ANNÉE 2016

UNIVERSITÉ DE RENNES 1

UNIVERSITE BRETAGNE LOIRE

THÈSE / UNIVERSITÉ DE RENNES 1

sous le sceau de l'Université Bretagne Loire

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Informatique

Ecole doctorale Matisse

présentée par

Christel Chamaret

préparée à l'IRISA
(Institut de recherches en informatique et systèmes aléatoires)
et Technicolor

Color Harmony: experimental and computational modeling.

Thèse soutenue à Rennes le 28 Avril 2016

devant le jury composé de :

Pr Alain Trémeau
Professeur, Université de Saint-Etienne / rapporteur

Dr Vincent Courboulay
Maître de conférences HDR, Université de La Rochelle / rapporteur

Pr Patrick Le Callet
Professeur, Université de Nantes / examinateur

Dr Frederic Devinck
Maître de conférences HDR, Université de Rennes 2 / examinateur

Pr Luce Morin
Professeur, INSA Rennes / examinateur

Dr Olivier Le Meur
Maître de conférences HDR, Université de Rennes 1 / directeur de thèse

Abstract

While the consumption of digital media exploded in the last decade, consequent improvements happened in the area of medical imaging, leading to a better understanding of vision mechanisms. More than ever, making aesthetic pictures quickly - with or without artistic expertise - is a research topic. Different axis of investigations remain possible: high resolution, high dynamic range or wide color gamut. Additionally to these objective image properties, more perceptual and artistic insights could be of benefit to any user manipulating pictures.

In such context, this thesis deals with the topic of ***Color Harmony***. The literature related to this topic is limited, but involves many different scientific areas: color science, image processing, psychology, biology and so on. The validity of collected data is questionable due to their limitation to two- or three-colors patches. The models extrapolated from these data remain non-exploitable on natural pictures. Other models depicting rules or areas on color wheel lack scientific guidelines for their utilization. Nonetheless, some algorithms employing color harmony theory and models as a core concept showed up in the literature, but suffered from being quantitatively tested and validated. This is typically due to the deficiency of available ground truth.

In this thesis, two views are put in perspective in order to respond to the previous statements: an experimental and a computational approaches. The conducted experiment allowed observing some effects with an eye-tracking protocol, never applied before with a task on color harmony assessment. From the collected data of our experimental work, we designed a method to generate a ground truth, which would serve to the validation of the two proposed computational methods.

First, we improved an existing architecture for automatic color harmonization and demonstrated exhaustively the benefit of our approach. As a second computational contribution, a novel quality metric is introduced that integrates the concepts of visual masking and color harmony. Thus, we may predict which areas would be perceived harmonious regarding its neighborhood and then the potential masking effects. As a last contribution, two editing tools made accessible the color harmony theory through a hidden formulation of it and a user-friendly and intuitive interface.

Résumé en Français

Contexte et Motivation

La couleur est un élément dominant de la vision humaine. Excluant la pathologie liée au daltonisme, la plupart des êtres humains - grâce à leur système visuel - sont capables de décoder et analyser leur environnement en interprétant les couleurs et leurs nuances. Ainsi, depuis le plus jeune âge, la connaissance des différentes couleurs composant notre ecosystème est transférée et la conscience d'un monde coloré est acquise.

En même temps que les premiers philosophes grecs commencèrent à se questionner sur leur environnement et ses fonctionnalités, l'association des couleurs et leur signification sont devenus un sujet majeur. Parce qu'il est intimement lié au domaine artistique, différentes communautés ont étudié l'association des couleurs, plus tard dénommée l'**Harmonie des couleurs**. Ainsi, des physiciens, peintres, designers, publicitaires, psychologues etc montrèrent un intérêt à la compréhension de l'harmonie des couleurs.

En 1704, Isaac Newton mis en évidence les propriétés circulaires des différentes longueurs d'onde et proposa un aménagemennt des couleurs sous forme de roue [181]. Consécutif à cette découverte, plusieurs personnages influents définirent leur propre roue des couleurs qui amenèrent à une caractérisation, parfois différente, de ce que sont les meilleurs combinaisons ou associations de couleurs [52, 177, 110, 26, 94]. C'est le début de la conceptualisation de l'harmonie des couleurs. Alors que le rôle et l'influence des artistes restèrent prépondérants aux 18^e et 19^e siècle, ce sujet fût soutenu plutôt par la communauté de *color science* au 20^e siècle [267].

A partir du milieu du 20^e siècle, la tendance était de collecter des données au travers de questionnaires reflettant l'opinion des participants à propos de la *bonne* association des couleurs. Ainsi, des protocoles expérimentaux, s'étendant parfois sur des années [168], ont été mis en oeuvre. Une fois collectées, ces données ont permis de construire une représentation empirique de l'harmonie des couleurs, malgré les limites évidentes de ce genre d'approche: diversité culturelle, nombre de stimuli, temps de collecte ou d'expérience... En parallèle, les techniques d'imagerie médicale se sont améliorées afin de mieux comprendre les mécanismes du cerveau et plus précisément le comportement du système visuel humain. Cependant, à la fin du 20e siècle, nous étions encore incapables de

mesurer et d'expliquer la notion d'harmonie des couleurs. Dans le même sens, les psychologues défendirent la théorie de l'idiosyncrasie rendant le problème beaucoup plus complexe et indéterminé. Ils insistèrent sur le fait que les différences culturelles, préférences personnelles, l'époque, l'humeur etc avaient un impact fort sur le jugement porté à l'harmonie des couleurs [193].

A partir du 21^{eme} siècle, le sujet revint au devant de la scène, de part l'expansion de nouvelles technologies. Un nouveau genre d'artiste est apparu dans les communautés de web design, photographie numérique, *computer graphic*. En conséquence, ils ont eu besoin d'outils efficaces et faciles à manipuler afin d'exploiter ces nouvelles opportunités numériques. De plus, internet et les réseaux sociaux ont facilité le déploiement d'expériences à large échelle dont les données sont collectées à la fois rapidement et à travers le monde.

Dans ce contexte, nous avons identifié une opportunité de contribuer à ce monde artistique émergeant en étudiant le thème de l'harmonie des couleurs sous une nouvelle perspective.

Une approche interdisciplinaire

Nous avons décidé d'aborder ce thème avec une approche interdisciplinaire telle que cela a été évoqué dans l'historique précédent. Ainsi, cette vue interdisciplinaire a été séparée en deux approches qui vont se rejoindre ensuite: des algorithmes et modèles numériques ont été conçus et implémentés en prenant en compte les observations et vérité terrain générés lors de notre approche expérimentale. Pendant ce travail de recherche, nous avons gardé à l'esprit les limitations existantes des outils d'édition d'image et proposer ainsi des interfaces et usage qui exploitent nos algorithmes et modèles de façon transparente pour l'utilisateur.

L'approache expérimentale

Dans cette partie, nous avons choisi de mettre en place deux protocoles distincts, jamais explorés dans le cadre de l'harmonie des couleurs.

Dans un premier temps, nous avons abordé cette problèmatique expérimentale **localement** en questionnant le rôle du système visuel humain. Il semblait raisonnable de faire l'hypothèse que le système visuel humain est partie prenante dans l'évaluation de l'harmonie des couleurs. Ainsi, comme l'attention visuelle est principalement mesurée à l'aide d'oculomètre, nous proposons de caractériser l'harmonie des couleurs au travers des mouvements oculaires mesurés avec un protocole faisant intervenir une tâche. En d'autres termes, ce qui attire l'attention va vraissemblablement déterminer ou influencer notre opinion à propos de la tâche considérée [275, 64, 248, 88].

Après avoir conclus sur les effets évalués à l'aide de mouvements oculaires, nous exploitons ces données collectées afin de concevoir une base de données qui servira à l'évaluation des modèles étudiés.

Dans un second temps, nous examinons le concept d'harmonie des couleurs **globalement**. Dans ce but, nous avons mis en place un protocole de type comparaison par paires, où il était demandé aux participants d'exprimer leur choix sur l'image la plus harmonieuse lors de la présentation de deux images côte à côte. Ensuite, toutes ces paires annotées sont classées sur une échelle commune d'harmonie, permettant ainsi de comparer tous les stimulis entre eux. Suivant la même méthodologie que dans l'étude locale de l'harmonie des couleurs, ces données collectées avaient deux objectifs: l'évaluation du facteur d'harmonie des couleurs (la cohérence entre les annotateurs, le rôle de la distribution couleur des stimulis...) et la constitution d'une vérité terrain pour l'évaluation des méthodes proposées.

L'approche algorithmique

La seconde approche développée dans ce mémoire concerne la conception d'algorithmes et/ou modèles. Nous avons conçu et développé principalement deux méthodes différentes dont les contributions se situent à différents niveaux:

1. un algorithme de traitement d'image dont le but est la recolorization automatique d'images,

2. un modèle perceptuel d'estimation de la qualité d'image;

tous deux basés sur les préceptes d'harmonie des couleurs.

En suivant l'hypothèse que le système visuel humain joue un rôle dans l'évaluation de l'harmonie des couleurs, un modèle d'attention visuel a été intégré à dessein dans la première architecture proposée dans l'état-de-l-art [54] qui harmonise automatiquement les couleurs d'une image. D'autres contributions algorithmiques, par exemple la fonction de réajustement des couleurs, la segmentation des zones homogènes, la fonction de coût pour le choix du modèle d'harmonie, ont permis d'améliorer significativement les résultats de l'état-de-l'art. Afin de démontrer ce gain, nous avons proposé une évaluation exhaustive des différentes étapes contituant l'architecture de cet algorithme. En plus de cela, nous avons démontré objectivement en utilisant le classement de l'expérience paire-à-paire le bénéfice de notre méthode.

La deuxième contribution algorithmique est une métrique de qualité perceptuelle prenant en compte les préceptes d'harmonie des couleurs. D'après notre étude de l'état-de-l'art, il n'y a pas eu de tentatives antérieures à l'élaboration d'une telle métrique incluant le concept d'harmonie des couleurs dans des architectures imitant le système visuel humain. Nous avons proposé une évaluation préliminaire de cette métrique en nous appuyant sur des cartes créées à partir des données oculaires collectées lors de nos expériences.

En résumé, ces deux contributions algorithmiques visent à ingérer des notions de perception visuelle dans les théories existantes d'harmonie des couleurs.

Application dans des outils d'édition

Les contributions proposées ne se limitent pas à une étude expérimentale et des méthodes algorithmiques. Un effort supplémentaire a été apporté afin d'offrir un cadre applicatif concret à ces algorithmes. Ainsi, ils ont été mis en oeuvre dans un contexte d'édition d'image. L'idée sous-jacente de ces outils d'édition est que ceux-ci bénéficient à tout utilisateur, quelque soit son expertise en traitement d'image, en harmonie des couleurs et son sens artistique. Plus généralement, nous souhaitons rendre accessible les théories et règles sur l'harmonie des couleurs à travers des interfaces intuitives et simples.

Dans cet esprit, deux outils d'édition sont décrits: le premier simplifie et guide l'utilisateur lors d'une tâche de retouche des couleurs par zone homogène; le second propose de biaiser ou influencer le traitement d'harmonisation automatique des couleurs à partir d'une base de données définies par l'utilisateur. Le premier outil affiche les cartes perceptuelles des zones les moins harmonieuses de l'image, ce qui permet à l'utilisateur de retoucher en conséquence la zone considérée à l'aide de la palette de couleurs proposée (garantissant une amélioration de l'harmonie globale). Le deuxième outil bénéficie des améliorations que nous avons apportées à l'algorithme d'harmonisation des couleurs et permet à l'utilisateur d'homogénéiser le look de toutes ses images tout en garantissant l'harmonie de son image.

Résultats et conclusion

A partir des expériences menées sur l'harmonie des couleurs, plusieurs hypothèses sont validées.

L'utilisation d'un protocole de tâche avec un oculomètre pour mesurer l'harmonie des couleurs a conduit à une cohérence inter-observateur acceptable (confirmée par plusieurs métriques).

Il y a une influence non négligeable de la distribution couleur pour l'évaluation de l'harmonie. Les images avec une large variété de couleurs se situent au bas de l'échelle d'harmonie. Plutôt que la diversité (nombre de teintes différentes présentes), l'agencement et l'aggregation spatiale dans l'image semblent être le facteur influant.

Les données collectées à l'aide de l'expérience d'*eye-tracking* (à savoir les mouvements oculaires sur les images et ceux sur la roue des teintes) nous ont permis de créer une vérité terrain, c'est-à-dire des cartes avec les zones non-harmonieuses.

Nous concluons que le concept d'harmonie des couleurs est bien compris par les observateurs et homogène entre les utilisateurs. Ainsi, nous sommes convaincus que le concept est assez universel pour envisager la conception de modèles numériques qui pourront prédire, jusqu'à un certain degré, un comportement humain moyen.

Dans la partie analytique, un algorithme d'harmonisation automatique des couleurs, a été développé et validé. S'appuyant sur le travail de Cohen-Or *et*

al. [54], cette méthode propose trois contributions principales dont le but est de réduire les distortions, d'obtenir un résultat visuellement cohérent et de favoriser la fidélité au contenu original.

Premièrement, une fonction de coût mesurant la distance statistique entre l'histogramme des teintes et tous les *templates* d'harmonie possibles a été changé et utilisé conjointement avec l'information de saillance dans l'image. Par conséquent, le template d'harmonie finalement sélectionné est mieux appareillé à la distribution couleur de l'image originale.

Deuxièmement, une nouvelle fonction permet de moins contracter les teintes lors celles-ci sont réajustées dans les zones dites harmonieuses du template. Ainsi, une plus grande variété de teintes est préservée comparé à l'utilisation de la fonction originale.

Troisièmement, l'algorithme de segmentation couleur que nous utilisons permet de générer une segmentation fine des zones avec de nombreux modes, ne nécessitant pas de paramétrage *adhoc* à l'image considérée.

Egalement dans ce chapitre de la thèse, nous avons tenté de convaincre que cet algorithme obtient de bonnes performances en investiguant différentes pistes de validation. Nous avons fourni des résultats qualitatifs qui permettent de visuellement apprécier l'apport de chaque étape de l'algorithme et de traiter des cas particuliers. Cela nous a permis de résoudre les problèmes levés lors de l'exposition des limitations de ce genre d'algorithme fournie en début de chapitre. Une contribution majeure dans l'évaluation de ce genre d'algorithme réside dans l'introduction du protocole d'annotation par paires. En effet, nous avons démontré que les images harmonisées avec notre algorithme étaient objectivement annotées plus harmonieuses que leur version originale sur une échelle commune d'harmonie (19 sur 23 stimuli).

Dans un autre chapitre de la partie algorithmique, une métrique de qualité perceptuelle guidée par l'harmonie a été conçue et proposée comme une nouvelle *caractéristique* ou *propriété* pour évaluer la qualité des images. Elle s'appuye à la fois sur les effets connus de masquage perceptuel qui sont des propriétés importantes du système visuel humain et sur les *templates* d'harmonie couleur. L'intégration de ces deux concepts permet de générer, en sortie du modèle, une carte perceptuelle de zones non harmonieuses et un score associé par image. Les résultats qualitatifs montrent que les cartes d'harmonie reflètent la perception des utilisateurs. En effet, l'impact de certaines zones initialement annotées non harmonieuses est minimisé par la métrique lorsque l'on en prend en compte les effets de masquage de l'oeil humain. L'analyse quantitative a révélé une corrélation significative entre les sorties de ce modèle et deux vérité terrain issues des expériences menées dans l'approche expérimentale.

Finalement, dans la partie algorithmique, nous avons introduit deux outils d'édition d'image. Ceux-ci ont rendu accessible aux utilisateurs la théorie et les règles d'harmonie des couleurs au travers d'une formulation intuitive et cachée de ce concept, tout en laissant une part importante de choix et de créativité à l'utilisateur.

Ceci nous permet d'examiner le concept d'harmonie des couleurs au travers d'un champ de recherche large: de l'expérimental vers des outils applicatifs.

Publications

This thesis reflects the work published in five papers during international conferences and workshops. Patents have also been filed and are currently examined by patent offices.

Peer-reviewed international conferences and workshops

- C. Chamaret, "Color impact in visual attention deployment considering emotional images", *Proc. SPIE. 8291, Human Vision and Electronic Imaging XVII* 82911T (February 9, 2012).
- C. Chamaret, F. Urban, J. Lepinel, "Creating experimental color harmony map", *Proc. SPIE 9014, Human Vision and Electronic Imaging XIX*, 901410 (25 February 2014).
- Y. Baveye, F. Urban, C. Chamaret, V. Demoulin, P. Hellier, "Saliency-Guided Consistent Color Harmonization", *Computational Color Imaging Worshop, Lecture Notes in Computer Science*, Springer p.105-118, 2012.
- C. Chamaret, F. Urban, "No-reference Harmony-Guided Quality Assessment", *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2013 IEEE Conference on , vol., no., pp.961-967, 23-28 June 2013.
- C. Chamaret, F. Urban, and L. Oisel, "Harmony-guided image editing", in *IEEE International Conference on Image Processing 2014 (ICIP 2014)*, Paris, France, Oct. 2014, pp. 2176-2178.

Remerciements

Cette page n'est sans doute pas la plus facile à écrire. Aussi sinueux fût-il, le chemin de la thèse en formation continue (ou VAE) a heureusement une fin et de nombreux acteurs m'ont aidé et ont participé à cet épisode important de ma vie. Je les en remercie aujourd'hui.

Tout d'abord, je remercie les membres de mon jury d'avoir accepté, de rapporter, Messieurs Alain Trémeau et Vincent Courboulay, sur ma thèse et de participer à mon jury, Messieurs Patrick Le Callet, Frederic Devinck et Madame Luce Morin.

Je salue aussi l'encadrement d'Olivier Le Meur, mon directeur de thèse. Merci pour tes retours constructifs, ta patience et tes coups de pouce aux moments cruciaux. Merci d'avoir accepté d'encadrer ma thèse en VAE, malgré les contraintes.

Un petit mot de remerciement à Technicolor ainsi que les membres du bureau de formation continue et plus particulièrement Florence Aubrée, qui m'a guidé avec ses conseils pertinents dans la rédaction du rapport de VAE sur mon parcours et mes compétences.

Une pensée pour mes ex-étudiants et collègues, qui ont contribué à mes travaux mais également à me soutenir, m'encourager, me montrer la voix: Yoann, Fabrice, Lionel (et Simon), Philippe G., Pierre, Philippe B., tous les membres du TA, Patrick et Valérie.

Enfin, il me reste à aborder la dimension personnelle, qui n'en est pas la moins chargée émotionnellement. Cette thèse, je l'ai fait en partie pour mes enfants, Glen et Olivia; pour leur montrer la valeur et la satisfaction liée au travail. Même s'ils sont encore jeunes, j'espère qu'ils comprendront mieux mes motivations plus tard. Merci à Julien et son soutien indéfectible: ça y est tu peux souffler, c'est fini ! Je ne serais pas arrivée au bout sans ton soutien logistique, mais surtout moral.

Une pensée pour mon père, qui a failli être encore plus fier de ses filles; le temps et le destin sont des données impossibles à contrôler. Mais je n'ai pas de regret, car tu as été là pour d'autres évènements beaucoup plus importants dans ma vie.

Même si c'est moi la grande soeur, je l'aurais passée après toi cette *** thèse, tendresse à Cécile (Seb. et Romane). C'est plutôt toi qui m'as toujours tiré vers le haut.

Finalement, j'aimerais remercier ma mère. Cela va te surprendre, mais c'est

toi qui fus mon modèle dans cette épopée. J'ai longtemps gardé l'image de toi qui avait repris tes études à plus de 30 ans avec tes deux enfants à charge, pour donner un vrai tournant et du souffle à ta carrière. Cette peine et cet effort, je les ai vus, enregistrés et reproduits d'une certaine façon. Ce fut le moteur de ce travail.

Contents

List of Figures

List of Tables

Notations

1D, 2D	1, 2 Dimensions
AOI	Area Of Interest
AUC	Area Under the ROC Curve
CC	Correlation Coefficient
CDF	Cohen-Daubechies-Feauveau
CR	Cornea reflex
CSF	Contrast Sensitivity Function
DWT	Discrete Wavelet Transform
EOG	Electro-OculoGraphy
FIT	Feature Integration Theory
FN	False Negative
FP	False Positiive
FPR	False Positive Rate
FV, H conditions	Free Viewing, Harmony conditions
HA	High Agreement
HSV	Hue, Saturation, Value
HVS	Human Visual System
IR	Infra Red
KLD	Kullback-Leibler Divergence
KS test	Kolmogorov-Smirnov test
LA	Low Agreement
LGN	Lateral Geniculate Nucleus
NSS	Normalized Scanpath Saliency
RED	Remote Eyetracking Device
RMS	Root Mean Square
ROC	Receiver Operating Characteristic
TN	True Negative
TP	True Positive
TPR	True Positive Rate
WC	Wheel Center

To Glen and Olivia

Chapter 1

General Introduction

This thesis deals with the concept of ***Color Harmony***. What a satisfying feeling to choose after several years the exploration of such an elegant notion.

Color is my day-long obsession, joy and torment.

Claude Monet

1.1 Context and Problem

Color is a dominant feature of human vision. Apart from the well-studied case of color blindness, most humans - thanks to their visual system - are able to decode and analyze their surrounding environment by interpreting the colors and its nuances. While some vegetables or animals develop camouflage properties to hide from other predators, humans are instinctively sensitive to color contrasts [238]. From the earliest days, the knowledge of the different colors composing our surrounding is transferred to virgin minds and the awareness of a colored world is acquired. For example, it is remarkably common that the adjectives related to colors are present in the first hundred words pronounced by a child [271].

Since the first Greek philosophers started to question their human environment and its functions, color association and their connotative meaning became a major topic. Since it relates to the aesthetic and art domains, many fields investigated the study of color association, being later referenced as *Color Harmony*. Hence, physicians, painters, chemists, color theorists/scientists, designers, advertisers and so on showed an interest in the understanding of color harmony.

In 1704, Isaac Newton highlighted the circular properties of wavelengths and proposed the wheel arrangement of colors [181]. Following such precursory work, many key figures designed their own color wheels, leading to different characterization of what are the best color combinations or associations [52, 177, 110, 26, 94]. This formed the beginning of the conceptualization of color

harmony. While the role and influence of artists remain fundamental in the 18^{th} and 19^{th} centuries, this topic was more supported by the color science community in the 20^{th} century [267].

