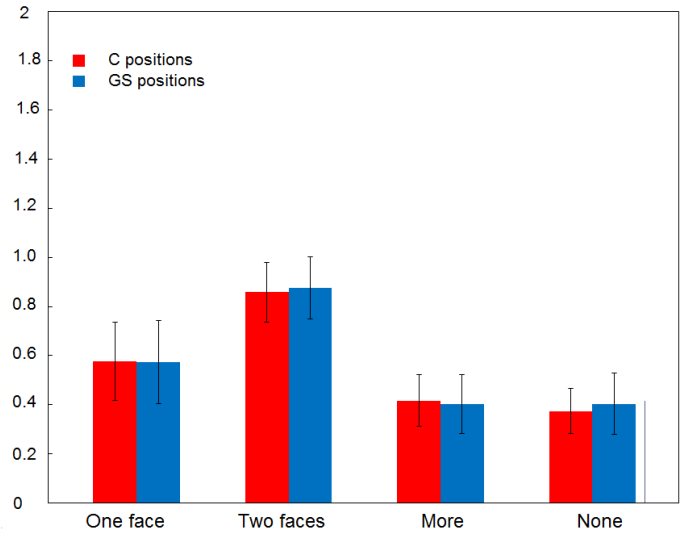
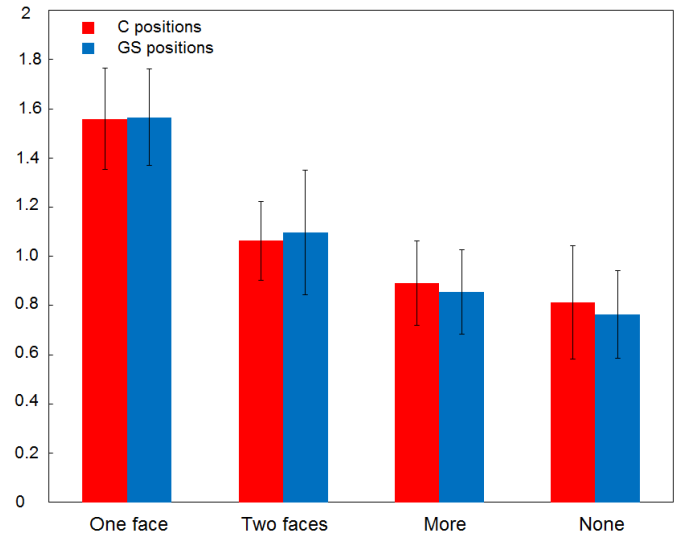


Figure 5.9: mean NSS value per category for stimuli of experiment A, (a) mean NSS values for luminance-based saliency model, (b) mean NSS values for colour-wise saliency model.



(a)



(b)

Figure 5.10: mean NSS value per category for stimuli of experiment B, (a) mean NSS values for luminance-based saliency model, (b) mean NSS values for colour-wise saliency model..

5.7 Conclusion

In the present chapter we have presented luminance-based saliency model of Marat and colleagues [Mar+09]. We have incorporated a chrominance pathway to the model. We have used eye tracking data of experiments A and B 4; these data allows us to validate the proposed saliency model and more specifically to quantify the contribution of colour in the saliency model to predict eye fixations.

Results show that indeed colour information improves significantly the performance of the model in predicting eye positions for both grayscale and colour stimuli only for the stimuli of experiment B 4.2. However, a better prediction power was expected for colour stimuli. This might be due to the fact that the major regions of interest are common in both stimuli conditions, but are better enhanced when employing colour processing steps. But, for the stimuli of experiment A 4.1 we do not observe such improvement.

The incorporation of colour information into the model is not optimized. Because the regions of interest are not always located on coloured zones, but their neighbouring [LMLCB07]. Whether reinforcement of luminance saliency according to the colour information of neighbouring zones can improve the predictive power of saliency model remains to be determined.

6 Conclusions and perspectives

Studies conducted in this thesis focus on colour information and visual attention. We are interested to better understand the influence of colour information on the visual attention, to propose a visual saliency model that optimizes the use of colour information. Throughout the thesis, we concentrate on the question, "How people explore dynamic visual scenes, how the different visual features are modelled to mimic the eye movements of people, in particular, what is the influence of colour information?". To answer these questions we set up eye-tracking experiments to analyse the eye movements of observers.

In this thesis we have used the results of these analysis to determine the factors that must be considered to propose a saliency model that predicts eye movements, and finally, regarding these factors, we have integrated colour information to a luminance-based saliency model.

6.1 Key contributions

In chapter 4, we have evaluated the influence of colour information on the eye movements when free-viewing videos. We have compared the gaze patterns of participants who explored freely colour video stimuli with those who explored grayscale video stimuli.