From the middle of the 20^{th} century, the trend was to collect data through questionnaires about users' opinion on the *right* color associations. Thus, experimental protocols, sometimes lasting years [168], were set up and the collected data built an empirical representation of color harmony, despite the potential limitations of such approach (cultural diversity, number of stimuli, processing time and so on). In the meantime, medical imaging improved to better understand the brain mechanisms and more precisely the behavior of the human visual system. However, at the end of the 20^{th} century, we were still far away from measuring and explaining the notion of color harmony. Moreover, psychologists strongly supported the theory that cultural differences, preferences, time, mood and other factors had a clear impact on color harmony assessment [193]. They argue that such concept should be carefully measured while controlling the preceding factors.

Since the beginning of the 21^{st} century, the cause regain attention, due to the expansion of new computer-aided technology. New kind of artists appeared in the digital world of web design, photography, computer graphics and so on. As a matter of fact, they needed efficient and user-friendly tools to exploit the digital opportunities. In addition, the internet and social networks made possible large scale experiments with data collected quickly and from many places. Digital consumption is now becoming a mass market. Consequently, the production of digital media aims at charming as many people as possible or at least a specific community, identified by its age, culture or taste.

In such a context, we identified an opportunity to contribute to this emerging artistic world by reviving the topic of color harmony from a new perspective.

1.2 An Interdisciplinary Approach

Following the exposed legacy of color harmony approaches, we also come up to this topic with a cross-disciplinary view. We tackle the topic by proposing an experimental approach, as well as building computational models which benefit from experimentally collected data. Along this research, we keep in mind potential limitations of the current editing tools and propose new user-friendly usage of such tools.

Experimental approach

Following an experimental approach, we choose to set up two distinct protocols which were unexplored to measure the factor of Color Harmony.

First, we take up this experimental problem by involving visual attention mechanisms. Despite the fact that other senses may contribute to the evaluation of Harmony, we focused on the vision since color is processed at a very early stage in the brain (already along the retina) [97]. As developed later in the

manuscript, Color Harmony is often described as a subtle equilibrium, a balance, a visual interest or even a correct order... These terms may be linked to the vision field. Thus, we found reasonable to assume the human visual system being stakeholder in the evaluation of color harmony.

Since visual attention is mainly measured by eye-trackers, we propose to characterize color harmony by means of eye movements information measured with a task protocol. In simple words, what catches your attention is likely to determine or influence your opinion about the considered task [275, 64, 248, 88]. However, we are aware that this is not enough to characterize completely a complex notion (so-called factor) as Color Harmony.

After concluding on the effects evaluated on the eye movements, we take advantage of the collected data to design a new dataset used to evaluate our computational models.

Second, we investigate globally Color Harmony, after exploring the local aspect of eye fixations. To do so, we set up a pairwise protocol, where participants were asked to express their opinion about the most harmonious picture over the two presented stimuli. Subsequently, all pair annotations were numerically ranked on a harmony scale. All involved stimuli could then be compared with each other. Once again, the collected data serve two purposes: they are employed to evaluate the factor of color harmony (the inter-rater consistency, the color distribution role...) and provide a ground truth for the benchmark of designed methods.

Computational approach

The second perspective, developed in this thesis, deals with the design of a computational model and a harmonization algorithm. We mainly investigate two different methods with some contributions positioned at different levels: an image processing algorithm for automatic color harmonization and a model for perceptual quality assessment. Those two will be referred to *models* or *methods* indistinctly.

Following the assumption that visual attention plays a role in the assessment of color harmony, we revisit a state-of-the-art architecture which automatically harmonizes the colors of any picture by ingesting a visual attention model, among other contributions. We undertake an exhaustive assessment of the different steps of our method. In addition, we demonstrate objectively with the pairwise protocol the benefit of automatically harmonizing the pictures following such method.

In the second proposed method, we design an innovative quality metric involving the harmony precept. As far as we know, this is the first attempt at merging such concepts. We propose a preliminary validation of this metric by relying on the ground truth maps, designed from the eye movement data.

In a nutshell, these contributions aim at integrating visual perception into existing theory of color harmony.

Application and Use case

On top of these experimental and computational contributions, a supplemental effort has tead us to create a concrete framework that would benefit any user, whatever his/her expertise in image, color and artistic manipulation. The idea behind this approach is to make the theory and rules of color harmony more accessible through intuitive and user-friendly interface.

In such a context, two editing tools were created: the first simplifies and guides the user whilst retouching step-by-step the color areas of a picture, the second proposes to bias or influence the color harmonization processing using other user's pictures. In the first tool, we display the computed harmony quality map as a guidance for disharmonious areas. The second tool benefits from the improvement we brought in the color harmonization process.

From artistic and expert to naive users which experienced and discussed the computational methods, we observe a common feeling about such research: the same frustration of not integrating a personal taste or touch. Hence, these editing tools also address the need to take into account creative intent.

1.3 Organization and Contributions

As a guideline for the reader, the remaining of this thesis is structured into three parts:

Part I: Literature Review.

Chapter 2 draws a picture of the main principles and mechanisms of visual attention. In addition, it also depicts the architecture of the visual attention model employed in the following contributions.

Chapter 3 depicts the main concept, material, data, protocols and metrics which are involved in eye-tracking.

Chapter 4 starts with an introduction to Color Harmony, its origin and definitions. Then, the theories and models, coming from the color science community, are described deeply, as they are referenced exhaustively throughout this thesis. The chapter keeps developing in another section their different use and implementation in the context of image processing.

Chapter 5 discusses the approaches presented in the literature. Also, it depicts the different issues and limitations of previous work.

Part II: Experimental Approach: attention, color and harmony.

Chapter 6 introduces the experimental approach. At the end, the hypothesis and expected conclusions regarding the experimental part are formulated.

Chapter 7 describes the first experiment which is a preamble to the central topic of color harmony developed in the next chapter. It studies the color factor with eye fixations, recorded on color stimuli and their grayscale counterpart. In addition, the prediction of two visual attention models and their performances regarding these two conditions are provided.

Chapter 8 investigates the concept of Color Harmony by means of eye fixation recording. More precisely, we present the gaze data statistics by proposing a qualitative and quantitative dataset analysis, some statistics on eye fixations and an analysis of the inter-observer agreement. As a second contribution, we carefully select the involved stimuli in order to study the behavior of specific color distributions.

Chapter 9 proposes to exploit pragmatically the collected data from the previous chapter. We aims at creating a dataset by post-processing these previously and experimentally collected fixation data. Such dataset serves to the validation of computational methods developed in the part III of the manuscript.

Chapter 10 discusses the different findings of the three preceding chapters. The different hypothesis formulated in Introduction (Chapter 6) are challenged to conclude on their validity.

Part III: Models and Applications.

Chapter 11 explains our approach by depicting the computational perspective, such as done for the experimental part. Then, it draws a picture of the existing algorithms and finally describes the formulation used along the three remaining chapters.

Chapter 12 describes the design and implementation of a computational method for Color Harmonization. It brings two main contributions. First, it improves the visual rendering of such method, by introducing mainly the use of a visual attention model. Second, it proposes both a detailed evaluation of the different steps of the method and a global evaluation of the gain brought by harmonizing a picture.

Chapter 13 proposes a new approach for assessing what the quality of a picture is. The proposed computational method relies on perceptual masking and color harmony models to output a no-reference perceptual harmony-guided quality map as well as a score of disharmony. An evaluation is proposed by means of collected data from Chapter 8.

Chapter 14 concludes the computational contributions of this thesis with two tools that directly implement the previously proposed computational methods. These latter aim at guiding the end user and introduce some interactivity in a task of image editing.

Chapter 15 concludes and summarizes the contributions of this thesis. It explores directions for future work.

1.4 Beyond the scene

In order to complete the introduction of this thesis, we would like to establish the context in which it took place. Even though this thesis focus on the topic of Color Harmony through a visual attention perspective, it takes advantage of the expertise I built for several years on different topics, related to Perception.

I started working on visual attention models aiming at measuring [48], improving performances of prediction on 2D content [45, 23] and simplifying the biological architecture [256] to ensure real-time industrial application [48, 44]. Then, I studied also the modeling and potential application of visual attention models for stereoscopic content [47]. In such context, I developed an expertise and a real taste for experimental protocols, always performing both the experiments and algorithms in accordance. I attempted to apply this philosophy when I addressed the topic of Color Harmony in this thesis.

Afterward, I got interested in the prediction of induced emotion through the supervision of a phd student around this topic. Once again, the experimental approach serves the computational work and each component was equally balanced and developed by the student in his thesis [18, 22, 20, 21, 19, 151].

Then, I also worked on image processing algorithms, more precisely color mapping [76], such as:

- Inverse Tone Mapping of legacy content to produce High Dynamic Range video content [212, 221, 222],
- Colorization of black and white pictures from a reference picture [135],

I am currently investigating new promising and difficult ways about perceptual model. Through a large campaign of different experiments, we attempt to characterize the concept of interestingness. What makes an image interesting? We define the notions of intra- and inter-interestingness to study more precisely the local and global aspects of such concept. Our first experimental conclusions [46] conduct us to model such concept for an average observer in upcoming publications.

In such rich context of experiments and algorithms, I developed the topic of Color Harmony while relying on my expertise in visual attention.

Part I

Literature Review

Chapter 2

Visual Attention

This chapter reviews the main principles of visual attention. We have not the ambition to develop an exhaustive explanation of the literature, but we prefer focusing on key mechanisms that are the foundations of visual attention. The mentioned principles are a preamble to the visual attention modeling which contributes to two proposed computational methods (Chapter 12 and 13).

2.1 Introduction

The surrounding world is evolving exponentially in terms of technology. While the breakthrough related to the color introduction in television standard dates back to 1960, in fifteen years many new revolutions, related to screens, devices, portability, content, accessibility, have changed our consumption habits. This is not untypical to observe either a person watching several screens at a time or another one reading the news or playing a game while walking in the street. The cognitive activity involved in daily digital consumption is getting more and more important. Thus, we can easily imagine that the brain, the cells, the mechanisms involved in such daily "fight" are evolving and adapting to survive.

The visual attention is challenged every day by an environment more and more complex and a bigger offer for digital entertainment. The competition seems unequal and the visual attention system has no other choice that being *selective*. Also, no doubt the visual attention is also influenced by other modalities that only visual stimuli: the audio signal obviously plays a role in attentional mechanisms that can not be neglected [29, 80], also internal human status or mood as well as other senses (smell, touch, taste) are relevant [258]. The human is full of specialized sensory receptors that transmit to the brain the sensation of the environment. Those are influencing our perception and our attention. In this Chapter, we will not tackle the broad and complex topic of sensation and perception. Only the field of attention related to the visual modality is investigated and discussed.

By essence, the *attention* word sounds as the discrimination of contrasted phenomena. The stimuli is partially treated by the brain that pre-processed the most relevant features, activity or meaning to be used for the global context understanding. The visual attention suggests the analysis of the perceived environment with or without the instruction of a specific task. Despite the main involvement of psychologists for promoting visual attention (Section 3.5), there is an intensive research activity where different communities has brought some considerable contributions:

- biology: identification and role of cells (Section 2.2.2),
- neuroscience: sophisticated neuroimaging technology for measuring stimuli impact and identifying involved parts of the brain (Section 2.2),
- physic and optic: for providing apparatus such as eye-tracking that measure eye movements (Chapter 3),
- signal processing: for having proposed and implemented modeling of visual attention (Section 2.3).

The scientific work on visual attention exploded from the 1980's consistently with the maturity and accuracy of measurement apparatus. The advances in functional magnetic resonance imaging (fMRI), event-related potentials (ERP) and magnetoencephalography (MEG) lead to significant breakthroughs in the characterization of visual attention. In addition, the progress made in the accuracy and reliability of the eye-tracking apparatus also contributed largely to the understanding of visual attention deployment by studying eye movements. Another factor favored the expansion of the research in this field: the computational modeling takes advantage of biologically-plausible brain mechanisms. Indeed, the findings of neuroscience experiments have been purposely reused for building biological-inspired models [42].

Many remarkable psychologists participated early to the coming of visual attention theory. While William James (1842-1910) promoted the voluntary and active features of visual attention controlled by a purpose [118], earlier Hermann Von Helmholtz (1821-1894) rather believed in a fine and detailed observation of objects or a precise area [97]. Both psychologists laid the foundations of contemporary modeling of visual attention.

Following this introduction, we introduce the main principles of visual attention (Section 2.2) where we develop the main concepts (Section 2.2.1), the biological concepts (Section 2.2.2), the color vision theory (Section 2.2.3) and the processing inferred in the brain (Section 2.2.4). Thus, in a section dedicated to visual attention modeling (Section 2.3), we describe the different approaches (Section 2.3.1) then we focus on cognitive models (Section 2.3.2) and finally we depict a simplified model (Section 2.3.3) which is later used in the contributions of this thesis. The last section summarizes the different notions presented in this chapter (Section 2.4).

2.2 Mechanisms of Visual Attention

2.2.1 Main concepts

The field of visual attention received a lot of attention and substantial efforts have been made the last decades by the researchers of different communities to understand and characterize its mechanisms. However, there is no consensus on how visual attention works precisely. Despite that, we are introducing the most commonly adopted principles related to visual attention.

First, the visual attention is *selective*. This is clearly due to the limited capacities of the human cells to process complex stimuli and information. The visual attention selection consist of two stages: a *pre-attentive* and an *attentive* processing [257]. Both stages are considered independent from each other and have a different role. The *pre-attentive* stage has unlimited capacity and operates simultaneously or in parallel in the visual field. On contrary, the *attentive* processing with limited capacity focus on one item for a better inspection of its features. More specifically, we can say that the *pre-attentive* stage detects low-level features or characteristics of the scene (such as color, edges, orientation, motion...), that are integrated afterward in the *attentive* stage for a complete understanding and awareness of the scene. Such concept referred to the Feature Integration Theory (FIT) developed in 1980 by Anne Treisman [254] and will play a fundamental role in the computational modeling of visual attention (Section 2.3).

In the literature, the spatial aspect of attention is also often discussed. Two well-known models, conceptually similar, evoke a delimited spatial region where the processing is located. The *spotlight* model of Posner *et al.* [210] describes three areas: the focus, the fringe and the margin (Figure 2.1). Typically, in the focus area the processing is performed at a high resolution, while in the fringe section less receptors are involved leading to a low resolution analysis. Beyond the margin area, the processing is not engaged. The *zoom lens* model from Eriksen and Yeh [73] also relies on the same metaphor and spatial organization (focus, fringe and margin). Their new contribution deals with the adaptation of the region size or radius. Following the idea of a camera and having the attentional resources fixed, the size of the focus area directly influences the quality of the processing (meaning the resolution available for analysis). Then, if the focus radius is large, the processing is likely to be less efficient. This theory has to be linked with the notion of fixation measured by eye-tracker and discussed in Chapter 3.

The concept of *overt* versus *covert* visual attention has to be specified, especially because eye movements data are manipulated along this manuscript. Covert visual attention does not involve eye movements. Typically, the subject focus on one area and monitors the surroundings in order to search for an object, to guide its movement in space [207]. Once the covert visual attention has ful-

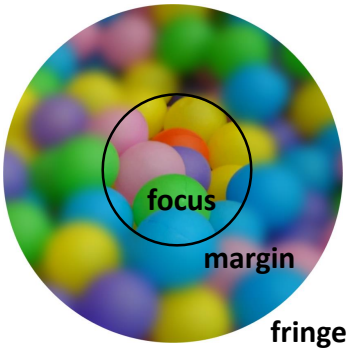


Figure 2.1: Spatial representation of attention by Posner [210]. Three spatial areas are identified: focus (high resolution), fringe (low resolution) and margin (not visible).

filled its function, it engages the eye movements in the scanning of a particular target, this is the overt visual attention [74, 211]. This latter is characterized by the different *varieties* of eye movements (fixations, saccades, smooth pursuit) [41] and remains the center of this manuscript. A set of fixations (area which is fixated) and saccades (amplitude of displacement between two fixations) creates a scanpath [188]. Note that these notions are discussed and illustrated in Chapter 3.

Pioneer work of Yarbus [275] is well-known from the community: he showed the evidence of task-related scanpaths for identical stimuli. Even if the experiment condition as well as the conclusion have been recently revisited [64, 248, 88], he introduced the notion of task dependency later investigated under different conditions. From these observations, visual attention can be categorized once again. *Bottom-up* (or *exogenous*) attention is assimilated to a reflex, an involuntary and unconscious movement. Usually, it is a very fast phenomena not motivated by a task. At the opposite, the *top-down* (or *endogenous*) attention is deployed voluntary and consciously. It is a slow action originally following a dedicated task.

There are other categories of visual attention depicting other conditions, environment (e.g. multi-tasks), stimuli (e.g. aesthetic pictures) or processing in the brain (e.g. memory). However, we will not discussed these different notions which are not related to the proposed work.

2.2.2 Biological concepts

Along the wide electromagnetic spectrum, only few wavelengths are perceptible and translatable by the eye. These visible wavelengths are the light and the colors as we may described them; they are spread only from 380 to 760 nanometers (10^{-9}), while at the boundary, gamma rays are around 1.10^{-14} and long

waves around 1.10^4 meters. This short visible band is only what the specialized receptor cells in our eye are capable to transform into a neural signal into the brain.

At this stage, it is interesting to make the link between the physical signal and color science. As mentioned previously, different values of wavelength depict a perceived color or a *hue*, e.g. a wavelength around 430-460 nm is a blue light. The amplitude of this wavelength represents the *brightness*, high wave amplitude means a brighter color. Also, the noise or the interference in the waveform shows the color purity, or its *saturation*. Hue, saturation and brightness are directly linked to the visible spectrum of wavelengths.

The eye demonstrates some properties due to its physiological organization. We do not detail the complete structure of the eye, since all elements are not participating to the visual attention. However, it is interesting to focus and notice the key role of the *retina* and *fovea*. The retina, located at the back of the eye, is a multilayer surface that records the wavelengths entering through the pupil. Such energy is stored and transformed on the retina into a neural signal or response for immediate processing in the brain. By means of the millions of photoreceptors constituting its surface, the retina is able to achieve this fundamental task.

The *rods* and *cones* are the two kinds of photoreceptors that are processing the input wavelengths. They have complementary properties and sensitivity to the signal. Coarsely, rods are specialized to process light and respond to low illumination while cones are rather dedicated to color vision and respond to high illumination. A small area, named *fovea*, on the retina plays a key role for the vision. This is the place where the vision is the best and which enclose only the cones. The rods are spread all along the retina except on the fovea (Figure 2.2). Once the *transduction* step has been performed by the two specialized photoreceptors, the produced signal is transmitted to the bipolar cells and after to the ganglion cells through the optic nerve which carries it to the brain.

2.2.3 Color vision

As mentioned in section 2.2.2, the specialized photoreceptors named *cones* are in charge of translating the wavelengths into neural information related to color vision. The explanation about the way they operate has opposed two theories that reveal to be both valid and complementary: the *trichromatic* theory and the *opponent-process* theory.

Color vision has a long history of psychological experiments to characterize color perception at the eye level. Originally, the *trichromatic* theory has been introduced by Thomas Young in 1802 and refined in 1852 by Hermann von Helmholtz [97] (already mentioned in section 2.1 for his work on visual attention). They found out that there exist three classes of cone: red, green and blue. They experienced the human capability to recreate an unique color wavelength from three others and derived the conclusion about the existence of three varieties of cones. However, such findings was not strong enough to explain some

responses of people having color abnormalities. Later in 1892, Hewald Hering [101] proposed the *opponent-process* theory that could answer some questions raised by the trichromatic theory. He believed in four kinds of cone that work by complementarity (red-green and blue-yellow). Thus, each cone is excited by the color 1 and inhibited by the color 2, both forming the considered pair.

The community gave support to both theories that turned out to be correct and complementary. Three varieties of cones - red, green and blue - (trichromatic theory) respond to a color difference following a three opponent mechanism (opponent-process theory). Thus, the first kind measures the differences in the response of the red and green, the second kind measures also the differences between the blue cones and the sum of the red and green (yellow) cones and the third kind measures the differences in luminance due to its achromatic nature.

This theory has been naturally *translated* in the computational models of visual attention.

2.2.4 Processing the information

Once the light information captured from the eye has been translated into a neuronal information, the brain role becomes predominant. A dedicated area for visual processing has been identified at the back of skull, the visual cortex. Through different areas, this latter analyzes some specific characteristics of the visual information. These visual signal being so rich and complex, the brain owns specialized cells, areas to decompose the signal characteristics (color, shape, motion, face and so on). However, such processing cannot be sequential to consider a fast response and understanding of human. The processing of information is designed in parallel, with different *pathways*. The two main categories of pathways refer to the *what* and *where* pathways [231, 85, 169], also known as the ventral stream and the dorsal stream, respectively. As their name suggests, the *what* pathway deals with the stimulus features (e.g shape, color) and the identity of an object, while the *where* pathway has more concerns about the spatial information of an object (e.g motion and depth). This concept is illustrated in Figure 2.2.

The optic nerve projects the neuronal signal to the *dorsal lateral geniculate nucleus* (LGN). Mainly, this area gathers the information coming from each eye and push them to the visual cortex and more specifically to the *primary visual cortex* (also called area V1). The visual cortex is made of six layers, whose the most well-known and studied is the area V1. This area proceeds as detecting edge and encoding them in spatial frequencies rather than encoding the complete received image. This is an efficient and natural compression of the information. The area V4 has been identified to provide a high attentional activity [175]. It deals with the following features: orientation, spatial frequencies and form recognition. Additionally, it is suspected to play a key role for color processing [83]. The cells present in V5 deal with the speed and direction of moving stimuli. Globally, this is not obvious to split the role of each area according to the addressed features.

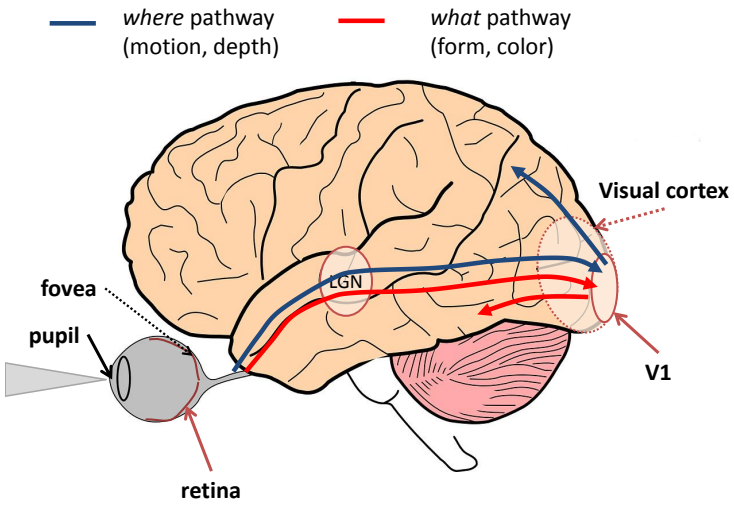


Figure 2.2: Eye and brain anatomy. what and where pathways in the brain dealing with the processing of different features

2.3 Models of Visual Attention

2.3.1 Approaches

The computational modeling of visual attention has been widely explored in the literature. The approaches are almost as vast as the number of techniques used in image processing. Also, computer graphics, computer vision, machine learning have brought their contribution in the detection of region-of-interest. A visual attention model is an algorithm that inputs a visual stimulus, as an image or a video sequence, and outputs a *saliency* map (one per frame if considering a video source). The *saliency* map concept has been introduced by Koch and Ullman in the context of their feed-forward model [131]. It describes if a pixel is conspicuous (visually relevant) with regards to its neighborhood. In other words, a greyscale map, having values from 0 to 255 (low to high interest) for each pixel, is estimated. This is the prediction map computationally estimated by the model. Note that human-made map (with homogeneous dimensions) are usually created experimentally in order to confront the model to a ground truth (Section 3.4).

In a complete review, Borji and Itti [30] provided to the community an exhaustive classification and benchmark of the different approaches. They discussed the prediction quality through an exhaustive list of dataset and the employed metrics. We can differentiate the models that have a biological architecture (meaning inspired from the findings of psychologists) from those purely computational. However, it does not mean that the "non-biological" models perform

worse than their biological homologue. They are also based on acceptable and intuitive arguments that statistically are good enough to compete with the simplified or adapted architecture derived from biological concepts.

In [30], different two-by-two comparisons are discussed to compare the different approaches, e.g. the pixel- versus object-based models or the bottom-up versus top-down approaches. Finally, the authors went for a classification related to the architecture or methods employed and pointed out eight clusters: cognitive models, bayesian models, decision theoretic models, pattern classification models, spectral analysis models, graphical models, information theoretic models and a miscellaneous category. In the next section, we will focus on the cognitive models and more specifically on Le Meur's model [170, 144].

Despite the different methods used for predicting the salient areas, all are motivated from the same theory already mentioned in section 2.2: the Feature Integration Theory [254]. In her theory, Anne Treisman distinguised two stages, the *feature search* and the *conjunction search* that also confirm the idea of a *preattentive* stage and an *attentive* stage. During the early *feature search*, the object characteristics, such as color, shape, orientation and movement are detailed individually in different areas of the brain (section 2.2.4). In the late *conjunction search*, the previous information are combined to make a complete perception of the considered object through a *master map* of locations. This latter includes all locations in which features have been detected. These locations point to the multiple feature maps.
All visual attention models are based upon this theory, because they usually proceed in two steps:

- First, they detect, extract image features. These cues may be directly linked to the FIT: shape, color, orientation, movement...or they may be more intuitive and high levels: face, scene classification, gist...
- Second, they integrate in a specific manner these features; e.g. either by combining linearly independent feature maps, by learning from a dataset and inferring a suitable feature combination or by using probabilistic model to estimate the likelihood of a feature usability.

In the next section, we are focusing on cognitive models and more particularly on the architecture of pioneer models that integrates only the visual modality. However, some work also evaluated the sound influence in eye movements [59] and introduced the audio modality as a key feature in the modeling of visual attention [58].

2.3.2 Cognitive models

Itti *et al.* [113] were pioneer in implementing a complete version of the feed-forward model designed by Koch and Ullman [131]. Moreover, this first implementation was made available to the community, then it assured his popularity for the two following purposes: to benchmark any new implementations and

to be used as a preprocessing or a core technology in the framework of other algorithms, e.g. compression [69], retargeting [6], surveillance [215], robotics [8], virtual agents [112] etc. Since we proposed a saliency-guided method in Chapter 12, we focus here on the cognitive model involved in this contribution.

The saliency model used in this work is a simplified version of Le Meur's model [170, 144]. His model followed the Itti's modeling, but proposed some additional contributions: a more biological plausible modeling [170] and a temporal modeling of the attention evolution [144].

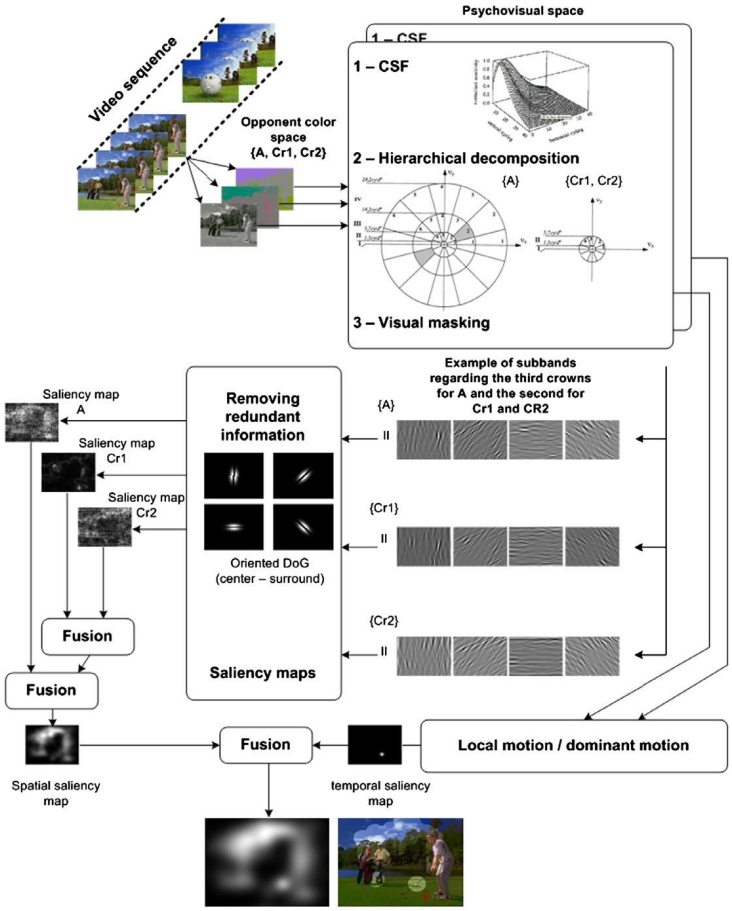


Figure 2.3: Architecture of the spatio-temporal model of visual attention by Le Meur *et al.* [144]

Figure 2.3 depicts the Le Meur's model whose the main steps are:

- **Color projection:** the input stimulus is an image composed of three color components. The first step of the model consists in projecting it into a

color space that best follows the eye behavior. Inspired by the opponent-process theory (section 2.2.3), the signal is rearranged into the Krauskopf's color space *(A,Cr1,Cr2)* [132]: the *A* component is the achromatic signal, *Cr1* is the red-green opponent and *Cr2* is the red-blue opponent.

- **Frequency thresholding:** this steps consists in selecting samples visible by the eye for the three color components. Contrast Sensitivity Functions (CSF) have been designed by Daly [61] for the luminance and later extended by Le Callet [141, 142] for chromatic components. By means of psychophysical experiments, they provide visibility thresholds that report the eye sensitivity and selectivity to a set of spatial frequencies. Thus, a perceptual subband decomposition is performed to split the image into perceptual spatial frequencies. In this case, a Fourier transform decomposes the three components into amplitudes as a function of spatial frequencies. Those latter are weighted according to the visibility thresholds: if the amplitude of the considered spatial frequency is higher than its visible threshold, these one is perceived by the eye and then amplified by means of the corresponding CSF.

- **Hierarchical decomposition:** In order to mimic the organization of the visual system [232] and to also take into account eye sensitivity with its neighborhood (called *visual masking*, see next step), the spatial frequencies are sorted into spatial radial frequency and orientation. This arrangement in channels allows simulating the functioning of cortical cells. This is illustrated with the crowns and subbands in Figure 2.3.

- **Visual masking:** This processing is complementary to CSF. While CSF describes the own perception of a spatial frequency independently of the other one, the concept of visual masking formalizes the perception of a spatial frequency with respect to its surrounding. A visually rich region (e.g. the audience during a football match) tends to produce a favorable ground for high visual masking, while an homogenous and flat region (e.g. the grass where players are running) is easy to analyse because less information must be processed. Several functions of visual masking have been designed experimentally [61, 265, 141]. Le Meur considered three different kinds of masking in [144]: intra-channel intra-component masking, inter-channel intra-component masking and inter-component masking. These masking refer to the interference of spatial frequencies within a channel, between channels and between the components. More details may be found in [170] and in Section 13 where the intra-channel intra-component masking has been employed.

- **Redundant information management:** In an effort to optimize the processing of information, an economical strategy is set up by the visual system. The center-surround organization of the cortical cell allows selecting relevant regions and removing redundant information. This human vision concept has been implemented by a difference of gaussian function.

- **Temporal saliency:** Following the findings of Itti [111], Le Meur *et al.* are interested in motion contrast to measure the temporal saliency. Rather than a plausible implementation, Le Meur *et al.* prefer follow the contrast idea supported also by Itti [111]. Thus, the temporal saliency prediction is formalized by the motion contrast. In other words, they do the difference between the local motion and the dominant motion in a scene. The local motion is a hierarchical block-matching algorithm, that estimates the motion vectors for each block between two frames. The dominant motion is the estimation of an affine (6 parameters) model from all pixels between two frames. At this stage, all early feature maps have been extracted. Nonetheless, a tricky step for pooling maps is required.

- **Fusion or pooling:** This step consists in merging all the available maps: the three spatial maps and the temporal saliency map. This tricky step is widely discussed in [144], where Le Meur *et al.* finally proposes to hierarchically merge the available maps, first the spatial maps and then the temporal map. They merge the maps two-by-two and derive an intra-map and inter-map competition to normalize the map and then promote the saliency peaks (while softening the lowest peaks). We discuss these fusion aspect through a study of different methods in [45].

2.3.3 Simplified cognitive model

Facing complexity issues, especially for deploying such model in real-time image processing techniques, we produced a *light* version of the Le Meur's model. Note that fast or real-time visual attention models have been investigated also in the literature [152, 57]. While maintaining the biological modeling of the original model architecture, we simplified most of the steps and even removed some of them. Mainly the computational complexity was related to the different filtering stages, thus we proposed the following adaptation of the Le Meur's model, also detailed in [256] and depicted in Figure 2.4:

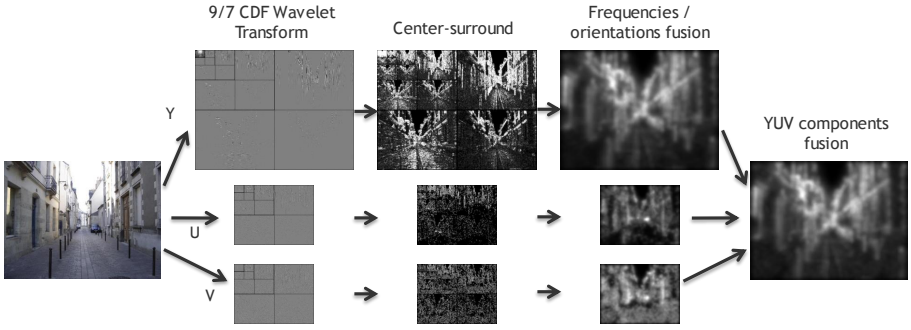


Figure 2.4: Architecture of the proposed simplified version of Le Meur's model

- **Color transform:** The YUV 4:2:0 color space replace the Krauskopf's color space (A,Cr1,Cr2) [132]. It separates achromatic (Y) and chromatic (U: green-magenta and V: orange-cyan) perceptual signals, with chromatic components having half the spatial resolution of the achromatic component. This color space has been chosen because it takes the human visual system into consideration and is commonly used in image and video processing, but it remains less biologically plausible than the Krauskopf's color space.

- **Hierarchical decomposition:** A wavelet decomposition is used here as an approximation of the original Fourier transform. In [185], Ninassi *et al.* brought the evidence that a wavelet transform could be a good compromise to approach frequentially a perceptual subband decomposition at a reduce computational cost. A 9/7 Cohen-Daubechies-Feauveau (CDF) wavelet transform is used to separate frequency bands and orientation ranges [12]. The wavelet transform separates frequencies with a bank of filters, i.e. low-pass and high-pass filters followed by a critical sub-sampling. The resulting multiscale decomposition consists of oriented contrast maps with limited frequency range and a low-resolution image. The number of wavelet levels is defined so that the last decomposition level contains details that can be captured by the fovea (1.5–2 degrees of visual angle). As explained later and depicted in Figure 3.3a, the fovea is able to analyze at a high resolution only a centric area with a radius inferior to 2 degrees of visual angle. Thus, there is no need to express the spatial frequencies smaller than this level.

- **Redundant information management:** In the same vein as Le Meur *et al.*, a Difference of Gaussian (DoG) modeling the center-surround response of visual cells is applied on each oriented contrast map (wavelet subbands). For each location, the center-surround filter is computed as the difference between current location and the average of surrounding. In our implementation, the surround area is a square of 5 x 5 pixels for each pyramid level.

- **Temporal saliency:** this step is similar to Le Meur *et al.*. The motion contrast is detected by subtracting the dominant motion to the local (block-based) motion.

- **Fusion or pooling:** Two fusions are successively applied: subbands corresponding to the different levels and the color channels. First, the orientation maps from each level are summed up together. Then, the maps obtained at each decomposition level are up-sampled using a bilinear filtering and a per-pixel addition is performed for each level to get the final map.

We can notice that the CSF and visual masking processing have been removed from this model. This is conceptually annoying. However, the prediction performances of this model are similar to Itti's model, such as quantified in [256]. A

large benchmark versus state-of-the-art models is provided in Annexe A. As expected the performances of the simplified model does not reach state-of-the-art models, especially the emerging deep learning approaches. However, regarding the strong simplification performed in this proposed architecture, the compromise between computational complexity and prediction quality is pretty good. We consider this implementation as a fast and efficient way to produce saliency map and we have based the contribution of Chapter 12 on this implementation.

2.4 Summary

Visual attention mechanisms have been widely explored, but the field did not manage to achieve a consensus. Advances in side technologies (eye-tracking, neuroimaging...) contribute substantially to the exploration of new intuition and the development of theories (Section 2.1). However, we presented the most commonly adopted theories: the *preattentive* and *attentive* stages of visual attention, the *Feature Integration Theory* as well as the eye organization and cells behavior (Section 2.2.1). This is a preamble to the topic of visual attention modeling.

As a natural extension to the understanding of visual attention mechanisms, the computational models of visual attention exploded the last decade (Section 2.3.1). Many approaches, biologically-inspired or not, appeared and demonstrated high prediction of visual attention, formalized by the saliency map. A complete review of visual attention models may be found in [30]. Focusing on biologically-inspired or cognitive models, the understanding of human visual system allows designing plausible architecture. While such models mimic the information processing from the retina to the visual cortex, the performances of these models are comparable to the one involving no biological architecture (machine learning or probabilistic approaches). Specifically, we described the biological-inspired architecture of the Le Meur's model (Section 2.3.2) as well as a simplified implementation we proposed in a strict context with real-time issue (Section 2.3.3).

There is also a field of investigations that is developed in the next chapter: the eye-tracking and the creation of ground truth. Obviously, the performances of visual attention models remain dependent to both the way the human-made saliency map are designed and the metrics used for comparison. This is discussed in the Chapter 3.

Chapter 3

Eye-Tracking

In this chapter, we review the topic of eye-tracking. As mentioned in the previous chapter, the advances in eye-tracking technology considerably have contributed to the understanding of visual attention. Not only the understanding of attentional mechanisms benefit from the high sampling rate and the data (pupil size, micro-saccades amplitudes...) provided by these apparatus. But also the ability to build up experimental saliency maps as a ground truth from accurate eye movements data contributes significantly to the development of visual attention models.

Nowadays, the technology constituting eye-tracking apparatus is mature and allows proposing a set of commercial solutions addressing different communities. We can find very accurate material for ophthalmology as well as user-friendly devices for marketing analysis. Thus, the range of devices goes from head-mounted apparatus to clinical head-constrained solution. However, the exploitation of eye movements data requires an expertise and rules and protocols should be followed to ensure a reliable conclusion. These aspects are discussed in the Chapter.

We propose in this thesis several contributions related to visual attention (Section 1.2). Relying on our expertise, we employed eye-tracking as a way 1/ to measure attentional mechanisms with a task related to color harmony, 2/ to create a ground truth for our computational model.

Consequently, we will introduce the main concepts hidden behind eye-tracking (Section 3.1) and we will detail what are its outputs (Section 3.3), how they are post-processed to finally produce exploitable data (Section 3.4). Also, we will describe briefly the two apparatus used in the mentioned experiments (Section 3.2) as well as the usual protocols (Section 3.5) when designing an eye-tracking experiment. Finally, we will introduce some metrics needed to quantify both the inter-observer consistency of recorded eye movement data and the performances of computational visual attention maps (Section 3.6).

3.1 Main Concept

Eye tracking is a technique whereby an observer's eye movements are measured so that one knows both where this person is looking at any given time and the sequence in which his eyes are shifting from one location to another. More specifically, it consists in measuring the point of gaze or the motion of an eye relative to the head. The estimation of gaze direction means the head movement in addition to eye movement is taken into consideration for estimating the area fixated by the observer. Thus, in apparatus where the head is constrained in a chin rest or the measure is performed directly on the head (as with Electro-OculoGraphy or glasses), the eye and gaze directions are the same.

An eye-tracking apparatus (or eye-tracker) typically measures the *overt* visual attention (Section 2.2.1) that is reflected through eye movements. The subject by means of eye movements, typically alternates between a *pause* on specific areas (fixation) and a *skip* to the next area of fixation (saccade). These notions are more specifically described in Section 3.3.

The first prototypes of eye-tracker date from the beginning of the 20th century, where more or less intrusive solutions were designed. While Huey (1908) built a contact lens with a hole for the pupil, Buswell (around 1922) was more concerned by the intrusiveness of any potential solutions. He used beams of light that were reflected on the eye and then recorded on film. From mid 20^{th} century till now, finding a good compromise between intrusiveness and accuracy of measures has occupied many researchers and commercial companies. There exists large number of hardware solutions whose the selection is clearly guided by the applicative context.

Generally speaking, there are two families of eye-tracker classified according to their monitoring conditions: those that measure the eye movement relatively to the head motion and those that estimate the direction of regard or the orientation of eye in space. The underlying technology involved in measurement also provides a good categorization: the Electro-OculoGraphy (EOG), the Contact Lens and the Pupil/Corneal reflection are the main way to measure and track eye movements.

The EOG technique expanded rapidly during the 1970s. This signal-based technique records the electrical potential differences on the skin surrounding the eyes. The contact lens techniques aim at originally improving the accuracy of first corneal reflection method (dating back early 1900s). Some devices, such as small mirrors or coils of wire, are attached to the contact lens. They count on the physical contact with eyeball to provide highly sensitive movement measurements. The main drawbacks of the two aforementioned techniques are their intrusiveness. While pretty accurate, the way to measure eye movement in this condition may lead to bias the occulometric behavior of the subjects. This reason probably makes more popular the corneal or pupil reflection method at the moment.

Infrared (IR) cameras (usually one for each eye) projected on subject creates a corneal reflection relatively measured to the pupil center. This is also known as the Purkinje Images [56]. Over the four available Purkinje Images formed by the eye, only two of them may be exploited by the eye-tracker. Requiring first an appropriate calibration, the direction of regard is estimated on a surface (typically the screen). Two points of reference during calibration on the eye are needed to separate eye movements from head movements. In fact, the measured differences between the pupil center and corneal reflection remain the same under head movements, but vary with eye movement. In many setups, the infra-red cameras are part of a fixed material leading to a known distance between eye and sensors. Thus, the measurement is stable, while the eye may scan the proposed stimuli and the subject may move his head to a certain extent.

Recently, another singular family of eye-tracking get developed by the image processing community. No specific hardware is required, the purpose is to provide a low cost eye-tracker. This solution are purely based on image processing and rely on any webcam signal. The face, then the eyes and finally the motion of pupil are detected and tracked by using usual algorithms (KLT for tracking, face detection, variance estimator...) and by being supported with visual attention models. Obviously, the accuracy of such algorithm is worse than professional costly solutions, but not ridiculously far. We implemented such kind of system and end up with an accuracy of about 3 degrees of visual angle [182].

During the last decade, the eye-tracking based experiments became more popular due to the simplicity of their hardware setup, the software effort to process the results and their accessible cost. They are not restricted anymore to the expert community dealing with vision and eye movements. Advertisers and designers concretely have employed the protocol of measuring gaze and their statistics to evaluate the potential of brand impact and product placement [154, 266].

3.2 Material

Different apparatus, measuring the eye fixations of any users for variable conditions (seating in front of a screen, walking in the street...) have been designed demonstrating different hardware setups and associated capabilities. While portable devices record the scene viewed by the observer, the on-site apparatus control itself the proposed stimuli. There are a bench of commercial solutions for eye-tracking; Figure 3.2 illustrates non-exhaustively the different solutions available on the market. We will not detail all these approaches, but rather focus on the specifications of the two eye-trackers that we have used in the following experiments.

Figure 3.1: Commercial eye-trackers, a bench of solutions: head mounted devices, embedded in car or glasses, professional and clinical solutions or non-intrusive on a regular PC.

SMI© iView X^{TM} RED

The SMI company proposes a large choice of eye-trackers with different hardware. The iView X^{TM} system is a dark pupil eye tracking system that uses IR illumination and computer-based image processing. Images of the eye are analyzed in real-time by detecting the pupil, calculating the center and removing artifacts. Once a calibration is performed, the pupil location is translated into gaze data.

In this solution, a dark pupil approach is employed. The eye is illuminated by IR light from a camera at a certain angle. The eye and face reflects this illumination but the pupil will absorb most IR light and appear as a high contrast dark ellipse. An image-analysis software determines where the center of the pupil is located and this is mapped to gaze position via an eye-tracking algorithm. Dark pupil approach is versatile, so potentially unstable, and easier to set up, but they also require some kind of head movement compensation.