- ❖ Colour influences gaze positions to some extents. We found that eye positions of observers across videos follow the same patterns for both colour and grayscale stimuli. If a snippet starts with a fixation cross, fixation dispersion among observers is very low and increases as the scene progress. If a snippet starts immediately after another snippet the fixations on the scene onset correspond to regions of interest in the previous scene, resulting in higher fixation dispersion. Dispersion for colour stimuli is slightly higher that grayscale stimuli, specially in the middle period of observation about one second after scene onset.

- ❖ Colour increases the number of the region of interest, especially in the middle period of observation. We identified the region of interest using a clustering method on eye positions. We found that, in colour stimuli, the number of the regions of interest is higher than in grayscale stimuli.

- ❖ Number of the regions of interest increase by time. The clustering of eye position shows that for both colour and grayscale stimuli the number of clusters (regions of interest) increases as the scene progress, which is coherent with higher dispersion among observers.

- ❖ Impact of colour on the eye positions is independent from the category of videos. We studied eye movements according to the category of the video stimuli. The evaluation using different comparison criteria, such as dispersion and number of clusters, shows that eye positions are independent from the categories of video stimuli, no matter the stimulus condition, except for *night-light* category. This is essentially related to the size of the illuminated region of the scene is smaller than other categories.

- ❖ Fixations are shorter in colour face videos. We observe that fixations are longer on videos with faces, but fixations in colour videos are shorter than grayscale videos for face category. However, the durations are shorter for other categories, no matter stimulus condition, when several regions of interest are competing for limited attentional resources.

In chapter 5, we used the observations about the influence of colour information in videos to integrate chrominance saliency to the saliency model of Marat and colleagues [Mar+09].

- ❖ We have proposed a colour-wise bottom-up visual saliency model that predicts eye positions using two pathways based on different types of visual features: static and dynamic. It is an extension of the saliency model proposed previously by [Mar+09]. The model is inspired by the biology of human visual system and proposes a simulation of first steps of the visual processing in human using retina-like filters and cortical-like filters. The original version of the model is based on the luminance information. The static pathway results in a static saliency map that extracts the texture information. The static pathway is improved by adding colour saliency maps to the luminance saliency maps. The dynamic pathway results in a dynamic saliency map that detects the moving objects in the scene. These two saliency maps are combined to provide a saliency map that enhances the regions of visual scene that attract attention in videos.

- ❖ We have evaluated the contribution of colour saliency map in the performance of the model against eye movement data from the psycho-visual experiments. We show that the inclusion of colour features improves the prediction power of the visual saliency model, specially for person-present scenes.

In conclusion, the eye-tracking experiments show a modest influence of colour information on the eye movements. However, incorporation of colour processing steps into a saliency model leads to higher efficiency.

6.2 Perspectives and future works

In view of cited contributions, we can say that the original objectives of this research have been met. We could evaluate the impact of colour information on the eye movements when freely viewing various video stimuli. We have incorporated a colour saliency pathway to the luminance-based model and we have obtained a higher performance. The following presents potential plans for future works, regarding colour videos and saliency model:

Eye tracking experiments

- ❖ We used video databases comprising short video exerts of various lengths to evaluate the influence of colour information on eye movements across time. We observed that the impact of colour information is higher in the middle period of observation. But, our experimental design did not allow us to identify, more precisely, when colour information interferes in guiding eye movements. It is important to set-up psycho-visual experiments to achieve this goal.

- ❖ We have studied the influence of colour information on the eye movements using short length video stimuli, when bottom-up attention is more involved in guiding eye movements. It is important to use longer videos to study how colour information intervene when top-down attention guides eye movements rather than bottom-up attention.

- ❖ Studies have investigated the contribution of colour in visual attention for static images [HPGDG12; FHK08] or dynamic scenes [Ham+15b]. It is interesting to conduct a study to analyse the influence of colour on eye movements for static and dynamic stimuli of the same visual scene.

- ❖ The main objective of the eye-tracking experiments conducted in this thesis was to compare the eye movement data of colour stimuli and grayscale stimuli. We proposed a display-dependent grayscale conversion method to minimize the intensity changes between colour and grayscale stimuli. A perspective is to study different grayscale conversion methods and evaluate the influence of these methods on the eye movements.

Saliency model

- ❖ We have incorporated colour features into a luminance-based saliency model. There are several criteria in literature that measure colourfulness of an image [Fai98; Fai10]. A colour-wise saliency model, requires an efficient method to estimate the colourfulness of input image regarding saliency features. Employing such criteria might simplify a model saliency by discarding input images with low colourful features.

- ❖ Saliency models of attention are compute-intensive. Incorporation of red-green and yellow-blue colour saliency maps increases the computation time. The application needs might elicit a preference to the luminance-based saliency model. A perspective is to improve saliency-preserving grayscale conversion methods. Such grayscale conversion methods might replace colour saliency steps to simplify the model while preserving its performance.

- ❖ One of the main objectives in this thesis was to propose a biologically plausible model of colour saliency. We used a simplified model of retina. A perspective is to approach the model to human visual system by simulating the random mosaic of photoreceptors and spatially variant structure of retina.