The RED (Remote Eyetracking Device) model (Figure 3.2a) is the one we used for the Experiment in Chapter 7. It is developed for absolutely contact-free measurement of eye movements with automatic head-movement compensation. Specifically, when compensating for head movements, it tracks the cornea reflex (CR) in relation to the static camera. The CR location in the eye changes with head position relative to the camera and it is used along with pupil location to determine the gaze point in the stimulus.

The product specifications claim a gaze accuracy inferior to 0.5 degree of visual angle.

SR© EyeLink 1000 Remote/Head Free

This eye-tracker works on the same principles as RED. However, it provides likely a higher accuracy due to its high sampling rate (1000 Hertz vs 50 Hertz for the RED).

The Remote/Head Free camera upgrade for the EyeLink 1000 system adds the ability to use the eye tracker as a fully remote system that does not require head stabilization (Figure 3.2b). The Remote Camera Upgrade is designed for areas of eye tracking research where a head rest or head mount is not desirable, but high accuracy and resolution are still important.

This system is more dedicated to quantitative eye movement analysis than the previous eye-tracker, even if the RED provides already satisfactory results. More sophisticated hardware allows a better accuracy. However, the calibration of subjects proved to be long, it is more tricky to manipulate and additional piece of software needed to be written to adequately control the environment and the experiment. Exported raw data may be of high interest for scientists who wish carefully to denoise and analyze the pupil movements. This eye-tracker has been employed in the experiment described in Chapter 8.

(a) (b)

Figure 3.2: Two eye-tracking apparatus used in our experiments. (a) SMI© iView X^{TM} RED, (b) Eye Link 1000 Remote/Head Free

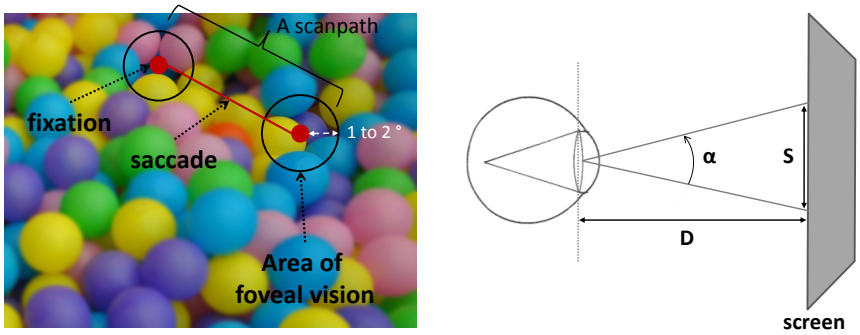
3.3 What is measured?

Sophisticated software solutions have putting eye-tracking within everyone's reach; it provides post-processing of these low level data and a more exploitable presentation of them, e.g. heat maps, scan paths..., aggregated for all users or a subset, provided for the complete pictures or only for a pre-determined region-of-interest. The raw outputs of an eye-tracker are **eye fixation** and **saccade**. Saccade and fixation are basically the two states of eye movement such as depicted in Figure 3.3a:

- **Fixation:** it occurs when the gaze direction has a stable position, permitting the eye to extract, encode and process information. Typically, the fixation duration ranges from 200 ms to 300 ms [217], but can get to 600 ms. Note that 90% of viewing time is allocated to fixations [109].
- **Saccade:** this is the shift or the fastest movement of the gaze between two fixations. Typically, a saccade lasts from 10 to 100 ms and have amplitude range from 1 to 40 degrees of visual angle (Figure 3.3b) with velocities that can reach up to 900 degrees per second [217].

Other kinds of eye movements have been identified in the literature, but have less interest in the context of eye-tracking and visual attention. *Convergence* is a motion of both eyes relative to each other that ensures that an object is still foveated by both eyes when its distance to the observer is changing. This movement can be voluntarily controlled, but is normally due to a moving stimulus. *Pursuit motion* is a much smoother, slower movement than a saccade; it acts to keep a moving object foveated. It cannot be induced voluntarily, but requires a moving object in the visual field. Finally, *microsaccades* occur during fixations and consist of slow drifts followed by very small saccades. These movements are involuntary.

First introduced by Noton and Stark [189, 190], a *scanpath* refers to a temporal sequence of fixations and saccades. The properties that are usually exploited for statistical analysis of scanpaths are: the fixation duration and location, the saccade amplitude and duration.



(a) (b)

Figure 3.3: Illustration of saccade, fixation on image and visual angle. (a) Spatial representation of saccade and fixation. The radius of the fovea region ranges from 1 to 2 degrees of visual angle. (b) Notations to compute the visual angle α.

When referring to vision, eye movement and eye-tracker, the indicator for accuracy is usually expressed in visual angle (degrees), such as depicted in Figure 3.3b for the different vision areas. This measure is the trigonometric relationship between the distance of the user to the screen and the size (radius) of the projection on the screen (Equation 3.1). The lower the angle of dispersion, the higher the accuracy.

$$\alpha = 2 \cdot arctan\left(\frac{S}{2D}\right) \tag{3.1}$$

This is not uncommon to find different report on statistics about fixations and saccades. Being not only dependent on the apparatus accuracy, the task of the experiment influences clearly the recording and conclusion. During a visual-search task, the involved attentional mechanisms will not be the same as a free viewing exploration [275]. More concretely, a reading task will not produce the same statistics as a gaming activity. While uninitiated person may

directly exploit *clean* data and maps processed by a software, experts would prefer extract raw data and post-processed them.

3.4 How to exploit the data?

In this section, we will not detail the methods used to specify and differentiate fixations from saccades. Basically, the stationary characteristics of the measured signal is analyzed (internally by the eye-tracking solution) in order to detect abrupt changes (saccades). More details about all possible techniques may be found in [226]. Nonetheless, we can notice that there exist two main approaches: the *dwell-time* method of fixation determination averages the temporal signal over time and determine based on variance statistics and empirical thresholds if a fixation may be considered. The *velocity-based* method computes the velocity of the signal measured between two successive samples and also determines the existence of a saccade by confronting the velocity against an empirical threshold.

Denoising the data

Some apparatus also apply a **denoising** stage in order to eliminate abnormal values. Whatever the internal method employed to provide raw fixation and saccade data, it is judicious to post-process the data in order to clean them up. When observing raw data coming out of the material, one may observe five source of abnormal values that could be reasonably corrected.

First, negative or out of resolution range of fixation locations occur, especially when the user fixated out of the stimuli but on the screen (the stimuli has a ratio inferior to the display, grey or black stripes fill in the remaining pixels).

Second, it may happen that experimenters remove the first fixation of each observer from the collected data; this fixation being related to center bias (natural tendency of human to watch the center of screen).

A third post-processing of data usually consist in removing aberrant range of values for fixations and saccades by defining a minimal and a maximal acceptable thresholds for the two data. For example, fixation duration below 100 milliseconds and above 400 milliseconds are considered atypical [226] and caused by a bad recording of the system or the user fatigue.

Fourth, it is also usual to appreciate the scanpath characteristics. Too short scanpaths testify to a visual fatigue or missing recorded data. A post-processing consists in removing scanpaths having less than a determined number of fixations. In [256], we considered 5 fixations as being long enough for a complete visual processing of categorized images.

A five post-processing may involve statistics of fixations. In order to guarantee consistent data and reduce measurement noise [217] due to abnormal subjects, it may be relevant to remove fixations having duration superior to the fixation duration average $\pm n \times$standard deviation.

Projection to spatial map

Once the raw data have been clean up, the scanpaths characteristics can be statistically treated to end up with a conclusion on attentional mechanisms. Many psychological studies abort at this stage. Nonetheless, a spatial representation of scanpath is also relevant when investigating e.g visual attention modeling: this is the concept of *fixation map* introduced by Wooding [272]. He argued that the processing of eye-tracking data should not be restricted to only pure *eye movement patterns*, but that regions-of-interest, coverage and similarity of maps are also relevant for quantifying eye movement recordings. However, the fixation maps remain computed rather in a context of visual attention modeling.

The fixation locations are projected onto the original stimulus in order to get both *aesthetic* saliency map, for qualitative appreciation, and spatial 2D experimental saliency map, for quantitative appreciation. Note that some metrics are dedicated to the comparison of spatial 2D maps when confronting ground truth to models and others rather process the eye movement patterns (Section 3.6).

From the raw fixation data, the pixel position is formalized by one value in horizontal direction and another one in vertical direction. In order to simulate foveal vision, Wooding [272] approximated the fixation behavior with a Gaussian representation. A 2D isotropic Gaussian filtering is applied on the pixel coordinates of the fixation. Naturally, the Gaussian width follows the fovea specificity (Figure 3.3a): a typical value spreads from 1 to 2 degrees of visual angle, that should also depend on the task and stimulus according to Wooding [272]. Then, this value is mapped in pixel domain according to the experimental setup and the equation 3.1. The contribution of all fixations f are usually considered as being the same, even if intuitively the difference in fixation duration attests to irregularities in the processing of information [100]. Also, the fixations belonging to several observers o for a common stimulus i are usually gathered to produce the final experimental saliency map E_i. The Equation 3.2 integrates the aforementioned concepts.

$$E_i\left(m,n\right) = \frac{1}{O}\sum_{o=1}^{O}\frac{1}{F}\sum_{f=1}^{F} exp\left[-\frac{\left(x_{f,o}-m\right)^2+\left(y_{f,o}-n\right)^2}{\sigma^2}\right] \tag{3.2}$$

At each site (m,n) of the experimental saliency map E with $m \in [0, M-1]$ and $n \in [0, N-1]$, M and N being respectively the width and height of the original stimulus i, the contribution of all fixations f, having their location in (x_f, y_f) are summed up for all subjects o. O and F are respectively the total number of observers and of fixations. σ is the fovea size. Note that a third dimension could be added to the Gaussian representation [272, 209], by setting a specific height to the Gaussian representation; e.g. Pomplum *et al.* [209] vary the Gaussian height with the fixation duration.

For visualization purpose, it is required to normalize the experimental map according to its maximum and the capabilities of the viewer (8 bits - 0 to 255). An interesting discussion, more related to the aesthetic or qualitative appre-

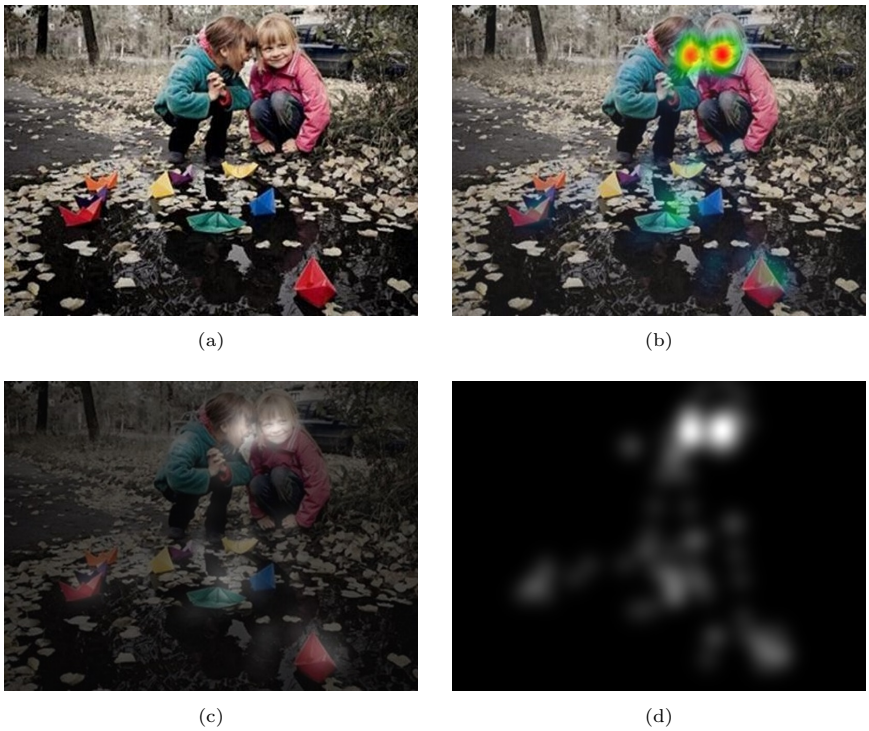


Figure 3.4: Illustration of fixation maps created from eye movement patterns. (a) Original stimulus. (b) Heat map with *hot* spots in red. (c) Original Luminance unmasked by the values in fixation map. (d) Greyscale fixation map.

ciation of experimental saliency map, occurred about the graphical mode of representation [272]. There is no ideal map, this is at the scientist's appreciation to design the map that will better serve his purpose. Original luminance of the stimulus may be unmasked weighted by S, while color heat map may be superimposed on original luminance (Figure 3.4). Also, we can cite flooding map that could be derived into binary region-of-interest map. The map representation used for statistical analysis is usually a one component map (greyscale) having value from 0 to 255 (low to high conspicuous value of pixel). A recent complete review of visualization techniques for eye tracking data may be found in [27].

3.5 Protocols

The setup of a reliable eye-tracking experiment requires to follow a methodology. Like for any experiment types, it is important to provide to the community enough details to demonstrate the validity of the findings, but also to ensure the duplication of the experiment. Previous statement of the literature in the field supports an adaptation of existing protocols and avoids the *faux pas*. Thus,

a number of decision have to be made for the following criteria: stimuli choice, participants, subject's guidance (task), viewing duration. But at first, it is fundamental to formulate a research question.

Formulating a hypothesis

It consists in measuring the effect or impact of a specific variable. Typically, the formulation of a null hypothesis is the reference for having no measured effect under the designed conditions. Having set most of variables to a specific statement, one can have one independent variable to which some manipulations are applied to highlight the targeted effect. In the context of eye-tracking protocol, it may consist in asking for two tasks, or having two (or more) type of stimuli (e.g. image categories) or two categories of observers (e.g. male versus female) and so on. Ideally, the statistical analysis of eye movement patterns would differentiate a behavior collected for each condition or each set of stimuli.

Subject's task

The task asked to the participants is crucial in an eye-tracking experiment. As widely demonstrated in literature [275, 64, 43, 171], the eye movement patterns are task-dependent. Thus, the formulation of the question delivered to participants for a task protocol should be meticulously refined. In addition, it raises the question of having a task or a free viewing protocol. Free viewing of a set of stimuli let the observers exploring the scene with no specific purpose and then less involvement of complex attentional mechanisms. Intuitively, it may lead to a high variability between observers, but it is not that critical, since the top-down mechanisms are minimized, especially for the first seconds of observation. A free viewing pass may sometimes serve as the way to validate the null hypothesis. It does demonstrate an effect when confronted to a second pass with a specific task [237].

Stimuli

An eye-tracking experiment always needs a set of visual stimuli to be presented at each participant. Each of them has to be carefully chosen with ideally a controlled metric that could ensure the dataset consistency. As an example, let us have this original hypothesis: a specific range of spatial frequencies are fast analyzed by cortical cells and cause significant different eye movement patterns. Thus, all involved stimuli should be controlled numerically in terms of visual frequencies, having several pools representing different spatial frequencies to be compared to and no additional factors that could disturb the deployment of visual attention (e.g. objects, face etc.).

Participants

The choice of participants may bias the results of an eye-tracking experiment. The age [7, 98] and gender (first category) as well as the personal experi-

ence/culture and mood (second category) may influence the deployment of visual attention. Age and gender attest to different physiological cells status or organization which lead to a variability in bottom-up attention. Potentially also dependent to these aforementioned factors, the mood and personal experience would rather cause a bias in top-down attention, since the purpose of people is variable regarding this criteria. For example, a person being afraid of water, would have tendency to fixate on water part of the stimuli if any, while this area will not attract the attention of another one. However, this second category of parameters are difficult to control; it seems tricky to ask a person for all details of his past life. Nonetheless, the first category of participants criteria is straightforward to measure. But in which extent should all participants have the same age and gender. If only men between 20 and 30 years old participate to the experiment, the significant measured effect may be associated to only this class and not generalized to the entire population. From another angle, the specific and unpredictable behavior of a category could generate noise in the analysis of the complete pool of observers.

Additionally, the number of participants of an experiment is also a point that should not be neglected. Since the collected eye data will be processed to find out statistically significant effects, a minimal number of participants is required to ensure the validity of any statistical tests (t-test, ANOVA...) when performing a power analysis. There exist some analytic demonstration to determine the minimal number of participants when setting a fixed power. More details may be found in book addressing this topic [219].

Viewing duration

In eye-tracking experiment, the viewing duration per stimulus should be considered carefully. Depending on the tested factor (related to bottom-up or top-down mechanisms), this value should be adjusted. The typical viewing time associated to the stimulus presentation ranges from 2 to 10 seconds, also depending on the task. There is no clear rules for the perfect time to be associated to the specificity of an experiment. The longer the presentation, the more variable the scanpaths between participants. Usually, the viewing duration is fixed during the design stage of the experiment, but the statistical post-experiment analysis will guide the selection of an appropriate subset of data within this total viewing duration.

Rule of thumbs

In addition to previous topics, we would like to remind a bunch of rule of thumbs. There might look obvious, but "who can do more, can do less". A *black* room with no exterior light coming in is advised as well as a noise free environment. No external stimuli should disturb the participants. Also, it is common sense to limit the total time of the experiment. Usually, we consider that 20 to 30 minutes is a maximum duration before getting the participants bored and tired. A way to maintain attention to the eye-tracking task is to ask the participant

for a fictional question between some series of stimuli. Last but not least, it is also important to get information about the correction of participants; if they can use it for the experiment and if they do not have vision abnormalities (color blindness...). Note that we experienced difficulties to track people with multifocal lens.

Even if most of remote material claim for compensating head movement, it is wise to use a chinrest in order to stabilize the subject's head when performing the experiment.

3.6 Metrics

In visual attention modeling, it is conventional to confront the *computational* saliency map C to the *experimental* saliency map E obtained for one stimulus i. To do so, several similarity metrics, dedicated to this context or not, are regularly employed in the literature [143]: Linear Coefficient Correlation (CC), Kullback-Leibler Divergence (KLD), Receiver Operating Characteristic Analysis (ROC), Normalized Scanpath Saliency (NSS). They provide a score of (dis-)similarity between the maps and then, they assess the relevance of the prediction for the considered model. These *attention map* metrics are a first family of metrics. We can cite two other families of metrics: the *string edit* based methods and the *geometric* methods. These two latter manipulate eye movement data rather than spatial maps; thus, they take advantage of the temporal aspect of visual attention deployment. String edit methods focus on the order or temporal arrangement in a sequence of fixations, while the geometric methods exploit other dimensions inherent to the scanpaths, such as fixation duration, shape etc.

In addition to the aforementioned map-oriented metrics, we can cite the main contributions for these two families: String Edit method [147], Mannan's metric [164, 165], Cristinao's metric (ScanMatch) [60], Mathôt's metric (Eyenalysis) [167] and the vector-based similarities [99, 119, 68].

The String Edit method (based on Levenstein distance) maps a fixation into an Area Of Interest (AOI), translates the scanpath information as a string of symbol (having a meaningfull sequential order) and assess the similarity between two sequences by computing the *editing cost* through *insertion, deletion, and substitution* of characters.

The ScanMatch metric builds on the Levenstein distance and improves the alignment of strings [60]. However, the limitations raised by the community regarding the mapping within an AOI is still present in this metric. Two fixations associated to one AOI would be always more similar than if they were in two different AOIs but spatially closer when measured with e.g. an euclidean distance.

Mannan's metric attempts to escape the AOI issue. It basically computes the euclidean distance between one fixation location and its nearest neighboring fixation, building a *similarity index*. However, the metric remains also questionable, mainly about the potential mapping of several fixations of the first

scanpath to the same fixation in the second compared scanpath. This may overestimate the impact of the candidate fixation in the second scanpath. Also, comparing two scanpaths of different length is an issue.

The Mathôt's metric try to overcome these two problems. The vector-based method is a new class of promising method that compares the spatial and temporal characteristics of two scanpaths through their arrangement in geometric vectors, later analyzed in multiple dimensions. They take into account *shape, position, duration and order* in their similarity metric.

We will not detail more all these contributions, but we will focus on the Mathôt's metric (Section 3.6.5) that we employed in our future analysis. Since we designed some computational methods that produce spatial maps, we have focused on metrics that deal with a score of spatial similarity.

3.6.1 Correlation Coefficient

The most straghtforward similarity measure is the Pearson Correlation Coefficient. It expresses the linear relationship between two set of variables, by using the covariance between the sets and the standard deviation of each set, such as expressed in Equation 3.3. C and E are the two set of values (e.g. the saliency map coefficients), $cov()$ their associated covariance and σ_C and σ_E stand for respectively the standard deviation of C and E.

$$S_{CC}(C,E) = \frac{cov(C,E)}{\sigma_C \, \sigma_E} \tag{3.3}$$

This measure ranges from -1 to 1, where -1 means a total dissimilarity, 0 no similarity and 1 a total similarity between the two set of samples. This measure is not sensitive to scale, thus the compared maps do not need to be normalized. However, singular aberrant values influence considerably the score, even if most of values are consistent. The interpretation of such metric may be tricky in the context of saliency map. This is not rare to end up with singular fixations (due to visual fatigue, top-down mechanism or noisy recording). Nonetheless, we expect their contribution to be averaged and minimized when creating the experimental saliency map.

3.6.2 Kullback-Leibler Divergence

Complementary to the correlation coefficient, the Kullback-Leibler Divergence expresses a dissimilarity between two probability distributions. It can not be considered as an *effective* distance measure (mathematically spoken), because it is intrinsically asymmetrical. Considering the notations explained below: $S_{KLD}(C,E) \neq S_{KLD}(E,C)$. Indeed, in information theory, a model or a reference distribution is confronted to observation samples or an experimental distribution. Thus, it is expressed as the divergence of an observed distribution D knowing the theory distribution T: $S_{KLD}(D|T)$. In our case, the reference distribution and the compared samples are respectively the experimental saliency

map E and the computational saliency map C, such as formalized in Equation 3.4:

$$S_{KLD}\left(C|E\right) = \sum_{i} C(i) ln\left(\frac{C(i)}{E(i)}\right) \tag{3.4}$$

Note that the two-dimensional saliency maps may be transformed into a one-dimension probability function or kept as a 2D distribution to maintain the spatial information, where i is the current pixel location being processed. Also, the two probability density have to be normalized to 1, meaning each pixel value is divided by the sum of values, in order to satisfy the constraint on probability density (the distribution area should be equal to 1). KLD is always positive (0 means the two probability distributions are the same), but has no upper boundary. This is a strong limitation in the interpretation of results. This point is discussed also in the context of the Eyenalysis similarity (Section 3.6.5).

3.6.3 Receiver Operating Characteristic Analysis

Originally, this metric is employed to assess the performances of a binary classifier, knowing a thresholding step is required. Usually, the Receiver Operating Characteristic (ROC) Analysis provides a curve with the true positive rate as a function of the false positive rate. The points fitting the curve are obtained for different values of the mentioned threshold. Good performance of the measured data happens when the curve stands in the upper left section (tends to True Positive Rate (TPR) = 1 and False Positive Rate (FPR) = 0).

In the context of visual attention map, these metric is popular and has been adapted to the non-binary nature of saliency maps. The true positive (TP) and true negative (TN) cases correspond to an agreement between the two maps: saliency has a label equal to 1 if both pixels are fixated (or to 0 if none are fixated) in each map. The other way around for the false negative (FN) and false positive (FP) cases: if one label equal to 1 (salient region) and the label in the second map is 0, there is no agreement. Also, the two saliency maps are simultaneously binarized with a varying threshold. If the saliency value exceeds the threshold, the pixel is labeled at 1, this for different values of threshold at the appreciation of the scientist. To build the ROC curve, the True Positive Rate and the False Positive Rate are expressed respectively as: $TPR = \frac{TP}{TP+FN}$ and $FPR = \frac{FP}{TP+FN}$.

At this point, we do not have a numerical value to be appreciated for assessing the prediction quality of any model. Conveniently, the ROC curve is usually translated into a unique value; more precisely the Area Under the ROC Curve (AUC) is a good indicator of performances [78]. The higher the area, the closer the curve to the upper boundary: a value of 1 means a perfect prediction, while 0.5 means the chance level. The area may be computed with different methods; e.g. the well-known method of Riemann sum consist in extrapolating shapes, e.g. a rectangle, under the curve, from true or approximate points of the ROC curve and then to sum the rectangular area of each shape.

This metric provides an interesting interpretation of results, since the AUC indicator has a theoretic and meaningful limit. However, the reliability of provided results may be influenced by an inadequate setting of parameters. The thresholds used in the ROC curve design is a tricky point that has been discussed in the literature, especially for the application in saliency map [249]. Also, the fitting model to extrapolate ROC curve, if not many samples, may have a negative effect. Finally, the method to compute the AUC is also a ground where an approximation is done.

3.6.4 Normalized Saliency Scanpaths

Normalized Saliency Scanpaths was first introduced by Peters *et al.* [205]. It can be apprehended as a mixed method, because it compares the eye movement data (from the experimental ground truth) to a saliency map (most likely from a computational model). It has the advantage to normalize the salience per scanpath: scanpaths with different number of fixations have the same weight. In other words, every observer has the same impact on salience. Moreover, the NSS gives more weight to areas more often fixated.

The NSS is the average value of the saliency map (represented by a set of pixels p) at each fixation normalized per scanpath. First, the saliency map C is normalized to have zero mean and unit standard deviation, such as described in Equation 3.5, where μ_C and σ_C are respectively the mean and standard deviation of map C. Then, the per-fixation salience value NSS_f is computed as the average of the saliency map on the projection of the fixation. A disk D with a radius r of 1 to 2 degrees of visual angle is usually used to project each fixation p_f (Equation 3.6). Per-observer o NSS is computed as the average of per-fixation salience value along the scanpath. The final NSS measure S_{NSS} is the average of the per observer NSS (Equation 3.7).

$$C_{norm}(p) = \frac{C(p) - \mu_C}{\sigma_C} \tag{3.5}$$

$$NSS(p_f) = \sum_{p \in C} D_r(p_f - p)\, C_{norm}(p) \tag{3.6}$$

$$S_{NSS}(o, f) = \frac{1}{O} \sum_{o=1}^{O} \frac{1}{F} \sum_{f=1}^{F} NSS(p_f) \tag{3.7}$$

F is the total number of fixations for an observer o and O is the total number of observer. A negative value of NSS means a low correlation, i.e. the distance is large between the considered saliency map and the averaged saliency maps. A positive value of NSS means a good match.

Interestingly, all these mentioned metrics could be used to assess the inter-congruency between observers [143]. Thus, by comparing the scanpath (or saliency map) of a specific observer t (with $o \neq t$) to a global saliency map extrapolated from all fixations of all the other observers SM, the likeness of an

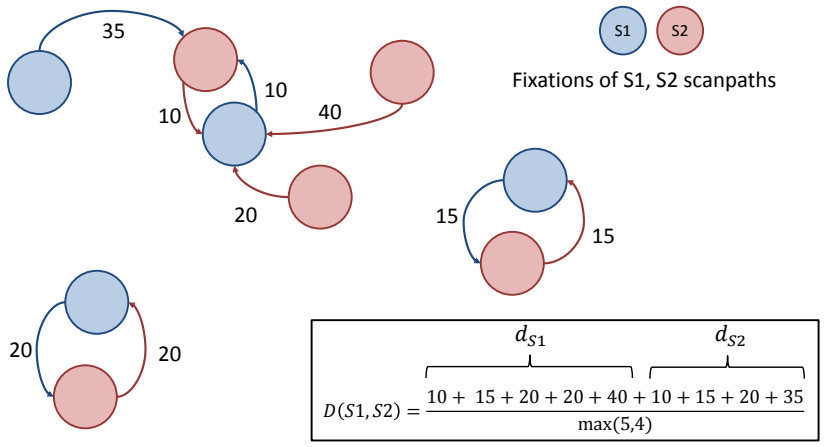


Figure 3.5: Illustration of the double mapping technique between two scanpaths. Also, the similarity between the twoscanpaths is computed from Equation 3.9

observer scanpath (or map) to those of a group of individual is assessed. These is also named *leave one out* technique [253].

3.6.5 Eyenalysis Similarity

This method is purely based on eye movement data and belongs to the geometric family of methods. To circumvent the limitations of similar method, Mathôt *et al.* [167] proposed the Eyenalysis metric.

Let us consider two scanpaths, S1 and S2, the idea of the method is to build a reciprocal mapping between S1 and S2, i.e. each point u belonging to S1 is mapped to at least one point v within S2, and reciprocally. Then, the sum of distances related to all mappings is minimized in order to obtain the optimal mapping, meaning the optimal pairs between S1 and S2 regarding a distance d measure (Equation 3.8). Note that in essence the mapping may deal with any properties i (maximum is P) of the scanpath.

$$d(u,v) = \sqrt{\sum_{i}^{P} (u_i - v_i)^2} \tag{3.8}$$

There is no known solution for such kind mapping. The authors went for the double mapping technique (Figure 3.5) which is a good compromise in terms of complexity and efficiency. Using the Euclidean distance, each point from S1 is mapped to its nearest neighbor in S2 and reciprocally. Thus, some mapping reveals common for some points in both directions, as illustrated in Figure 3.5. Such technique may the advantage of mapping many points to an unique one, but as compensation, it allows mapping of sequences with different lengths. A global distance is defined as the normalized sum of all the point-to-point

mappings (Equation 3.9):

$$D(S1, S2) = \frac{\sum_u^K d_{S1 \to S2_{v'}}(u) + \sum_v^L d_{S2 \to S1_{u'}}(v)}{max(K, L)} \tag{3.9}$$

where $d_{S1 \to S2_{v'}}(u)$ is the Euclidean distance of point u in S1 to the nearest neighbor v' in S2 and $d_{S2 \to S1_{u'}}(v)$ is the euclidean distance of point v in S2 to the nearest neighbor u'. K and L are the total number of points in respectively S1 and S2.

This similarity metric do not have an asymptotic limit, such as the NSS and KLD. Consequently, the values have no meaning (except the higher, the better) if considered alone. Usually, the average similarity of a first set of scanpaths should be compared to a second set, formalizing a reference for testing the original hypothesis (validating or unvalidating the null hypothesis), such as suggested by the authors [167].

3.7 Summary

Actual solutions to setup an eye-tracking experiment gain in practicality and propose software packages to post-process the data at an accessible price. Thus, the eye-tracking technology becomes more popular and is employed in a broad field of applications: marketing, webpage design, human-computer interface, gaming, mobile scrolling and so on (Section 3.1). Not only restricted to the communities of psychology, neuroimaging, ophthalmology, any experimenters may either define a set of stimulus to be analyzed in the friendly software interface or uses a purely image processing algorithm to get informed about raw gaze position of observers [182].

However, good practices exist and inexperienced user has to follow rules of thumb if one wants to exploit statistically the collected eye movement patterns (Section 3.5). Different outputs may be exploited from an eye-tracking solution. While fixations and saccades are directly measured by the system (Section 3.3), experimental saliency maps may be derived from such raw data (Section 3.4). These latter provide qualitatively a first impression of the regions watched by the observers. More quantitatively, they also are confronted to computational saliency maps coming from visual attention models by means of similarity metrics (Section 3.6). Interestingly, the metrics could be used to assess the inter-congruency between observers by comparing the scanpath of a specific observer to a global human saliency map extrapolated from all fixations of all the other observers.

Chapter 4

Color Harmony

Color harmony, and more precisely the rules surrounding its definition, understanding and modeling, has occupied the time of many artists and scientists for the last centuries. Despite this substantial energy applied to formulate the problem of color harmony, the concept is far from encountering a consensus in the *definition, representation and modeling* domains. Although the historical theories are based on intuitive observations, specific color space and representation, they provided the necessary basis for designing later some models. These latter supplied the community with a comprehensive implementation for being applied in concrete applications.

We tentatively expose in this chapter the accomplished work by dividing the chapter into four different sections related to the future development of this thesis. First, the origins of color harmony are introduced to provide an overview of historical sources (Section 4.1). Second, we propose a non-exhaustive list of models, classified into three classes (Section 4.2). Third, we rather focus on the applications in image processing and editing that directly implement the cited models (Section 4.3). Finally, we summarize the main points (Section 4.4).

4.1 The origins

When investigating the notion of Color Harmony, we have some questions in mind: **what is the substance behind *Color Harmony*?** Can it be classified as a hard, pure and exact science? or Can it be described only as purely empirical, intuitive and experimental? When searching for a strict definition of Color Harmony, we started answering the question. Artists, psychologists, color scientists and mathematician built and enriched the notion of Color Harmony. Everyone has an idea, an own feeling of what is the color harmony in his/her surrounding environment, but finally we ended up with a bench of definitions.

Later, our investigations concerning the history, the theories and the potential computational models evidence the following fact: depending on the original sensitivity composing the point of view, the color harmony definition

flirts with the notions of aesthetic, emotion, artistic, preference, culture. Despite this statement, we found out full of sense and fundamental experiments that lead to the modeling of color harmony.

In conclusion to this reflexion, Color Harmony suffers from limited but reasonable investigations and needs new psychological breakthroughs to better understand the involved attentional mechanisms. Even if some attempted to design algorithms and automatic solutions to benefit from the concept, this is still a conceptual notion that need some clarifications to be considered as a promising scientific field.

4.1.1 Definition(s)

All people involved in Color Harmony experiments target the same Grail: *How are colors perceived together? How do they interact? How do they harmonize?* [234]. Those questions have been the starting assumptions for many researches in the field and highlight one fundamental point: the color cannot be perceived alone, it is part of a global perception. This follows the findings on how the human visual system typically behaves: color variation or contrast determines our perception of the environment (Section 2.2.1).

Integrating such reflection, the notion of *complementarity* emerged to refine the definition (Section 4.1.2). Indeed, the complementary colors intuitively reflects the notion of "most beautiful as they together form a whole" [234]. Here, the definition is refined, but having exposed this statement, have we really step forward? There are still many objective formulations of what complementary colors could be.

Another limitation for investigations shows up at this point: how many colors to be involved? which attributes of colors? Quantitatively measuring colors is already a large research problem, that may suffer from adding the notion of color harmony as a new variable.

Beyond their numbers, also the question of color representation must be raised. Can a circle, volume, any spatial, geometric representation support the findings and definitions of color harmony?

Facing the difficulty of characterizing the notion of color harmony, Granville [87] and Kuehni [134] brought another argument related to contingency: age, gender, cultural heritage and mood probably influence the perception of color harmony of any observers.

Despite that, the definition of color harmony raises also the question of involving aesthetic or not. Historically, the aesthetic notion was at the heart of the definition, thus scientists wondered whether color harmony is related to the pleasing use of colors in a specific artistic analysis or if it is only a matter of combining colors [87].

Below is a bench of definitions annotated along the different readings:

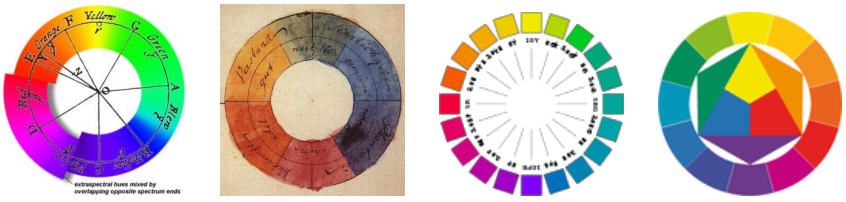
- Burchett [37], p.28: *Colors seen together to produce a pleasing affective response are said to be in harmony*

- Arnheim [13]: *Visually right or wrong*
- Granville [87]: *Color harmony is color usage that pleases people*
- Judd and Wyszecki [122]: *When two or more colors seen in neighbouring areas produce a pleasing effect, they are said to produce a color harmony.*

It seems that a positive feeling and a homogeneous and balanced arrangement are the key concepts behind these definitions.

4.1.2 Theories and Color wheels

Complementary to his work on electromagnetic waves, Isaac Newton (1643-1727) plays a key role in the achievement of color theory by originating the color circle (Figure 4.1a). He identified seven distinct spectral colors that he placed on a hue circle. Thus, he invented the geometric color models (also referred as *Hue Wheel*) that is the foundation for many models of color harmony (Section 4.2.1). By just looking at the color spectrum, he observed the logical circular arrangement of hues and also evidence the principle of light mixtures, i.e. the yellow results from mixing the red and green light sources.



(a) 1611 (b) 1810 (c) 1905 (d) 1961

Figure 4.1: Color wheels designed along centuries. (a) Newton's color circle. (b) Goethe's color wheel. (c) Munsell's color wheel. (d)Itten's color wheel.

Opposed to Newton's theory, Wolfgang Goethe (1749-1832), poet and philosopher, refused strongly that science and mathematics interfere in the theory of color. However, he contributed to the problem by conceptualizing the notion of complementary colors [84]. Interestingly, Goethe anticipated the well-known findings of Hering about *opponent process* theory (Section 2.2.3). By studying the physiological response to opponent colors, he designed a symmetrical arrangement of the colors on his wheel (Figure 4.1b): yellow opposed to purple, green to red and orange to blue. Eugène Chevreul (1786-1889) [52] also investigated this aspect and highlighted that complementary colors are more easily distinguishable than adjacent colors that interfere with each other [267].

This disagreement between communities on complementary color and primaries led to a bench of color wheel representation (Figure 4.1). Following Newton's original idea and the current findings on subtractive or additive primary colors, scientists and artists tended to accurately design color wheel whose the

properties are the number of primaries and their proportion on this circular representation.

The concept of color harmony has been explored through different approaches which accompanied some adjacent scientific discoveries. In other words, each implementation also is the reflection of the period. As mentioned previously, this is a field where opposite communities confronted each other: painters and color scientists, chemists, psychologists...

The vast developed theories implement different primaries, color components relying on different wheels or volume and consider different kinds of harmony measurement. However, we can extract common principles and methodologies between them. First, they manipulate three color components; while two of them remains fixed, they measure the produced harmony when having the third one varying along a line or a plane. Second, they all identify the harmony as a contrast or a ratio between the employed primaries. Third, they also mention the spatial location on the hue wheel as being a key harmony identifier.

More particularly, we are focusing on Munsell's and Itten's theories, since they would be the foundation of harmony models developed in the next section.

Munsell's Perceptual Harmonies

Munsell's vision of Color Harmony deals with the idea of *balance*. He intuitively suggested the artists to balance the color strength and size in their composition. Thus, he defined color strength as the product of value and chroma in his Color System [177]. Then the relationship between color strength (CS_n) and color area (A_n), where n is the considered color, is defined as:

$$\frac{A_2}{A_1} = \frac{V_1 \cdot C_1}{V_2 \cdot C_2} = \frac{CS_1}{CS_2} \tag{4.1}$$

where A, V and C are the area size, value and chroma (in Munsell system) of colors 1 and 2. Note that the area ratio is inverted against the color ratio: the color with the weaker color strength is assigned to the larger visual area. The Munsell's criterion of harmony has been used widely [172] and generalized in such way:

$$\sum_{n=1}^{M} CS_n \cdot A_n = 0 \tag{4.2}$$

where M is the total number of colors in the image, CS_n (the color strength of color n) = chroma of color n x value of color n, A is the area of color n.

Itten's Contrasts

Another popular approach is the onne originated from Johannes Itten in 1916. His perception of color harmony took inspiration from the contrast notion. For him, everything is always a matter of contrast. A high color harmony must be the fact of a high contrast. However, this is not straithforward to design which kind of contrast is relevant in the context of color harmony. He defined

seven relevant color contrasts, as listed below, where ΔC stands for a contrast between two quantities i and j in HSV color model ($h \in 0° \dots 359°$, $s \in \Re$, $0 \leq s \leq 1$ and $v \in \Re$, $0 \leq v \leq 1$). The three first one relates to the contrast on the components he used (HSV color space).

1. Contrast of value: $\Delta C_v = v_i - v_j$
2. Contrast of saturation: : $\Delta C_s = s_i - s_j$
3. Contrast of hue: : $\Delta C_h = h_i - h_j$
4. Contrast of extension (or proportion): related to the size of each compared color in the picture. According to Goethe and Itten, the perfect proportions are the followings: Yellow: 9; Orange: 8; Red: 6; Violet: 3; Blue: 4; Green: 6; as illustrated in the work of Sauvaget *et al.* [227] in Figure 4.5.
5. Contrast of warm/cool: formed by the opposition of hues considered *warm* or *cool*, where $\Delta C_{cw} = \frac{\omega_i - \omega_j}{180°}$ and $\omega_i = |\omega_i - 180°| = (h_i + 60°) \ mod \ 360°$
6. Contrast of complements: when two opposite hues on the hue circle are confronted, where $\Delta C_c = \frac{k_i - k_j}{180°}$ and $k_i = (h_i - 180°) \ mod \ 360°$
7. Simultaneous contrast: when two colors attract the other one to its complement. It creates illusion of motion or depth within the picture.

These contrasts rely on the hue wheel purposely designed by Itten (Figure 4.1d). Such as Munsell, this theory influenced number of colorists, designers. Some image processing algorithm translated the proposed contrasts into computational method for color harmonization [227].

4.2 Models of Color Harmony

Some models emerged from the middle of the 20^{th} century. We propose to classify them into three categories: the *geometric* model (first introduced by Moon and Spencer [172]), the *numerical* models and a new contemporary approach, the *contingent* models.

The geometric approach follows the historical wave by taking advantage of color wheel concept and by numerically formalizing their use (Section 4.2.1). On the other hand, the numerical approach tackles the problem completely empirically. This last category rather perform user experiments to characterize the degree of color harmony encountered by their stimuli; they are then able to **model, predict** and finally **generalize** the color harmony concept (Section 4.2.2). The reliability of models is discussed in Section 5.3.

We differentiated these models from the historical color harmony theories (such as the Munsell' and Itten's one) based on two observations: first, these models have been published quite recently (after 1950); second, they have a strong computational aspect that makes them realistically applicable to image

processing (Section 4.3). Relevant to this purpose, Sivik and Hard wrote in "Some reflections on Studying Color Combinations" [234]: *The dream was to be able to express the relationship between stimulus and experience in a mathematical formula with few variables as possible.*

In the last decade, even a new conceptual approach for color harmony modeling appeared; it translates new progresses made by the psychology field. This contingent-based modeling is described in Section 4.2.3.

4.2.1 Geometric model

Moon and Spencer, 1944

The mathematicians Moon and Spencer [172] decided to apply their knowledge to aesthetics and more precisely to the topic of color harmony, originally formulated by Munsell [177]. They proposed a mathematical interpretation of Munsell perceptual harmonies (Section 4.1.2) that able to predict by a number the aesthetic value of any color combinations. Their concept or model relies with the notion of **ambiguity**. They considered that if the viewer is *confused* by the color combination, then the feeling of color harmony is low. In other words, their model determine the harmony likeliness in an *ambiguous interval* (or color difference). More precisely, they differentiated two cases of ambiguity: 1/ when the subject is uncertain if the colors are *identical* or *similar*; 2/ when he hesitates whether two colors are *similar* or *contrasting*. They refined the Munsell equation of proportion (Equation 4.1):

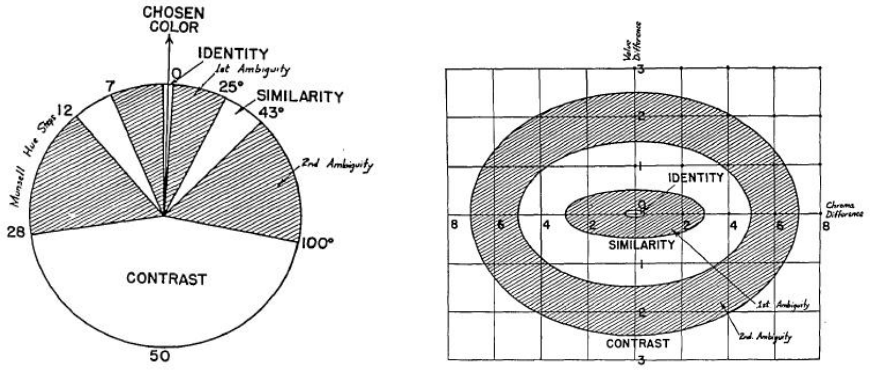
$$\frac{A_2}{A_1} = \frac{\sqrt{C_1^2 + 64 \cdot (V_1 - 5)^2}}{\sqrt{C_2^2 + 64 \cdot (V_2 - 5)^2}} \tag{4.3}$$

They proposed also an understandable schematic representation of color harmony areas based on the aforementioned concepts and illustrated in Figure 4.2. Their geometric model is applicable to hue, lightness and chroma and for a chosen color defines quantitatively four areas on the hue wheel: *Similarity*, *Identity* and *Contrast* areas are considered harmonious, while *Ambiguity* area is disharmonious. Thus, their geometric model outputs a binary decision/prediction (harmony or non-harmony) when combining at least two colors.