- ❖ We integrated a GPU implementation of colour saliency model to the existent GPU-based visual saliency model. The next generation of graphics cards extends the computational capabilities of the hardware. The implemented saliency model could be ported on to the rapidly improving GPU technology.

A

Résumé en français

A.1 Contexte

Face à la l'énorme quantité de l'information visuelle qui nous entoure, notre système visuel a les ressources biologiques et sensorielles limités. Cependant, le système visuel humain (SVH) effectue une perception visuelle plutôt efficace de notre environnement. La perception visuelle correspond à la faculté de système visuel humain dans l'interprétation et l'exploration de l'information visuelle brute, de l'acquisition de l'image par rétine au traitement cortical. Pour faire face à l'énorme quantité d'informations visuelles, notre système visuel est capable de sélectionner l'information la plus pertinente qui parvient à la rétine de l'ensemble des stimuli situé dans le champ visuel. Cette capacité est appelée l'attention visuelle. L'attention visuelle est en corrélation avec les mouvements oculaires. Une séquence des saccades et des fixations apporte une zone particulière de la scène visuelle à la fovéa, où les dispositions sensorielles de l'œil sont concentrées à fin d'effectuer un traitement adéquat de l'emplacement de focus de regarde. Le choix de l'emplacement, qui doit être regardé, implique deux mécanisme de l'attention sélective: un mécanisme inconscient et exogène appelé également l'attention ascendant (bottom-up) et un mécanisme endogène et conscient aussi connu comme l'attention descendant (top-down). L'attention ascendant, qui est stimulée par des caractéristiques de bas niveau des stimuli, permet le traitement de l'information visuelle rapidement et sans impliquant toutes les ressources attentionnelles. L'attention de top-down est consciente, contrôlée, dépendante de La tache de l'observateur et implique la plupart des ressources sensoriels et cognitifs. Modélisation du mécanisme de l'attention visuelle sélective est l'un des domaines de recherche actifs de la vision par ordinateur ainsi que des sciences cognitives. En raison de la très grande complexité de l'attention visuelle due à des interactions et dépendances entre l'attention ascendante et descendant, la modélisation du mécanisme de l'attention visuelle dans son intégralité est peu réalisable avec les technologies existantes. Ainsi, les chercheurs ont tendance à diviser les modèles de l'attention en en deux catégories : les modèles de l'attention ascendante et des modèles de l'attention descendant. À la base des modèles de l'attention il y a des théories comme le "filter model" [Bro58] et la " feature integration th Feature Integration Theory " (FIT) [de TG80]. Cette dernière est l'une des théories les plus cités de l'attention, et divise les processus d'attention en deux étapes: un pré-attentive et une autre ciblée. Selon la FIT, caractéristiques visuelles élémentaires tels que

l'intensité, la couleur et l'orientation sont traitées en parallèle à une phase de pré-attentif, et ensuite combinés pour conduire le centre d'attention. Plus tard en 1985, basée sur la Feature Integration Theory , Koch et Ullman [KU85] ont développé l'un des premiers modèles de calcul de l'attention qui a été inspiré de la biologie du système visuel humain. Pour la première fois le terme de carte de saillance est apparu dans ce travail. Une carte de saillance a été définie comme une représentation visuelle de la scène, dans laquelle les régions les plus intéressants sont améliorées. La FIT et l'architecture de calcul de cette théorie proposée par Koch et Ullman [KU85] ont été la source d'inspiration pour de nombreux autres modèles de calcul de l'attention, tels que le modèle proposé par Itti et ses collègues [IKN98], qui est un modèle de référence dans le domaine des modèles de calcul de l'attention. Plupart de ces modèles, calculent une carte de saillance des stimuli visuels en fonction de leurs caractéristiques de bas niveau, tels que, la couleur, l'intensité, l'orientation, la fréquence, le mouvement, etc. La contribution de ces fonctionnalités pour le déploiement de l'attention a été examinée sur les stimuli synthétiques [WH04]. La couleur en plus d'autres fonctions sont identifié comme les attributs qui guide l'attention lorsque vous effectuez une recherche visuelle, par exemple trouver une barre rouge horizontale entre les barres verticales vertes. Pourtant, la faculté de guidance des caractéristiques de couleur lors de l'exploration des scènes naturelles est questionnée.