Moon and Spencer acquired a lot of attention in the field, since they scientifically proposes an applicable model for harmony prediction. Despite that, the weak theoretical basis and the lack of relevance in the definition of their concept was pointed out. However, their digest model for predicting color harmony is still nowadays inspiring image processing techniques and user interfaces (Section 4.3).

Complementary to this work, Moon and Spencer derived a formula for rating color harmonies which is perceived more as an aesthetic measure (M) [173]:

$$M = \frac{O}{C} = \frac{O}{N_C + N_{hdif} + N_{vdif} + N_{cdif}} \tag{4.4}$$



(a) (b)

Figure 4.2: Moon and Spencer's color harmony model. Geometric model and representation for hue, lightness and chroma (expressed in the Munsell color system). *Similarity*, *Identity* and *Contrast* areas are considered harmonious, while *Ambiguity* area is disharmonious. Original illustrations are extracted from [172]. In the original cylindrical color volume, two surfaces are extracted: (A) Regions of similarity and contrast in the plane with constant value, (B) Regions of similarity and contrast in the plane with constant hue.

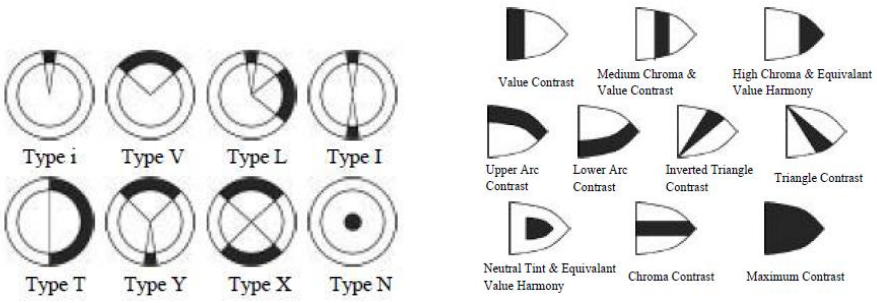
where O is the number of elements: the color identity (basic color category), similarity, contrasts of hue, value and chroma. C is the factor of complexity where N_C is the number of colors, N_{hdif} the number of pairs with hue difference, N_{vdif} the number of pairs with value difference and N_{cdif} the number of pairs with chroma difference.

Matsuda, 1995

In the same vein as Moon and Spencer, Matsuda Color Coordination [168] is a geometric method that formulates color harmony rules in Munsell color space. By gathering continuously the questionnaires of students during nine years, he could design his model on specific stimuli, mainly print clothes and dresses. However, the applicative field has been so largely extended to other stimuli type, that it seems to be adopted for any kind of content. A non-exhaustive review of application in image processing and computer graphics using this model is provided in Section 4.3.

Matsuda's method goes one step further than the modeling of Moon and Spencer. Although he also predicts harmonious and non-harmonious areas using Munsell color space, he proposes eight geometric representations for hue distribution (Figure 4.3a) and ten for tone distribution (Figure 4.3b), that will be called **color harmony templates** later by Cohen-Or [54]. Matsuda rather formulates it as *color scheme*. A color scheme expressed the relationship between hue and tone components, then there are a total of eighty color scheme types.

In his experiments, Matsuda observed the geometric relationship of combinations, already foreseen by Munsell and Itten (Section 4.1.2): opposite, similar



(a) (b)

Figure 4.3: Matsuda's color harmony model. Geometric model for hue and tone components in the Munsell color system: (A) The circular types of Hue distribution (B) The polygonal types of Tone distribution. Black areas are the considered harmonious parts. (Original illustrations extracted from Tokumaru [252])

and orthogonal arrangements are the key features of his color schemes.

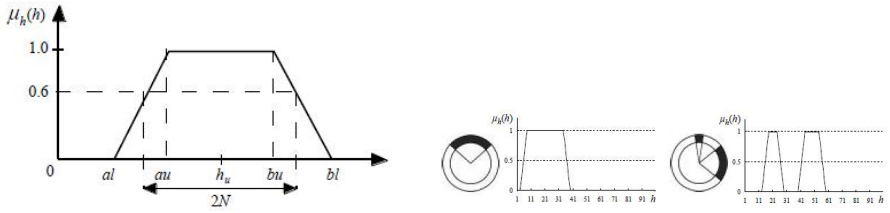
Tokumaru, 2002

In 2002, Tokumaru [252] proposed a comprehensive version of Matsuda's color schemes (Figure 4.3). In addition to the english translation of the original japanese text, he defined membership functions for 1D mathematical distribution. It has opened the way for many interpretations and implementations of the Matsuda's color schemes in image processing. His original intent was to build a system for supporting color design based on harmony consideration and user's word.

To do so, he employs fuzzy logic that purposely catches the degree of uncertainty around the Matsuda's color schemes. Indeed, he defines membership functions for each color scheme type describing the (non-)harmonious areas. For the hue distribution, one or two trapezoids represent the color schemes depending on the number of sectors in the original color scheme. Figure 4.4 depicts two membership functions about the hue distribution for a color scheme of one area or two areas. For the tone distribution, their curved shape (Figure 4.3b) makes them more difficult to be represented by a membership function. Despite that, Tokumaru proposes a membership function about tone distribution for nine color schemes.

When a user provides an input color C of components (h_u, v_u, c_u) standing for the hue, the value and the chroma, the membership function, (a trapezoid following the design of Matsuda, with a sector width equal to N and the characteristics expressed in Equation (4.5), in the case of h_u), gets align on C. Then, the system evaluates the considered color scheme by applying the fuzzy rules of Equation (4.6):

$$\begin{aligned} |au - al| = |bu - bl| = 5 \\ \mu_h\left(h_u + N\right) = \mu_h\left(h_u - N\right) = 0.6 \end{aligned} \tag{4.5}$$



(a) (b)

Figure 4.4: Tokumaru' membership functions about the hue distribution for the Matsuda's color schemes. Note that in the Munsell color space, hue goes from 1 through 100, values from 1 through 10 and chromas from 1 through 16. (A) Trapezoid and its notations for defining the membership function. (B) Two examples of trapezoid centered on the hue 21. (Original illustrations extracted from Tokumaru [252])

$$\mu_p = \mu_h\left(h_i\right) \wedge \mu_t\left(v_i, c_i\right) \wedge \mu_t\left(v_u, c_u\right) \tag{4.6}$$

μ_p is the final evaluation, while μ_h and μ_t are the local evaluation or the membershift function for the hue and tone distribution. The index u stands for the color inputted by the user, while the index i represents all colors in the Munsell database.

Sauvaget, 2010

Sauvaget and Boyer [227] proposed more recently a harmony formulation based on Itten's contrasts (Section 4.1.2) and Goethe's theory [84]. Relying on the Itten's color wheel, formed of 6 distinct hues (red, orange, yellow, green, blue and violet), they exploit ideal color proportions between hue areas; ideal meaning that engage the most in a harmony feeling (Figure 4.5, bottom left wheel). Thus, for example the optimal harmonious ratio between yellow and violet is 1:3. When transferring this principle onto the 6 distonct hues located on the color wheel, the optimal proportions are: 60 degrees for red, 40 degrees for orange, 30 degrees for yellow, 60 degrees for green, 80 degrees for blue and 90 degrees for violet. Interestingly, Sauvaget's model allows analyzing the reciprocal harmony from a subset of hues, from 2 to 6 hues. In other words, the harmony analysis can be considered for only two colors, if the user wishes it.

In addition to the formulation of color harmony from Itten's contrasts, they proposed to perform an automatic color harmonization of pictures. Their complete framework consists in 1) measuring the original hue proportions of the considered image following the defined range for each hue and 2) reassigning over-proportion of pixels to the most *correct* sector to tend to the ideal proportions for each sector. This is illustrated in Figure 4.5.

They initialized what are the usual range for the 6 distinct hues on the color wheel, based on mean values provided by users. Thus, red stands from 340 to 10 degrees, orange from 10 to 40 degrees and so on, such as illustrated on the right-hand side wheel of Figure 4.5 (dotted lines).

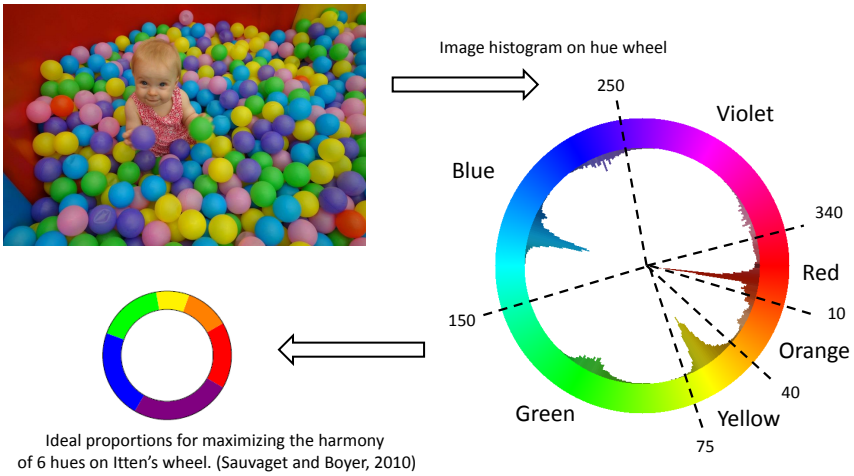


Figure 4.5: The harmonic model of Sauvaget and Boyer [227]. The main idea is to tend to ideal hue proportions (left-hand side wheel). The right-hand side wheel depicts the hue histogram of the picture and the proportion (pixel histogram) for each hue sector.

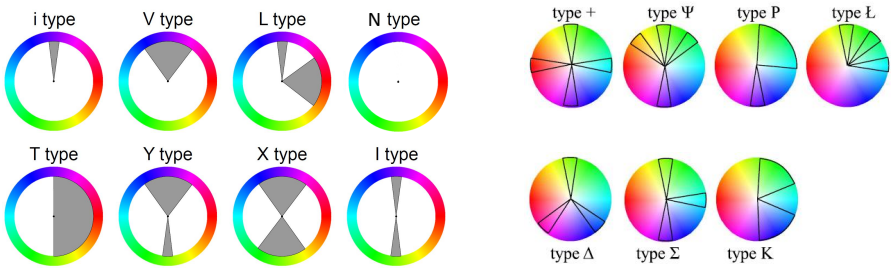
Large dataset exploitation

With the arrival of new social websites dedicated to color theme design, such as COLOURLovers [2] and Adobe Kuler [1], huge communal dataset of beautiful, harmonious, aesthetic and consistent color palettes have been made accessible to anyone. Those platforms propose to explore, create and share color themes. No expertise level is required to create a theme, however the community is able to rate the color themes. Thus, hundreds of thousands themes are available by means of XML API and updated every day by thousands of users.

While such powerful and social information are riding a wave of success, O'Donovan *et al.* designed a machine learning based model to predict color compatibility [194]. They took advantage of the huge number of available color themes to learn a model, but also to evidence potential bias or limitations of existing color harmony theories (discussed in Section 5.3). They rather preferred the term *Color Compatibility* to *Color Harmony*, but this can be considered as very similar notions.

In the same vein, Skurowski *et al.* [235] investigated such huge dataset to derive new Matsuda-like templates. By means of fuzzy logic and clustering, they could aggregate the color themes onto the hue wheel and removed noisy samples. On the contrary of [194], they finally confirmed the use of original Matsuda's templates (Figure 4.6a) and end up with new templates design (Figure 4.6b).

Recent studies kept working on the large dataset track in order to *discover* the *underlying principles* that constitute color harmony and then infer high aesthetic score [155, 157]. Following the approach of O'Donovan *et al.* [194], they also aimed at finding the relevant color combinations that predict aesthetic scores, but went one step further by applying such hierarchical color harmony model on natural pictures. In [156], the same authors deal with this approach in



(a) (b) Extracted from [235]

Figure 4.6: Large dataset exploitation has introduced new color harmony schemes. (a) Modern representation of Matsuda's templates, (b) New templates introduced by [235].

depth and proposed a Bayesian framework that reconciles the empirical theories and the large scale approach. Indeed, the Matsuda's templates as well as the Moon and Spencer theory are employed as a prior while the color combinations learned from the large dataset are injected as the likelihood.

4.2.2 Numerical model

Previous section introduced the geometric models that tend to define geometric patterns on color wheel that demonstrate harmonious and disharmonious areas. In this section, we are interested in the model whose the authors use the word, *quantitative*. We rather propose to name them **numerical model** due to their computational nature.

Sivik and Hard [234] as well as Burchett [36] introduced the typical attributes that could directly derive a color harmony response. From their reflections, they strongly encouraged scientist in investigating their design. Carefully, they mentioned the foreseen problems, such as the choice of color stimuli and the potential bias related to context, time, fashion, culture...

Ou and Luo, 2006

The Ou and Luo's model [195] is well-known of the color science community. It was precursor in the modeling of color harmony extrapolated from user experiments with controlled stimuli. The problem of infinite number of colors, conditions discouraged quantity of scientists in their investigation of color harmony, such as pointed out by Sivik and Hard [234]. Due to progress in modern color science, especially about color measurement, uniform color space and color difference formulations, a new way to design models empirically from user experiments started recently.

Ou and Luo adopted a pair-wise protocol where they asked to participants for rating the considered colors pairs from *extremely harmonious* to *extremely disharmonious*. 1431 colors pairs were investigated and generated using 54 colors

extrated from CIELAB color space. The color pairs were presented side-by-side against a medium gray background on a cathode ray tube in a dark room. The rating scale had 10 levels of appreciation. From this experiment, they found trends (by applying regression algorithms) from which they established a computational formulation for three color components: hue, lightness and chroma.

17 observers, 11 males and 6 females, participated in the experiment. Their methodology deals with the estimation of functions for the different features, that they isolated, such as difference and sum of lightness, hue and chroma difference. These three functions are computed from a regression extrapolated on the samples that they gathered from the experiment, such as illustrated in Figure 4.7. Finally, they linearly integrated the previous functions to obtain a final score of harmony between two colors. Complete details about the equations of the three functions may be found in [195].

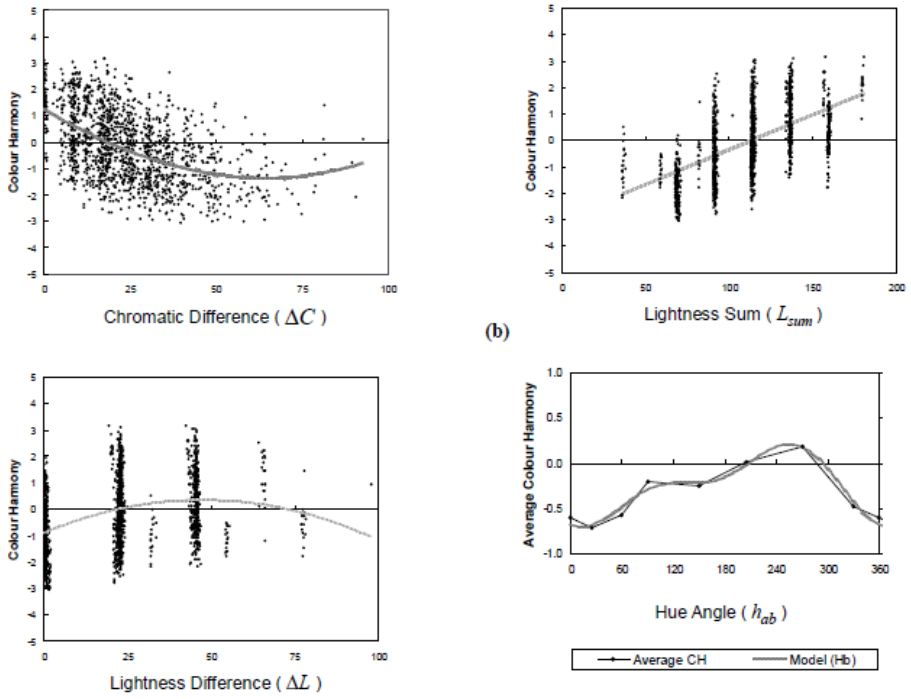


Figure 4.7: Ou's color harmony model: estimated functions to predict features characterizing harmony. (Extracted from [195])

They cross-validated the model by applying another dataset of color pairs, elaborated in different conditions (screen, color space, nationality...). They obtain a good correlation on both their own dataset ($R^2 = 0.73$) and the one of Gurura ($R^2 = 0.75$).

Szabo, 2009

In the same vein as Ou and Luo [195], Szabo *et al* [241] investigated the rating of patches of color combinations to derive a mathematical model. They proceeded with a similar methodology to design their model and to compare its performances against historical color harmony theory and Ou's model. As an additional contribution to Ou's model, they proposed a numerical model for three-color combinations.

They used the CIECAM color space where they derive 2346 two-colors combinations, 3222 monochromatic three-color combinations and 2x2000 randomly chosen trichromatic three-color combinations. These stimuli were presented to a limited number of observers (less than 10 for the different experiments). They use a 11 levels in their harmony scale for the annotation where they added the notion of *neutral color harmony impression* at the middle. Such as Ou and Luo, they assumed an additive contribution of the sums and differences of the perceptual color attributes. From the gathered scores, they observed tendencies that they translated into mathematical formulae (Complete equations are presented in [241]). More precisely, they ended up with several color harmony models depending on the considered association: *monochromatic two-color combinations, dichromatic two-color combinations, monochromatic three-color combinations and trichromatic three-color combinations.*

Szabo *et al* found similar tendencies compared to Ou and Luo [195] except for the hue preferences function that was not confirmed by correlation. They credited the different ethnic origin of observers as a major factor for such analysis. Satisfactory correlation around 0.70 were found regarding their dataset for the two- and three-color models.

They evaluated their numerical model against Ou's model [195] as well as three color harmony theories: Munsell ([177]), Coloroid ([180]) and RAL design [4]. Their model outperformed the Ou's prediction slightly, but substantially justify the introduction of numerical models. Indeed, it demonstrated much higher skill for prediction than any previous approaches.

Solli and Lenz, 2009

Solli and Lenz [239] designed a predictive model for estimating the perceived overall harmony in pictures by extending the two-color model originated by Ou and Luo to a multicolored model. They applied a color segmentation algorithm in order to set up the pair combinations that will feed the Ou's model. Since color segmentation is usually supervised, it seems tricky to determine a target suitable number of color clusters. To circumvent this observation, the authors defined a generic framework where they applied two passes of color segmentation leading to a coarse and a fine segmentation. The mean shift based image segmentation method [55] is applied twice with two different settings and the two sets R_{small}, R_{large} of regions r are described with a constraint on their area

a_r:

$$\begin{aligned} R_{large} &= \{r : a_r \geq 0.025q\} \\ R_{small} &= \{r : a_r \geq 0.0025q\} \end{aligned} \tag{4.7}$$

where q is the number of pixels in the image. Thus, they applied two strategies depending if they considered the large regions in R_{large} of size n or the small regions in R_{small} of size m.

They derived the harmony score (from Ou's model) for every unique color combination $r_{large} = 1, \ldots, \hat{n}$, with $\hat{n} = n(n-1)/2$, that they stored in the vector L of length $\hat{n}$. Then, the harmony score for the entire image is the minimum of such vector: $l = \min_{r_{large}} L$.

The case of small regions is more tricky since there are much more numerical values to be computed and potentially not relevant. They restricted the evaluated pairs to the five closest neighbors $N_{1\ldots5}$ of regions $r_{small} = 1, \ldots, m$. They stored the results in a vector $U(r_{small}, N_{1\ldots5})$. As done for the large regions, the smallest value in U provides the harmony score for the entire picture: $s = \min_{r_{small}, N} U$.

In addition, they also derived another useful information for the assessment of harmony. They assumed that a high harmony disagreement in neighboring regions means a low local harmony. Thus, they computed the variances over the neighboring regions, formalized through the vector U: $\sigma^2(r_{small}) = Var(U(r_{small}, N_{1\ldots5}))$. The variance of the entire image is the maximum over $\sigma^2(r_{small})$: $\sigma_s^2 = \max_{r_{small}} \sigma^2$.

They used a weighting combination of the previous local harmony scores for deriving the final harmony score:

$$h = \left[1, l, s, \sigma_s^2\right] \cdot W' \tag{4.8}$$

W' is determined by a linear regression from the scores annotated by users during an experiment. Finally, they ended up with a correlation score of 0.84 between their predictive model and their dataset.

4.2.3 Contingent model

Not only the limitations of previous evoked models, but also new trends in psychology, have gradually germinated the seed of a theory: color harmony is not only stimuli-dependent, but could also be related to other factors, such as environment, age, gender, origin, task...

In 1994, Sivik and Hard already mentioned the problem of having constant responses to color harmony depending on *time, fashion, culture* [234]. Anat Lechner originated the idea of *color harmony* ***contingencies*** [145]. More precisely, he argue for distinguishing the concepts of *harmony-within* and *harmony-between*. While *harmony-within* is purely related to stimuli properties, i.e. hue,

saturation and value components, *harmony-between* identifies critical elements that correlate to color harmony evaluation. We can easily see the parallel with human vision (Section 2.2), where exogenous (involuntarily) attention is stimuli-driven and endogenous attention is conceptually-driven [17]; referring also to bottom-up and top-down attentional mechanisms.

In her paper "'Color Harmony Revisited" [193], Zena O'Connor discussed several aspects highly relevant to this section. First, she highlighted the well-known problem stated around definitions and the corresponding approaches to model and to generalize the elicitation of color harmony. She argued that the universal and deterministic laws of color harmony can not hold up. Many previous studies made the evidence of aesthetic response (strongly related to color harmony) changing over time, space, humans. Thus, the response to color harmony is indivisible, according to her, of the *individual affective state, cultural differences* and also *contextual, perceptual and temporal factors.*

She also claimed that the contingent approach is a response to the lack of consensus for the color harmony definition. In such context, she formulated her color harmony conceptual model as:

$$\begin{aligned} Color\ Harmony &= f\left(Col_{1,2..n}\right) \times f\left(Cont_{1,2..5}\right) \\ &= f\left(Col_{1,2..n}\right) \times (ID + CE + CX + P + T) \end{aligned} \tag{4.9}$$

where basically, the color harmony response is the product of two functions, one related to the stimulus and its constituent colors Col_i, and one related to exogenous factors, such as individual differences (ID: age, gender, personality and affective state), cultural experience (CE), the context (CX: the environment), perceptual effects (P) and the effects of Time (T: social and design trends). No further investigations have been done to specify the different attributes; however a general framework is installed. Any numerical models mentioned in the previous section could feed the first term of the formula.