A.2 Des défis

Dans cette thèse, nous nous intéressons au rôle de la couleur dans l'attention visuelle, des mouvements oculaires ainsi que dans les modèles. Le premier défi est d'enquêter sur la faculté' de guidance des caractéristiques de couleur dans les stimuli vidéo, en utilisant des expériences oculométrie. La question principale est de savoir si la couleur influences, dans le moins, les mouvements des yeux et la position de focus d'attention lorsque on regarde librement les stimuli vidéos. Il y a aussi plusieurs questions quant à savoir si l'influence de la couleur sur l'attention visuelle est corrélée à la catégorie des stimuli. Est-ce que la couleur guide l'attention Lors d'observation des stimuli vidéo naturels par exemple des paysages? Est-ce que la contribution de la couleur dans l'orientation de l'attention varie entre les scènes artificielles, telles que les routes urbaines et des scènes d'intérieur, et des paysages? Qu'en est-il des scènes sur présentant les visages? IL est montré que les visages dans une scène guident l'attention visuelle rapidement et indépendant de la tâche. Ne diffère l'attribution de l'attention sur les visages sur des stimuli de couleur qu'en niveaux de gris? Le deuxième défi est d'intégrer les résultats des expériences et des évaluations dans un modèle de calcul de l'attention basé sur luminance. Dans cette thèse, le modèle ascendant de l'attention proposée précédemment par Marat et ses collègues [Mar + 09] est amélioré. Le modèle original calcule les cartes de saillance visuelles d'une vidéo par des voies statiques et dynamiques pour des stimuli en niveaux de gris. Nous essayons d'améliorer la performance du modèle à l'aide des informations de couleur.

A.3 Des objectifs

En ce qui concerne les défis décrits ci-dessus, l'objectif de cette thèse est double. D'une part, nous étudions et comparons le comportement des observateurs lors de l'affichage des stimuli vidés couleurs et niveaux de gris. D'autre part, nous aimerions inclure des caractéristiques de couleur à un modèle de calcul de saillance. La première étape consiste à réaliser des expériences oculométrie utilisant stimuli vidéo avec divers contenus. Les expériences nous

permettraient d'identifier divers facteurs liés à l'impact des caractéristiques de couleur sur l'attention visuelle. La deuxième étape consiste à la modélisation informatique et à l'incorporation des caractéristiques de couleur dans un modèle de saillance biologiquement inspiré et moduler ces fonctionnalités selon ces facteurs. Un modèle de saillance sensible à la couleur pourrait être bénéfique pour les applications de vision par ordinateur, robots cognitifs, la reconnaissance d'objet et les dispositifs de contrôle de la qualité.

A.4 Principales contributions

Cette thèse porte sur la contribution de l'information de la couleur dans les mouvements oculaires d'un côté et dans la performance d'un modèle de saillance de l'autre côté. Ces deux objectifs sont atteints grâce à les principales contributions apportées dans cette thèse : Nous identifions l'impact de l'information couleur sur les mouvements des yeux lors de l'observation des stimuli vidéo, en termes de la congruence des positions du regard des observateurs, nombre de régions d'intérêt, la durée de fixation et l'amplitude de saccade en global et aussi en fonction du temps. Nous incorporons une carte couleur de saillance à un modèle existant de saillance basée luminance. Nous évaluons la performance du modèle par rapport aux modèles existants dans la littérature.

Acronyms

ANOVA	analysis of variance
AUC	area under the curve
AVI	audio video interleave
CIE	*commission international d'éclairage*
CSF	contrast sensibility function
CRT	cathod ray tube
CUDA	compute unified device architecture
cuFFT	NVIDIA CUDA fast Fourier transform library
DMAs	distributed associated memories
FCC	federal communication commission
FIT	feature integration theory
FOA	focus of attention
GHz	gigahertz
GIPSA-lab	Grenoble Images Parole Signal Automatique laboratory
GPU	graphics processing unit
HVS	human visual system
HSV	hue saturation value
IOR	inhibition-of-return
IT	inferotemporal cortex
KL	Kullback Leibler
MATLAB	matrix laboratory
mp4	MPEG-4 Part 14
NPL	national physical laboratory
NSS	normalized saliency scanpath
NTSC	national television standard committee
PAL	phase alternating line
SECAM	séquentiel couleur à mémoire
TC	Torralba's percentile criterion
V1	primary visual cortex
WTA	winner-takes-all

Bibliography

[AM12] David Alleysson and David Meary. "Neurogeometry of color vision". In: *Journal Of Physiology Paris* 106 (2012), pp. 284–296 (see p. 56).

[BT06a] R. J. Baddeley and B. W. Tatler. "High frequency edges (but not contrast) predict where we fixate: A Bayesian system identification analysis". In: *Vision Research* 46.18 (2006), pp. 2824–2833 (see pp. 15, 39, 59).

[BE04] R. Bala and N. Eschbach R. York. "Spatial Color-to-Grayscale Transform Preserving Chrominance Edge Information". In: *Proc ISTSIDs 12th Color Imaging Conference*. Vol. 100. 2. 2004, p. 4 (see p. 38).

[BB82] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall Professional Technical Reference, 1982 (see p. 7).

[BI11] F. Baluch and L. Itti. "Mechanisms of top-down attention." In: *Trends Neurosci.* 34.4 (2011), pp. 210–224 (see pp. 8, 59).

[BM05] W. H. A. Beaudot and K. T. Mullen. "Orientation selectivity in luminance and color vision assessed using 2-d bandpass filtered spatial noise." In: *Vision Research* 45(6) (2005), pp. 687–696 (see p. 64).