4.3 Derived Applications of Color Harmony

The geometric and numerical models of color harmony have been used in a context of image processing. The most noticeable breakthrough is in the context of automatic harmonization of pictures where a bench of papers investigated the automatic recolorization of content based on harmony criteria (Section 4.3.1). First, we will detail the the pioneer algorithm of Cohen-Or *et al.* [54], which is in relation to the contribution of Chapter 12.

The aforementioned models also played a role in metrics and models that involve high level and subjective notions, such as aesthetic metric, emotion modeling (Section 4.3.2).

4.3.1 Color harmonization

In 2006, Cohen-Or *et al* [54] introduced the first algorithm to automatically harmonize a color picture. They termed this specific image processing, *Color Harmonization*. This pioneer work has been later improved and extended to video content, but without any real breakthroughs. As additional contributions, Cohen-Or *et al.* suggested different color rendering usecases: compositing, foreground/background harmonization by introducing a manual mask, color design of illustrations, interior design, color transfer-like techniques...but the main point related to their paper relies on the **formulation of harmonic templates** originated from both Matsuda's color schemes [168] and the Tokumaru's friendly version [252].

The main idea of the automatic harmonization of colored pictures is to estimate the *closest* template to the hue distribution of the considered image and then, to remap the hues uncovered by the harmonious sector(s) of the candidate template inside this(ese) latter. This can be seen as a dedicated or constraint color mapping processing [76].

Harmony formulation

Figure 4.8a reintroduces a modern version of Matsuda's color schemes, with the specification on sector size, that are named *harmonic templates* by Cohen-Or. There are eight templates dedicated to hue components where grey sectors denote the sets of harmonic hues. They consist of one or two sectors which address different geometric arrangements of harmony: complementary for 180 degrees opposite sectors, analogous for similar hues in one sector and orthogonal for 90 degrees distant sectors. Such as formulated by Tokumaru [252], each grey sector relates to a simple membership function. Cohen-Or employed only the hue templates; he discarded tone harmonic schemes, certainly due to the complexity of their membership functions (Figure 4.4b).

The template formulation is initiated by expressing the link between any hue h to the closest sector (distance in arc length). Thus, they defined the set of templates T_m, such as $m \in \{i, I, L, T, V, X, Y, N\}$ and the sector border $E_{T_m(\alpha)}(p)$ of template T_m with orientation α the closest to the hue h of current pixel p. In their formulation, they found more convenient to manipulate the sector border rather than the considered sector as an unit. Note that we propose a slightly more comprehensive formulation of the templates in our implementation by indexing the considered sector (Section 12.3).

Given a membership function representing the different harmonic schemes T_m for all orientations α, the idea is to determine which template would be *the most aligned* on the considered picture distribution. In other words, it consists in finding the most appropriate template to the considered picture. The membership function or harmonic scheme is defined as the couple (m, α), where m is the template index and α its rotation angle on hue wheel. Given this, they defined a cost function F that measures the harmony of an image X,

$X : \Omega \subset \Re^2 \rightarrow \Re^3$. Note that this function rather catch the disharmony, since it is minimized afterward.

$$\forall p \in \Omega,\ F\left(X,(m,\alpha)\right) = \sum_{p\in\Omega} \left|h\left(p\right) - E_{T_m(\alpha)}(p)\right| \cdot s\left(p\right) \tag{4.10}$$

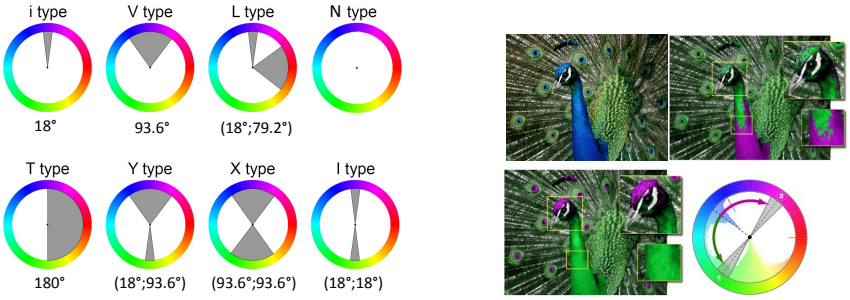
where h and s denote the hue and saturation channels, respectively. $|\cdot|$ refers to the hue distance on the hue wheel measured in arc length. This distance is equal to 0 when the considered hue is in the grey sector. This harmony distance may be apprehended as a simple average of all pixels weighted by the saturation.

Finding the best fit $B(X, T_{m_0})$ between the harmonic schemes (m, α) and the picture X consists in minimizing successively the two following expressions. In other words, finding first the optimal α_0 for each harmonic scheme T_m and after that the optimal m_0, knowing the optimal α_0:

$$\alpha_0 = arg\min_{\alpha} F\left(X,(m,\alpha)\right) \tag{4.11}$$

$$B\left(X,T_{m_0}\right) = (m_0,\alpha_0) \quad s.t. \quad m_0 = arg\min_{m} F\left(X,(m,\alpha_0)\right) \tag{4.12}$$

They used Brent's algorithm [35] to solve the minimization problem of orientation α. The Brent's method allows converging fast by choosing the most suitable methods over three available: the secant method, the bisection method and the inverse quadratic interpolation. The minimal distance F over m and α reflects the highest harmonic concordance between the picture X and the template representation.



(a) (b) from [54]

Figure 4.8: Cohen-Or 's Color Harmonization. (a) specification of Matsuda's templates. (b) Color shifting issue: without color segmentation, similar hues may be shited in two different sectors.

Algorithm

Once the best template (in sense of concordance) has been determined, the harmonization algorithm could be described by means of the aforementioned formulation (Equations (4.11) and (4.12)). It consists in shifting the unsatisfactory colors of X within the candidate harmonic scheme $T_{m_0}(\alpha_0)$. More

precisely, the idea is to remap inside the candidate template the disharmonious hue values. However, the approach of performing that only on the basis of the hue distribution is somehow naive: if no spatial coherency is taken into account, it may lead to different sectors (and colors) matching of original color, having e.g. gradual variation of hue as illustrated in Figure 4.8b.

To circumvent this problem, they employ a graph-cut optimization problem [33] for matching the hues not anymore to the closest sector, but to the one that ensures a better spatial consistency. To each pixel p is assigned a label $v(p)$ which reflects the binary decision between the two potential sectors. The optimal label assignment $V = \left\{v(p_1), \ldots, v(p_{|\Omega|})\right\}$ consists in minimizing the energy $E(V)$:

$$E\left(V\right) = \lambda \cdot E_1\left(V\right) + E_2\left(V\right) \tag{4.13}$$

where $E_1\left(V\right)$ stands for the distance between the hues $h(p)$ and $h(v(p))$, and $E_2\left(V\right)$ reflects color coherency in the neighboring pixels N (4- or 8- connected pixels in Ω) assigned to $v(p)$. Such as already done, the two expressions are weighted by the saturation to promote pixels with a high saturation:

$$E_1\left(V\right) = \sum_{i=1}^{|\Omega|} \left|h\left(p_i\right) - h\left(v\left(p_i\right)\right)\right| \cdot s\left(p_i\right) \tag{4.14}$$

$$E_2\left(V\right) = \sum_{\{p,q\} \in N} \frac{\delta\left(v\left(p\right), v\left(q\right)\right) \cdot s_{max}\left(p,q\right)}{\left|h\left(p\right) - h\left(q\right)\right|} \tag{4.15}$$

where $\delta(,)$ means $\delta(v(p), v(q)) = 1$ if $v(p) \neq v(q)$, otherwise $\delta(v(p), v(q)) = 0$. $s_{max}\left(p,q\right)$ is the maximal saturation between the points p and q. Once all pixels p have been binary labeled to an appropriate sector edge $E_{T_m(\alpha)}(p)$, they can be shifted or recolored with the appropriate hue $h'(p)$:

$$h'\left(p\right) = c\left(p\right) + \frac{w}{2} \cdot \left(1 - G_\sigma\left(\left|h\left(p\right) - c\left(p\right)\right|\right)\right) \tag{4.16}$$

where $c(p)$ is the central hue of the candidate sector, w is the arc width of this latter and G_σ is a normalized Gaussian function whose the standard deviation is σ.

Pioneer work of Cohen-Or formulates a comprehensive formulation and an applicative framework for Color Harmonization. Most of succeeding work built on it for improving the final color rendering.

Improving the Cohen-Or's work

There have been some work about harmonization in a wider sense, especially authors targeted the harmonization of more features than just color: tone [261, 242], grain and contrast [240]... We will not provide details about these implementations, but rather focus on the one related to color feature.

Pursuing the work of Cohen-Or, other authors proposed some improvements of the original formulation. All the following attempts relied on the same harmonic templates to recolor or recompose color within a picture [228, 262, 108, 243, 90, 245, 260]. Following the Cohen-Or methodology, their first step consists in determining the harmonious template type and its rotation angle that is the closest to the original picture by minimizing a cost function. Then, the colors are transformed so that the colors outside the harmonious sectors are mapped inside a harmonious sector. Color segmentation has been identified as a crucial pre-processing before color mapping because visible artifacts can appear when two close colors, eventually associated to the same object, are mapped to two different sectors. Strategies between the papers differ in terms of cost function for template selection, color segmentation and color mapping.

Huo *et al.* [108] improved the original algorithm by changing the choice of the best harmonic scheme and applying another strategy for the color shifting as they claimed. Differently to Cohen-Or, they define the optimal angle α_0 as the one corresponding to the peak in hue histogram (weighted by saturation). Thus, they do the same as Cohen-Or for template determination and color segmentation. They claimed also to simplify the Gaussian function used for color shifting.

Tang *et al.* [243] proposed more additional improvements in their implementation. A cost function is derived from the observation of the local smoothness of the hue values. Also, a new matching function (Equation 4.17) is introduced for the choice of the best harmonic schemes. Finally, they adapted the original architecture by adding a new component based on a pre-harmonization strategy to preserve the hue distribution of the harmonized images.

Following the same formulation as Cohen-Or, they defined a relative rather than an absolute distance for the matching function than an absolute distance:

$$F\left(X, T_{m,\alpha}\right) = \frac{\left(\sum_i h(i) \cdot T_{m,\alpha}(i)\right)^2}{h_{max} \cdot \sum_i T_{m,\alpha}(i)}, \quad 0 \leq i \leq M-1 \tag{4.17}$$

where h is the hue histogram of image X. M and h_{max} are the number of bins and the peak value of the hue distribution, respectively. $T_{m,\alpha}(i)$ is the i^{th} bin of the harmonic scheme T or may be seen as the membership function such as described by [252]. It is defined by a simple binary statement:

$$T_{m,\alpha}(i) = \begin{cases} 1, & \text{if } i \text{ is in the gray sector(s)} \\ 0, & \text{otherwise} \end{cases} \tag{4.18}$$

By proposing the ratio in Equation 4.17 as a matching function, the author favored the harmonic scheme where most of the pixels fall within its gray sector(s). The bigger the ratio, the more similar the hue distribution and harmonic scheme are.

Different from previous work of Cohen-Or, Tang *et al.* treat color harmonization as a hue value reassigning process rather than a color shifting process.

In such way, they solve the problem of spatial coherency and allow not changing the hue of pixels inside the gray sectors, which may be seen as a limitation of previous gaussian-based shifting functions (see [24] for details). They posed the problem of color reassignment as it is done also for colorization algorithm [148], i.e. as an objective function to be minimized by means of an affinity function ω_{rs} between pixels r and s (Equation 4.19)). Note that the reassignement step only concerns the set of pixels, denoted $\bar{\Omega}$, that have been previously identified as being outside the candidate template.

$$J(h) = \sum_{r \in \bar{\Omega}} \left(h(r) - \sum_{s \in N(r)} \omega_{rs} \cdot h(s) \right)^2, \tag{4.19}$$

where $N(r)$ is the neighboring pixels around r. The affinity function is defined by:

$$\omega_{rs} = \frac{1}{\sum_{s \in N(r)} e^{\frac{-(h(r)-h(s))^2}{2\sigma_r^2}}} \cdot e^{\frac{-(h(r)-h(s))^2}{2\sigma_r^2}}, \tag{4.20}$$

where σ_r is the standard deviation of the hue values in a window around pixel r.

A real-time color harmonization for video has been introduced by Sawant *et al.* [229] where a histogram splitting method is employed instead of a graph-cut approach to reduce computational cost. One dedicated template per group of frames is determined to guarantee temporal consistency. In [245], Tang *et al.* perform a foreground/background detection in addition and apply the same template determination as [229] for a coherent group of frames.

Finally, we can notice that Cohen-Or algorithm has been used in constraint environment, such as Augmented Reality [262, 90], where simplification was performed to be compliant with rendering issues.

Almost all works rely on the Cohen-Or approach for performing Color Harmonization. We can notice the work of Sauvaget and Boyer [227] formulating the color harmony principles from Itten's contrasts (Section 4.2.1). Similarly to Tang *et al.* [243], they estimate the best harmonic scheme (but based on Itten's ratio) and reassign the oversized or undersized hue regions. Their shifting strategy is somehow sophisticated, since they let the user to choose between four methods: basic hue attribution (hue limit of candidate sector), closest hue, preserving ratio distance and maintaining density distribution.

4.3.2 Color Harmony as an image feature

This section describes a bench of applications where computational color harmony has been used as an intrinsic description characterizing the image. Color harmony as a feature is estimated and input to infer aesthetic metrics, emotion estimation, image indexing...

Aesthetic Quality Assessment

Aesthetic metrics have recently attracted a lot of attention and many researchers have published their own model. More generally, high-level semantic concepts respond to the need of having models or metrics closer to human perception [120]. Color harmony has been also explored in such concept. Coming from the computer vision community, the aesthetic models rely on machine learning approaches which build a model or a prediction based on extracted features and annotated ground truth.

Luo *et al.* [158] and Li *et al.* [149] were precursors in aesthetic or appeal metric. While Li's model exploited fully the Matsuda's templates and provided evidence of its efficiency as a global feature in the context of machine learning, Luo could not extract from Matsuda's templates a numerical value or index that could help the prediction of aesthetic as a low level feature. Thus, he designed his own harmony feature by learning the best color combinations from the training dataset. Li's metric and dataset being devoted to painting material may explain the fact that the color harmony feature is more valuable. Later, Moorthy *et al.* [174] also investigated the Matsuda's templates as a feature of aesthetic, they were rather in accordance with Luo about the difficulties for exploiting color harmony roles.

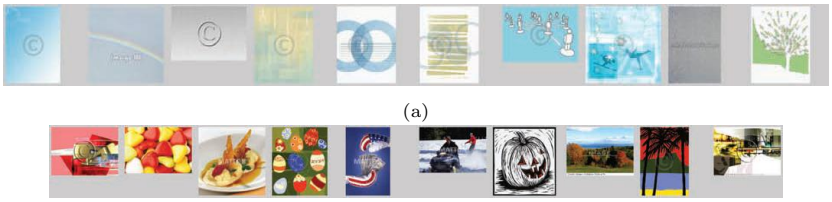
Last but not least about aesthetic quality metric, Nishiyama *et al.* [186] completely specialized their metric based on color harmony concept. They experienced the direct exploitation of geometric color harmony models (Moon and Spencer approach as well as Matsuda's templates) to predict the aesthetic quality and faced, according to them, a poor prediction due to the fact they are designed for simple color patterns, not for natural and complex pictures. Thus, without rejecting such models, they decided anyway to apply color harmony theory at a local scale, on small portion or areas of the pictures. The method extracts a sample of local regions, describes each of them by color harmony descriptors, quantizes these features and represents the picture as an histogram of quantized features in order to later apply a Support Vector Machine classifier. Their color harmony features derive from Moon and Spencer's model.

Emotion inference

Recent works about the extraction or prediction of emotions perceived by users or induced by stimuli get more and more attention. Basically, all these models rely on machine learning approaches [162, 18, 153] and need to compute features. Some of them take advantage of color harmony's theories to compute one or several features that could, at some extend, catch the emotion induced by the considered stimuli. This is quite similar to aesthetic metric, where a high score in color harmony potentially produce a positive assessment and emotion of the stimuli. In [162, 153], the authors prefer implementing Itten's contrast as their harmony feature, while in [18], we contribute to the emotion prediction by means of the Matsuda's color coordination system.

Image indexing

In Section 4.2.2, we described Solli and Lenz predictive model. They elaborated an interesting extension of the Ou and Luo's model, where they attempted to predict the color harmony of multi-color pictures. Originating from the image processing community, they targeted image indexing or classification as direct application of such model. They used an user experiment to tune the weights necessary to the combinations of several aspects of harmony (local, global). They achieved 0.84 and 0.49 of linear correlation between their predictive harmony score and the same user experiment, depending if they considered a *mean* observer or all observers data. Unfortunately, they did not perform additional experiment to cross-check the strength of their model prediction on another annotated dataset. They provided qualitative results by means of images extracted from their dataset on harmony score request (Figure 4.9). This qualitative appreciation lets the reader uncertain about the model performances, particularly potential bias about saturation.



(a)

(b)

Figure 4.9: Visual appreciation for the Solli's color harmony model [239]. (A) Most harmonious images, (B) Most disharmonious images. (Extracted from [239])

4.4 Summary

This chapter presented a status of the literature for the topic of Color Harmony. This topic occupied many thoughts since early time: already Newton during the 18^{th} century get interested in the color arrangement and their potential meaning. Interestingly, he evidenced the *circular* representation of colors and their complementarity. However, at this time many fields of thought claimed to be more competent and authorized for investigating color harmony. Thus, poets, philosophers and painters started implementing their vision, their color wheel and their theory of color harmony. All these deployed energies led to diverse interpretations, definitions, representations of color harmony as discussed in Section 4.1.

Later, during the 20^{th} century, some computational approaches and exploitable models got introduced by color scientists, mathematicians, psychologists. We classified them into three categories (Section 4.2). The *geometric* models build on the historical theories of harmony by defining rules on color wheel. The *nu-*

merical models are empirical and expressed the color harmony from regression applied on experimental data. Finally, the *contingent* models is a psychological approach that claims the involvement of stimuli but also the inference of other cues in the assessment of color harmony, such as the mood, the temporal factors, the user's characteristics and so on.

Having these models available, some researchers from the image processing field (Section 4.3) attempted to process automatically the colors of pictures to get them homogeneous, this is the *Color Harmonization* processing introduced by Cohen-Or *et al.*[54]. They proposed a harmony formulation and different frameworks to apply such kind of algorithms. Succeeding work improved the original implementation and proposed also to use the harmony formulation in other contexts, e.g. as a cue for quantifying the aesthetic and emotion of pictures.

Chapter 5

Discussion

This chapter discusses the relationships between the three previous chapters that depict the state-of-the-art of this thesis. If the Visual Attention and Eye-Tracking evolved in a strong partnership, Color Harmony is fairly a topic reserved to color scientists. At a first glance, there is no relationship between them.

In Section 5.1, we propose to decline the different color mechanisms and the related concepts. From this big picture, we attempt to position the concept of Color Harmony in such context (Section 5.2). In a third section, we discuss the mentioned color harmony models, their status, validity, applicable domain (Section 5.3). Finally, there is a big limitation in the study of color harmony: the ground truth aspect (Section 5.4).

5.1 Color Mechanisms

This section aims at exposing that colors act at different levels of processing in the brain. There are many concepts related to colors that are parts of human past, history, common knowledge and deals with top-down processing. In addition, there are low-level mechanisms that tend to be common between individuals, even if biological evolution creates specificity related to the environment and learning.

Color in the brain: low level processing

The involved processes in color vision become well-identified. The eye anatomy has been studied, dissected and clarified through many experiments. We know how the color are perceived when the light reaches the pupil and how this is translated into a neuronal signal to be processed later in different areas of the brain (Section 2.2.2) . This last point is more controversial. While the area V1 is clearly the center for vision processing, other areas may play also a role in the analysis of color features.

Even if V1 neurons have been identified as being the heart of color contrast [259], it seems that V2 and V4 areas participate also to the gradual analysis of color features. More specifically, V4 is the place where *Color Constancy* mechanisms are realized. Color constancy is the ability of our brain to identify a constant color associated to an object despite the viewing conditions (e.g. illumination, light reflected from the surface and object). This notion is fundamental for any living entity as a defending mechanism. Finally, the discovery of another area, named V8 or V4α, suggested that it is the site of neuronal processing for color constancy and the conscious perception of color [91]. Nowadays, the role of each area in color vision is not distinct, it can not be split into different tasks and concepts [102].

Additionally to theoretical and measured phenomena explained previously, color perception is also variable between individuals [224]: this is the inter-individual variability of color vision or observer metamerism. Proper physiological arrangement of cones and adaptation of human to surrounding environment lead to differences in cone functioning and performances that can be measured and detected. Typically, color blindness (due to the absence or deficiency of one or two category of cone(s)) is a critical pathology when measuring color factors. Less drastically interfering, the number and repartition of the three types of cones vary between individual [224] providing a variable perception of colors. Despite the progress in color science for measuring objectively the perception in color difference [5], the related experiments may be disturbed by these aforementioned factors.

Color-related notions: high level processing

Beyond the *Color Perception* referring to immediate processing in the brain, there are also higher levels of processing for the color feature. We will discuss here some concepts that are top-down mechanisms in relation with colors. We introduced *Color Cognition*, opposed to Color Perception [28].

For most of culture, colors are meaningful; they symbolize a mood, natural elements and infer a preference related to a memory. *Cognitive color* [67] goes further than color naming. It associates a semantic concept or object to a color. In other words, the human knowledge (due to long-term memory) of surrounding objects or our familiarity to them allows forming the following pairs: e.g. banana-yellow, blue-sky etc. More generally, there have been substantial research exploring both color and memory. Main hypothesis rely on the impact of color in a memory task. Several studies agreed on a faster recognition and higher memorability of stimuli having color properties (versus black and white stimuli) [77, 89, 268, 201]. Color seems to facilitate the memorization process by eliciting more attentional mechanisms. For a complete recent review of this topic, one can refer to [72].