[BK09] D. M. Beck and S. Kastner. "Top-down and bottom-up mechanisms in biasing competition in the human brain". In: *Vision Res.* 49.10 (2009), pp. 1154 –1165 (see p. 59).

[Ben+10] L. Benedetti, M. Corsini, P. Cignoni, M. Callieri, and R. Scopigno. "Color to gray conversions in the context of stereo matching algorithms". In: *Machine Vision and Applications* 57(2) (2010), pp. 254–348 (see p. 36).

[Bin10] M. Bindemann. "Scene and screen center bias early eye movements in scene viewing". In: *Vision Research* 50.23 (2010), pp. 2577–2587 (see pp. 8, 56).

[BI10] A. Borji and L. Itti. "State-of-the-art in Visual Attention Modeling". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 99 (2010) (see pp. 11, 14, 70).

[Bro58] D. E. Broadbent. *Perception and communication*. Ed. by J. E. Birren and K. W. Schaie. Pergamon Press, 1958 (see pp. 1, 8).

[BT06b] N. D. B. Bruce and J. K. Tsotsos. "Saliency based on information maximisation." In: *Advances in Neural Information ProcessingSystem* 18 (2006), 155162 (see p. 70).

[BP02] E. Bruno and D. Pellerin. "Robust motion estimation using spatial Gabor-like filters". In: *Signal Process.* 82 (2002), pp. 297–309 (see p. 63).

[Bun90] C. Bundesen. "A theory of visual attention". In: *Psychological Review* 97 (1990), pp. 523–547 (see p. 8).

[CI06] R. Carmi and L. Itti. "Visual causes versus correlates of attentional selection in dynamic scenes". In: *Vision Res.* 46.26 (2006), pp. 4333 –4345 (see pp. 8, 40).

[CFK09] M. Cerf, E. P. Frady, and C. Koch. "Faces and text attract gaze independent of the task: Experimental data and computer model." In: *J. Vision* 9.12 (2009), pp. 10.1–15 (see p. 39).

[Cha03] A. Chauvin. "Perception des scènes naturelles: étude et simulation du rôle de l'amplitude, de la phase et de la saillance dans la catégorisation et l'exploration des scènes naturelles". PhD thesis. Université Pierre Mendes-France, Grenoble, 2003 (see p. 13).

[CEY04] C. E. Connor, H. E. Egeth, and S. Yantis. "Visual attention: bottom-up versus top-down". In: *Current Biology* 14 (2004), pp. 850–852 (see p. 59).

[CPDS12] V. Courboulay and M. Perreira Da Silva. "Real-time computational attention model for dynamic scenes analysis: from implementation to evaluation." In: *SPIE Optics, Photonics and Digital Technologies for Multimedia Applications - Visual attention*. Ed. by SPIE. Brussels, Belgium, Apr. 2012, to be published (see p. 13).

[Cou+12] A. Coutrot, N. Guyader, G. Ionescu, and A. Caplier. "Influence of soundtrack on eye movements during video exploration." In: *J. Eye Mov. Res.* 5.4 (2012), pp. 1–10 (see pp. 8, 57).

[DMB10] M. Dorr, T. Martinetz, and E. Barth. "Variability of eye movements when viewing dynamic natural scenes". In: *Journal of Vision* 10.10 (2010), pp. 1–17 (see pp. 8, 15, 59).

[Fai98] M. D. Fairchild. *Color Appearance Models*. ISBN 0-201-63464-3. Addison-Wesley, Reading, MA., 1998 (see p. 77).

[Fai10] M. D. Fairchild. "Color appearance models and complex visual stimuli." In: *Journal of Dentistry* 38 (2010) (see p. 77).

[Fau79] O. D. Faugeras. "Digital color image processing within the framework of a human visual model". In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 27.4 (1979), pp. 380–393 (see p. 27).

[FLMB11] B. Follet, O. Le Meur, and T. Baccino. "New insights on ambient and focal visual fixations using an automatic classification algorithm". In: *iPerception* 2(6) (2011), pp. 592–610 (see pp. 45, 52).

[FHK08] H. P. Frey, C. Honey, and P. König. "Whats color got to do with it? The influence of color on visual attention in different categories". In: *Journal of Vision* 11(3) (2008), pp. 1–15 (see pp. 15, 39, 56, 57, 59, 77).

[Fri05] S. Frintrop. *VOCUS: A Visual Attention System for Object Detection and Goal-directed search*. Vol. 3899 / 2006. Lecture Notes in Artificial Intelligence. Springer Berlin/Heidelberg, 2005 (see pp. 11, 14).

[Fri06] S. Frintrop. "VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search". PhD thesis. Rheinische Friedrich-Wilhelms-Universität Für Informatik and Fraunhofer Institut Für Autonome Intelligente Systeme, 2006 (see pp. 8, 59).

[Geg03] K. R. Gegenfurtner. "Cortical mechanisms of colour vision." In: *Nature Reviews Neuroscience* 4(7) (2003), pp. 563–72 (see pp. 22, 65).

[Goo05] Olsen S. Tumblin J. & Gooch B. Gooch A. "Color2gray: salience-preserving color removal." In: *ACM Transactions on* 24.3 (2005), pp. 1–6 (see p. 38).

[HW71] A. Habibi and P. A. Wintz. "Image Coding by Linear Transformation and Block Quantization". In: *IEEE Transactions on Communication Technology* COM-19 (1971), pp. 50–61 (see p. 28).

[Ham+14] S. Hamel, N. Guyader, D. Pellerin, and D. Houzet. "Color information in a model of saliency". In: *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*. 14785751. IEEE, 2014, pp. 226 –230.

[Ham+15a] S. Hamel, N. Guyader, D. Pellerin, and D. Houzet. "Contribution of color in saliency model for videos". English. In: *Signal, Image and Video Processing* (2015), pp. 1–7. DOI: 10.1007/s11760-015-0765-5.

[Ham+15b] S. Hamel, D. Houzet, D. Pellerin, and N. Guyader. "Does color influence eye movements while exploring videos?" In: *Journal of Eye Movement Research* 8 (2015), pp. 1 –10 (see p. 77).

[HPGDG12] T. Ho-Phuoc, A. Guérin-Dugué, and N. Guyader. "When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements?" In: *Journal of Vision* 12(2) (2012), pp. 1–13 (see pp. 8, 15, 39, 56, 59, 77).

[HPGGD10] T. Ho-Phuoc, N. Guyader, and A. Guérin-Dugué. "A Functional and Statistical Bottom-Up Saliency Model to Reveal the Relative Contributions of Low-Level Visual Guiding Factors". In: *Cognitive Computation* 2(4) (2010), pp. 344–359 (see p. 59).

[HS95] J. E. Hoffman and B. Subramaniam. "The role of visual attention in saccadic eye movements". In: *Perception & Psychophysics* 57 (6) (1995), pp. 787–795 (see p. 8).

[IGGD09] G. Ionescu, N. Guyader, and A. Guérin-Dugué. "SoftEye software". In: *IDDN*. 2009 (see p. 43).

[Itt05] L. Itti. "Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes". In: *Vis. Cogn.* 12.6 (2005), pp. 1093–1123 (see pp. 40, 59, 70).

[Itt02] L. Itti. "Real-Time High-Performance Attention Focusing in Outdoors Color Video Streams". In: *Proc. SPIE Human Vision and Electronic Imaging VII (HVEI'02), San Jose, CA*. Ed. by B. Rogowitz and T. N. Pappas. SPIE Press, 2002, pp. 235–243 (see p. 12).

[IB09] L. Itti and P. Baldi. "Bayesian surprise attracts human attention". In: *Vision Res.* 49 (2009), pp. 1295–1306 (see p. 39).

[IK01] L. Itti and C. Koch. "Computational modelling of visual attention." In: *Nat. Rev. Neurosci.* 2.3 (2001), pp. 194–203 (see p. 11).

[IKN98] L. Itti, C. Koch, and E. Niebur. "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis". In: *IEEE T. Pattern. Anal. Mach. Intell.* 20 (1998), pp. 1254–1259 (see pp. 2, 12, 13, 59, 60, 63, 70, 71).

[Jud30] D. B. Judd. "Reduction of data on mixture of color stimuli". In: *Bureau of Standards J. Research* 4 (1930), pp. 515–548 (see p. 31).

[Kan+09] C. Kanan, M. H. Tong, L. Zhang, and G. W. Cottrell. "SUN: Top-down saliency using natural statistics". In: *Vis. Cogn.* 17.6-7 (2009), pp. 979–1003 (see p. 59).

[KY06] N. Kanwisher and G. Yovel. "The fusiform face area: a cortical region specialized for the perception of faces". In: *Philos. Trans. R. Soc. London, Ser. B* 361.1476 (2006), pp. 2109–2128 (see p. 49).

[KU00] S. Kastner and L. G. Ungerleider. "Mechanisms of visual attention in the human cortex." In: *Annu. Rev. Neurosci.* 23.1 (2000), pp. 315–341 (see p. 8).

[Kla] Klab. *http://www.klab.caltech.edu/ harel/share/gbvs.php* (see p. 70).

[KU85] C. Koch and S. Ullman. "Shifts in selective visual attention: towards the underlying neural circuitry". In: *Hum. Neurobiol.* 4 (1985), pp. 219–227 (see pp. 2, 11).

[KWH82] J. Krauskopf, D. R. Williams, and D. W. Heeley. "Cardinal direction of color space". In: *Vision Research* 22 (1982), pp. 1123–1131 (see pp. 27, 63).

[Lat88] C. R. Latimer. "Eye-movement data: cumulative fixation time and cluster-analysis." In: *Behavior Research Methods, Instruments, & Computers.* 20(5) (1988), pp. 437–470 (see p. 45).

[LM05] O. Le Meur. "Attention selective en visualisation d'images fixes et animées affichées sur ecran: Modèles et evaluation de performance–applications". PhD thesis. Ecoloe polytechnique de l'université de Nantes, 2005 (see p. 65).

[LMLCB07] O. Le Meur, P. Le Callet, and D. Barba. "Predicting visual fixations on video based on low-level visual features". In: *Vision Res.* 47.19 (2007), pp. 2483–2498 (see pp. 13, 59, 74).

[LRT77] J.O. Limb, C.B. Rubinstein, and J.E. Thompson. "Digital Coding of Color Video Signals – A Review". In: *IEEE Transactions on Communications* 25.11 (1977) (see pp. 25, 30–33).

[Mac70] D. L. MacAdam. *Sources of Color Science*. Ed. by MIT pr. Cambridge, Mass., 1970 (see p. 22).

[MLH02] R. Malach, I. Levy, and U. Hasson. "The topography of high-order human object areas." In: *Trends Cogn. Sci.* 6.4 (2002), pp. 176–184 (see p. 50).

[Mar10] S. Marat. "Modèles de saillance visuelle par fusion d'informations sur la luminance, le mouvement et les visages pour la prédiction de mouvements oculaires lors de l'exploration de vidéos." PhD thesis. Université Joseph-Fourier - Grenoble I, 2010 (see p. 13).

[Mar+13] S. Marat, A. Rahman, D. Pellerin, N. Guyader, and D. Houzet. "Improving visual saliency by adding 'face feature map' and 'center bias'". In: *Cogn. Comput.* 5.1 (2013), pp. 63–75 (see p. 56).

[Mar+09] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guérin-Dugué. "Modelling Spatio-Temporal Saliency to Predict Gaze Direction for Short Videos". In: *Int. J. Comput. Vision* 82 (2009), pp. 231–243 (see pp. 2, 39, 40, 44, 59, 60, 62, 67, 71, 74, 76).

[MMP05] M. Martelli, N. J. Majaj, and D.G. Pelli. "Are faces processed like words? A diagnostic test for recognition by parts." In: *J. Vision* 5.1 (2005), pp. 58–70 (see p. 39).

[MK11] S. McMains and S. Kastner. "Interactions of Top-Down and Bottom-Up Mechanisms in Human Visual Cortex". In: *J. Neurosci.* 31.2 (2011), pp. 587–597. eprint: http://www.jneurosci.org/content/31/2/587.full.pdf+html (see p. 59).

[Mil+94] R. Milanese, H. Wechsler, S. Gill, J.-M. Bost, and T. Pun. "Integration of bottom-up and top-down cues for visual attention using non-linear relaxation". In: *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR '94., 1994 IEEE Computer Society Conference on*. 1994, pp. 781 –785. DOI: 10.1109/CVPR.1994.323898 (see pp. 11, 12).

[Mit+11] P. K. Mital, T. J. Smith, R. L. Hill, and J. M. Henderson. "Clustering of Gaze During Dynamic Scene Viewing is Predicted by Motion". In: *Cognitive Computation* 3(1) (2011), pp. 5–24 (see pp. 8, 39).

[Nei67] U. Neisser. *Cognitive Psychology*. Appleton-century-Crofts, New York, 1967 (see p. 8).

[PLN02] D. Parkhurst, K. Law, and E. Niebur. "Modeling the role of salience in the allocation of overt visaul attention". In: *Vision Res.* 42 (2002), pp. 107–123 (see p. 8).

[PI08] R. J. Peters and L. Itti. "Applying computational tools to predict gaze direction in interactive visual environments". In: *ACM T. Appl. Percept.* 5.2 (2 2008), pp. 1–9 (see p. 70).

[Pet+05] R.J. Peters, A. Iyer, L. Itti, and C. Koch. "Components of bottom-up gaze allocation in natural images." In: *Vision Res.* 45.18 (2005), pp. 2397–2416 (see p. 8).

[PS00] C. M. Privitera and L. W. Stark. "Algorithms for defining visual regions-of-interest: comparison with eye fixations." In: *IEEE Trans Pattern Anal Mach Intell* 22(9) (2000), pp. 970 –82 (see p. 45).

[PJ01] D. Purves and Augustine G. J. *Types of Eye Movements and Their Functions*. Ed. by editors. Fitzpatrick D et al. Sunderland (MA): Sinauer Associates. http://www.ncbi.nlm.nih.gov/books/NBK10991/: Neuroscience. 2nd edition., 2001 (see p. 7).

[Rah13] A. Rahman. "Face perception in videos: Contributions to a visual saliency model and its implementation on GPUs". PhD thesis. Universit Grenoble, 2013 (see pp. 60, 67).

[RHP11] A. Rahman, D. Houzet, and D. Pellerin. "Visual Saliency Model on Multi-GPU". In: *GPU Computing Gems Emerald Edition*. Elsevier, 2011, pp. 451–472 (see p. 68).

[RPH14] A. Rahman, D. Pellerin, and D. Houzet. "Influence of number, location and size of faces on gaze in video". In: *Journal of Eye Movement Research* 7(2) (2014), pp. 1–11 (see p. 50).

[Riz+87] G. Rizzolatti, L. Riggio, I. Dascola, and C. Umiltá. "Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention". In: *Neuropsychologia* 25 (1987), pp. 31–40 (see p. 8).

[SG00] Dario Salvucci and Joseph H Goldberg. "Identifying fixations and saccades in eye-tracking protocols". In: *Proceedings of the symposium on Eye tracking research applications* 469.1 (2000). Ed. by AEditor Duchowski, pp. 71–78 (see p. 44).

[SD04] A. Santella and D. DeCarlo. "Robust Clustering of Eye Movement Recordings for Quantification of Visual Interest". In: *Eye Tracking Research and Applications (ETRA) Symposium*. 2004 (see pp. 8, 45).

[Sch04] S.H. Schwartz. *Visual perception: a clinical orientation*. McGraw-Hill, Health Pub. Division, 2004 (see p. 8).

[Sha03] G. Sharma. *Digital Color Imaging Hand book*. Xerox Corporation, Webster, NY: CRC press LLC, 2003 (see pp. 25, 34).

[Tat07] B. W. Tatler. "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions". In: *J. Vision* 7 (2007), pp. 4.1–17 (see p. 56).

[TBG05] B. W. Tatler, R. J. Baddeley, and I. D. Gilchrist. "Visual attention correlates of fixation selection: Effects of scale and time". In: *Vision Res.* 45 (2005), pp. 643–659 (see p. 70).

[TV08] B. W. Tatler and B. T. Vincent. "Systematic tendencies in scene viewing". In: *J. Eye Mov. Res.* 2(2):5 (2008), pp. 1–18 (see p. 8).

[Tor+06] A. Torralba, A. Oliva, M. S. Castelhano, and J. M. Henderson. "Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search." In: *Psychol. Rev.* 113.4 (2006), pp. 766–786 (see pp. 44, 50, 70).

[TG80] A. M. Treisman and G. Gelade. "A feature integration theory of attention." In: *Cognitive Psychol.* 12 (1980), pp. 97–136 (see pp. 1, 9–11, 59).

[TFMB04] Alain Trémeau, Christine Fernandez-Maloigne, and Pierre Bonton. *Image numérique couleur, de l'acquisition au traitement*. DUNOD, 2004 (see pp. 23, 24, 27, 28, 63).

[Tso11] J.K. Tsotsos. *A Computational Perspective on Visual Attention*. MIT Press, 2011 (see p. 11).

[Van00] N. Vandenbroucke. "Segmentation dimages couleur par classification de pixels dans des espaces dattributs colorimétriques adaptés. Application à lanalyse dimages de football." PhD thesis. 2000 (see pp. 18, 19, 27, 28).

[Vel+96] B. M. Velichkovsky, M. Pomplum, J. Rieser, and H. J. Ritter. *Attention and communication: Eye-movement-based research paradigms.* Ed. by Amsterdam: Elsevier. 1996 (see p. 70).

[Win51] W.T. Wintringham. "Color Television and Colorimetry". In: *Proceedings IRE* 30 (1951), pp. 141–164 (see p. 26).

[WG97] J. Wolfe and G. Gancarz. "Guided Search 3.0: A model of visual search catches up with Jay Enoch 40 years later". In: *Basic and Clinical Applications of Vision Science*. Ed. by V Lakshminarayanan. Norwell, MA: Kluwer Academic Publishers, 1997, pp. 189–192 (see p. 9).

[Wol94] J. M. Wolfe. "Guided Search 2.0: A revised model of visual search". In: *Psychon. B. Rev.* 1.2 (1994), pp. 202–238 (see p. 9).

[WCF89] J. M. Wolfe, K. R. Cave, and S. L. Franzel. "Guided search: an alternative to the feature integration model for visual search". In: *J. Exp. Psychol. Human* 15.3 (1989), pp. 419–433 (see pp. 9, 10, 59).

[WH04] J. M. Wolfe and T. S. Horowitz. "What attributes guide the deployment of visual attention and how do they do it?" In: *Nat. Rev. Neurosci.* 5 (2004), pp. 1–7 (see pp. 2, 39).

[Yar67] A. L. Yarbus. *Eye movements and vision*. Ed. by EnglishEditor Trans By L A Riggs. Vol. chapter VI. Plenum Press, 1967, p. 222 (see p. 7).

[You02] T. Young. “On the Theory of Light and Colors”. In: *Philos. Trans. R. Soc. London* 92 (1802), pp. 20–71 (see p. 27).