In the same vein, color mood or arousal has been explored. Many studies investigated the link between colors and elicited emotions [128, 197], e.g. it is agreed that black is associated to sadness, fear; blue to peace; yellow to happy and so on. More specifically, Kobayashi [130] defined a taxonomy between color

triplets and words related to high-level semantic concepts. To make the link with the previous topic, the memory effect is also variable according to the type of mood and thus to the considered color. For example, it has been found that emotional elicitation created from red is very high [115].

Color preferences is not a recent topic of investigations [75], even if findings in this area are recent. For such topic, several studies reported a variability in preferences due to age [218], gender [196] and culture [163]. Also, related to preferences is the notion of aesthetic. Color is known for contributing to quality and beauty [230, 66]. Thus, high colorfulness, right color balance or proportion may maximize the user experience. With this last high level notion related to color, we are approaching the Color Harmony definition. The distinction between *preferences, harmony* and *similarity* is thin [230], this requires a specific care on experimental protocols in order to measure the right notion, if any distinction can be done.

5.2 Color Harmony at which stage?

All the previous topics engage color processing in the brain. Thus, specific cells analyzing the color signal are involved at different levels of the brain depending on the color-related tasks. There are some high-level concepts engaging the color factor and its understanding.

Intuitively, we believe the color harmony concept being related to top-down mechanisms. Due to the no-consensus on its definition, it seems reasonable to think that the mechanisms related to this concept are inferred at a high level, such as it is the case for color preferences, color mood and so on.

But is Color Harmony too subjective, person-related, culture-related to be measured? We do not think so, because we believe the concept easier to apprehend than e.g. color preferences (having high inter-observer variability), or color mood. As mentioned previously, measuring the color preferences lead to high inter-observer variability, if not controlled. This concept is highly related to culture, age, personal feeling. Color mood is a difficult field that requires the mapping between two heterogeneous data: an emotion symbolized by a word and a color (1, 2 or 3 colors can be involved). Also, we need to all agree on emotion definitions and representations which is not ensured. However, many experiments explored these two fields. More specifically, it seems that the study of emotional mechanisms in the context of eye-tracking exploded recently (Section 10.1).

We believe Color Harmony is easier to investigate than Color Preferences, because 1/ there are empirical work and geometrical representation to rely on, 2/ we think people will more agree on such concept due to the presence of more universal rules. Regarding the topic of Color Mood, we expect Color Harmony meeting the same levels of limitations (Figure 5.1) for the inter-observer agreement.

However, this point should be monitored in any experiments to ensure that

participants understand the concept of Color Harmony and do not formulate it differently. Schloss and Palmer [230] having pointed out the protocol issues when assessing the color harmony of pair combinations elaborate three different ways of evaluating perceptual harmony: people's aesthetic preferences, perception of harmony and preference for a figural color when viewed against a colored background. Thus, they distinguish the notions of preference, harmony and similarity of color combinations (as being three different tasks), which have been so far fully included in the concept of color harmony. They also asked for a rate on a scale having different meaning depending on the three tasks.

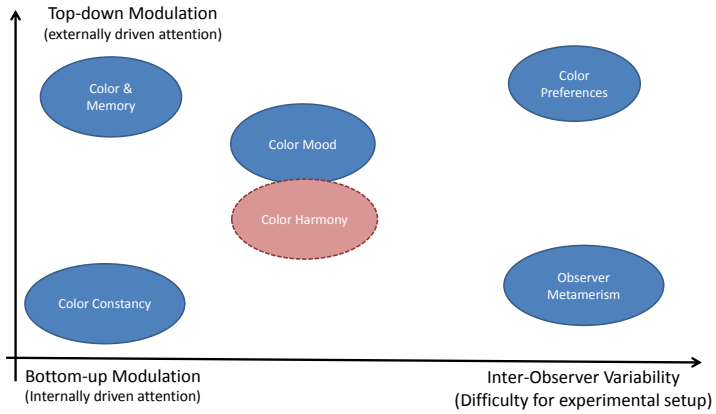


Figure 5.1: Color Harmony mechanisms: a tentative for positioning it in a 2D representation top-down/bottom-up modulation as a function of inter-observer variability.

In Figure 5.1, we tentatively depict a graphical representation of top-down/bottom-up modulation versus the inter-observer variability. Based on previous inputs, we venture the hypothesis of a "middle" positioning for the color harmony topic.

5.3 Model status

Color harmony theories over centuries inferred intuitive notions of color harmony that have been implemented and verified a posteriori in design and painting but mostly without rigorous scientific validation. Their implication in the contemporary status of color harmony field is huge, because they are the basis of existing computational models.

Validity of models

In the case of numerical models, the authors carefully controlled the input stimuli on which they extrapolated their mathematical functions. However, we can point out several limitations.

First, the measure of agreement for inter- and intra-observers is open to criticism. Ou and Szabo employed the Root Mean Square (RMS) measure

which limits the accuracy of agreement and is difficult to interpret anyway. Recent methods discussed the validity of how agreement should be measured with regard to the kind of involved samples, experiments [225]. The tendency is to include a chance factor to penalize the final agreement. We can cite the following coefficients that tend to generalize in many experimental approaches: Cohen's Kappa [53], Fleiss' Kappa [79], Krippendorff's alpha [133], Randolph's multirater Kappa [216]. However, such measures may require a minimal number of observers to exploit their power.

This leads us to another underlying limitation: the number of observers involved in the experiments of Ou and Szabo. They do not exceed 10 observations, but they asked for multiple rating of the same observers to check also the intra-observer consistency. Another inherent limitation of such small sampling is their origin, age and gender which may also restrict any form of generalization.

Second, even if authors carefully selected their color samples, a pair-wise or rating protocol intrinsically limits the number of pairs that could be annotated. Thus, the generalization of the model to other dataset is also a remaining question. Szabo *et al.* [241] depicted this problem by highlighting a small correlation ($R^2 = 0.30$) of Ou's model on their dataset. However, he mentioned the much higher correlation of the two numerical models compared to historical harmony theories. Even if they are questionable, the two numerical models of Ou and Szabo clearly have the merit of being designed experimentally and proposed to the community.

The geometric models are less disruptive from the tradition of color harmony history. Relying on color wheel, they precise the areas inferring color harmony for each specific components. However, they are open to interpretation since they basically either offer several harmonious areas at the same time on wheel [172] or several possible arrangements (templates) of areas on wheel [168]. While Moon and Spencer did not really empirically validated their approaches, Matsuda built its template approach entirely on samples that he gathered during several years on japanese students. A clear limitation is the bias in age, gender and also regarding the used stimuli, since this work was dedicated to fashion application. Despite that, some authors proposed a framework for application in color design and image editing [252, 54].

Recently, due to the possibility of large datasets designed by the web community [1, 2], we can notice some interesting attempts [194, 235] for confirming the relevancy or updating the design of Matsuda's templates. O'Donovan *et al.* clearly reject the interest of Matsuda's templates. While they evidence no correlation between Matsuda's templates and the color themes that designers agreed for real use, Skurowski *et al.* confirmed the intrinsic use of Matsuda's templates by designers and additionally proposed some new relevant templates. Despite a high concordance in the material they used, both investigations did not reach the same conclusion. We argue this is potentially due to the distance metric they used to model the likeliness between the Matsuda's templates and the 5-color themes dataset.

Applicable or not?

The numerical models take as input some features related to hue, value, chroma and provide a prediction for two- or three-color combinations. Unfortunately, there is no work that either extended this modeling to a higher number of color comparisons or proposed a generalization to n number of colors. It clearly restricts the applicative field of such model. Nonetheless, we can cite the work of Solli and Lenz [239] who derived a predictive model for estimating the perceived harmony of ordinary multi-colored images from Ou's model in the context of image indexing or classification.

On the other hand, the geometric models have been largely employed in image processing application and satisfactory results were demonstrated for natural pictures (Section 4.3). Authors' interpretation have made them applicable and useful to concrete application having complex stimuli. Nevertheless, semi-supervized or assisted human intervention seems more relevant to overcome the room they let for interpretation.

5.4 Ground truth issue

Thinking about the lack of robust validation of the models, this leads us to another limitation: the need for ground truth on complex/natural stimuli. Numerical models are based on simple patterns and it is difficult to guarantee they could be extended to complex or natural stimuli despite the attempt of Solli and Lenz [239]. Geometric models have been translated in image processing community, but the resulting algorithms suffer also from being quantitatively tested and validated.

A simple reason is behind this: how can we reasonably design a perfectly harmonious picture for any kind of content that would serve as benchmark? Can we ensure a high agreement without bias related to age, gender and so on. On one hand, annotations at large scale with crowdsourcing could be a solution. On the other hand, the context of annotation is not controlled for such protocol. Anyway, the qualitative and manual appreciation of a human remains the only way to assess the quality of such processing. Once again, we are facing the lack of scientific foundation at the origin of the color harmony theory.

In the field of visual attention, it is a regular practice to record eye fixations in order to understand the attentional brain mechanisms. By varying the protocol conditions, the stimuli, the observer's tasks, particular behavior can be observed and analyzed. Such mechanisms are not related to only low level actions, but may also be related to visual search, object recognition, memory task etc. Thus, we will intuitively turn to this technology and type of protocol to measure color harmony-related behavior (Chapter 8) and create an associated ground truth (Chapter 9).

Part II

Experimental approach: attention, color and harmony

Chapter 6

Introduction

The second part of this thesis consists of three chapters and aims at dealing with the concept of color harmony through an experimental and prospective study. The two first chapters are purely experimental by designing two distinct experiments with an eye-tracking: the first one related to the color factor (Chapter 7) and the second one to the color harmony factor (Chapter 8). Building on them, the third chapter (Chapter 9) aims at creating a ground truth by filtering and post-processing the data gathered in the experiment of Chapter 8.

In this chapter, we explain our approach by depicting a short state-of-the-art of the task protocols for the eye-tracking experiments (Section 6.1). In a second section, we describe briefly the recent debate about the implication of color in visual attention mechanisms (Section 6.2). In a third section, we tackle the color factor in the context of eye-tracking (Section 6.3), e.g. the involved protocols, tasks, stimuli. Finally, we explicitly provide the different hypothesis that we are going to test along this part (Section 6.4).

6.1 The experimental perspective

We tackle the experimental problem of color harmony as being related to visual attention and involving top-down mechanisms (Chapter 5.2). Since visual attention is mainly measured by eye-trackers, we propose an attempted paradigm to characterize color harmony by means of eye movements information measured with a task protocol (Chapter 8). The related works in color harmony are not so important and the experimental field remains widely unexplored. Main past contributions concern controlled and simple stimuli in experiments, e.g. patches of two or three colors arranged either side-by-side or contrasting in backward/forward presentation [195, 241, 230]. Considering such stimuli, the application and generalization of current funding turned out to be limited for natural and complex scenes. Consequently, substantial additional ways could be investigated in characterizing the concept of color harmony: having more complex stimuli for characterizing color harmony (how many colors, which pro-

portion? etc.), looking at the spatial combination and influence, defining a color harmony masking and so on.

To measure purposely color harmony effects, we thought a task protocol was necessary. In other words, would it be enough to ask participants for pointing out the harmonious color? This is a bet; if they can point it out consistently, we assumed they understood the concept and agreed on its assessment. On the other hand, if they disagreed, it would be reasonable to conclude that the concept is not universal (over our participants at least) and that such concept infers very high level attentional mechanisms related to personal taste, experience and so on. Thus, a non-exhaustive review of task protocols in eye-tracking experiments is provided in the following paragraphs.

Pioneer work of Yarbus [275] is well-known from the community: he showed the evidence of task-related scanpaths for identical stimuli. Even if the experimental condition as well as the conclusion have been recently revisited [64, 248, 88], he introduced the notion of task dependency later investigated under different conditions. DeAngelus and Pelz [64] found clearly that eye movements were task-dependant, but less drastically than Yarbus exposed it. In the same vein, Tatler *et al.* [248] drew the same conclusion as Yarbus experiment related to faces. One step further, Greene *et al.* [88] tentatively aim at predicting task conditions of different scanpaths. They highlighted evidences of differences between scanpaths but could not predict the task by using only eye movement patterns. Recently, Borji and Itti [31] as well as Kanan *et al.* [125, 124] reinterpreted the data provided by Greene *et al.* [88] and found that it is possible to decode and infer from machine learning techniques the task's observer from eye movement, significantly above the chance level. Borji and Itti [31] also evidenced the same conclusion on Yarbus' stimuli and some others, but more moderately. Yarbus' work motivated numbers of investigations on task role for eye gaze data. Tatler *et al.* [246] also investigated the task influence on spatial statistics and found unsurprisingly clear evidences of differences between free viewing and searching tasks on saccade amplitudes. They temperate their results based on potential influence of task nature and highlighted the need for complementary investigations of task bias especially in any saliency-based models.

Other studies proposed to extend Yarbus' conclusions to more complex tasks, environment and stimuli. Elaborated tasks, such as walking, driving [138] or washing hands [204] with non-static eye-tracking systems or non static scenes have concluded that a task-dependant strategy is elicited by the visual system and is not related to salience or conspicuity of objects in the scene. Same conclusion appeared on gaze data for the task of copying a colored pattern of building blocks [16]. Decision-making task reveals to be a favor ground for the investigation of the influence of eye movement control during complex cognitive tasks. A complete review of task diversity is proposed by Ballard and Hayhoe in [16] where they tentatively proposed a task-dependent approach modeling of gaze control and argue for abandon of image-based computational models.

Based on these studies, the task strategy of the visual system does exist but

its characterization or prediction depends on the complexity and the own nature of the task.

However, before investigating such experiment, we would like to clarify the color role in the guidance of visual attention. In Chapter 7, a first experiment involving the color factor in an eye-tracking experiment is detailed and serves as a support to the second experiment proposed in Chapter 8.

6.2 Color and Visual Attention

Chromatic feature is a predominant factor in the human visual attention system. Even if it does not represent all the information required for the understanding of a scene, it may intuitively be considered as a major factor that impacts the visual attention deployment.

Visual pop-out effect [254] occurs when a singular object is swamped in a homogeneous set of distractors. Transposed to the color case, it is obvious that a red line would attract immediately the attention when presented among several green lines. This pop-out phenomenon happens early during the pre-attentive stage. However, some studies [251, 238] got interested in whether such effect really occurs due to the color contrast or if a luminance difference between the two present colors is hidden and responsible for such attraction. In addition, the literature highlighted a paradox: while the *where* stream is recognized for carrying out attentional processing [210], these dorsal regions do not deal with color information (Section 2.2.4). Hence, the color feature should not be involved on its own in attentional guidance. Theeuwes [251] made an experience to state on the parvocellular (P or *what* stream) involvement in attentional guidance. They did not measure any pop-out effect between two isoluminant colors, confirming that the *what* stream (processing the color) is too slow to elicit pre-attentive mechanism. On the contrary, Snowden [238] evidenced with a similar paradigm that a chromatic cue attracts attention, arguing two possible reasons: either this is due to the fact that the M or *where* stream is not completely color-blind or the P (*what* stream) stream plays a role in attentional guidance as not assumed before.

Tatler *et al.* [249] studied the implication of luminance, contrast, chromaticity and edge-content in the visual attention deployment. Their precursor work demonstrated carefully the contribution of such features over time. They evidenced that discriminability between fixated and non-fixated areas is more important for the contrast and edge-content features than for the luminance and chromaticity. Also, they pointed out that such measured rates are surprisingly low (63% for contrast and edge-content features and 57% for the luminance and chromaticity features), showing that attention is far from being entirely driven by image statistics. This may lead to reconsider the key role of such features in the biological architecture of models (Section 2.3.2).

6.3 Color and Eye-tracking

In this section, a focus is set on existing experiments involving the color factor or color attributes in a context of eye-tracking. The purpose of such eye-tracking experimental methods may differ regarding their study context. However, setting-up an eye-tracking experiment is likely to be related to visual attention deployment.

Baik *et al.* [15] investigated the visual attention with low level color-related stimuli: mainly controlled patches in hue, saturation and brightness, having potentially a small object with other controlled color attributes. Observers had to watch randomly localized color patches. Through three eye-tracking experiments, they end up with four interesting conclusions: 1) subjects are more attracted by warm colors than by cool colors, 2) tone does not influence significantly the subjects' initial attention, 3) complementary color contrasts greatly drive the attention and 4) background and foreground contrast is not relevant in visual attention deployment.

Hu-Phuoc *et al.* [104] aims at understanding with experimental data the impact of color as a visual feature on natural scenes. Following the Feature Integration Theory (FIT) of Treisman and Gelade [254], they purposely want to estimate the relevancy of integrating such cue in a visual attention model [103]. To do so, they designed a free viewing eye-tracking experiment on three stimuli sets: color scenes, their counterparts in greyscale and abnormal colors. As first suggested by Tatler *et al.* [249], their analysis revealed that color information contribute little to eye fixations, and by contrast, high spatial frequency luminance plays a far more important role. This conclusion is also confirmed by Liu and Heynderickx [152] who give up the color channels in their computational model for decreasing complexity in a constraint context.

Earlier, Frey *et al.* [81] focused on the role of color features (saturation, Red-Green color contrast, Blue-Yellow color contrast) on visual attention for different image categories (Face, Rainforest, Man-made...). They also recorded eye fixations of the selected images for each category and their greyscale version. Since most of categories (four over seven) produced a significant difference between the fixations locations of the two conditions, they claimed a strong influence of color on human overt attention. Unfortunately, they did not provide any correlation numbers for the complete set of pictures. Later in [82], the same authors focused on the Rainforest category for which they found the highest color-contrasts at fixated locations. Going one step further, they manipulated the color properties of stimuli before presenting them to color-normal observers and deuteranope observers. This color deficiency as well as the color channels manipulation of stimuli allowed to evidence that Red-Green contrast has a significant influence in overt attention, much higher than Blue-Yellow contrast.

Recently, Amano *et al.* [10, 9] experimented a search and detection task, measured by eye-tracking, where the color properties were responsible for 57-60% of the variance in detection performance. This is reported as being as good as gaze position in free viewing. However, the authors pointed out also a

singular behavior of the performances related to the scene category [9].

Jost *et al.* [121] also investigated the contribution of color in the prediction of a visual attention model. In the same vein, Hamel *et al.* [92, 93] set up recently an eye-tracking experiment with colored and grey pictures. They measured a higher variability between observers for eye movement positions of the color stimuli than for those of grey stimuli [92]. One step further, they measured a slight influence of the color factor in eye movements, i.e. the number of fixated regions was slightly higher for the color stimuli than for the grayscale stimuli [93]. They demonstrated that having a luminance-based model is good enough to predict the saliency of color and grayscale stimuli.

Related to color but at a higher cognitive level, Lee *et al.* [146] studied the notion of color preference. They brought a new methodology for the understanding of this concept by the introduction of an eye-tracking experiment. They demonstrated a clear correlation between observers' color preferences and their return of fixation, fixation durations and fixation counts. Decorrelated from the color preference factor, they observed that more attention was paid to textured colors than non-textured colors.

Surprisingly, the work involving the color factor in an eye-tracking experiment is not so wide. Most of them are related to visual attention and the associated computational model. The conclusions are not shared enough to ensure a well-established basis on color role in viewing behavior. We decided to conduct our own experiment in eye-tracking and to estimate first the impact of color factor in visual attention deployment. We added another factor which was investigated first by Nummenmaa *et al.* [136, 191]: the emotional category of pictures. In addition to color factor, we would like to measure any potential effects of emotion on eye movements. We differentiate from Nummenmaa *et al.* by using other emotion categories, such as developed in Section 7.1. Based on the findings of this first experiment, which appeared to be later consistent with [104] about the color factor, we could set-up a second experiment with a color harmony task.

6.4 Hypothesis and approach

As far as we know, there is no tentative work for recording eye fixations targeting the color harmony factor. There are two goals we pursue: 1/ we want to better understand the visual attention mechanisms related to color harmony in order to model it; 2/ we want to create a ground truth in order to validate our model (Chapter 13) and to exploit this latter within an image processing editing tool (Chapter 14). To do so, we measured in a first experiment the surrounding bias related to color and emotional factors. In a second experiment, we instructed the observers with a color harmony task and recorded the eye fixations. Finally, in Chapter 9, we exploited the findings and data of the second experiment (Chapter 8) to derive a ground truth.

Experiment 1

Hypothesis 1: The color factor influences the visual attention deployment.

Hypothesis 2: The emotion factor plays a role in visual attention deployment.

Experiment 2

Hypothesis 3: During a task protocol with eye-tracking for measuring color harmony, the inter-observers consistency is high.

Hypothesis 4: There is an influence of color distribution and color diversity in the assessment of color harmony.

Ground truth

Hypothesis 5: The eye movements and individuals' assessment of color harmony are relevant enough to create a ground truth.

Global hypothesis: the concept of color harmony is well understood by observers, homogeneous between anyone and close to universality.

Chapter 7

Experiment 1: Color Factor

Contribution: *C. Chamaret, Color impact in visual attention deployment considering emotional images, Proc. SPIE. 8291, Human Vision and Electronic Imaging XVII 82911T (February 9, 2012)*

The proposed experiment is a preamble to the central problem of achieving the characterization of color harmony. This experiment gets the reader more familiar with eye-tracking experiment and lays the foundations for designing any protocol involving both eye-tracking and color harmony (Chapter 8).

7.1 Introduction

Biological mechanisms related to colors are well-known, e.g. in the eye, it is clearly the role of cones to translate the inputted wavelength signal into a neuronal signal that could be interpreted by brain areas (Section 2.2.3). However, it becomes less clear when studying the role of these different areas for decoding the color information (Section 2.2.4 and 6.2). Related to this statement, few experiments (in proportion to other features, such as orientation, spatial frequency...) really investigated the role of color in visual attention. Only recently, its contribution in eliciting attentional mechanisms has been debated [249, 129, 81, 103].

Some investigations focus on differentiating a specific behavior per image category. Thus, Parkhurst *et al.* [203] established that color plays a dominant role (compared to intensity and orientation) in the fractals and home interior category, while this is the luminance feature for the landscape and man-made categories. In the same vein, Frey *et al.* [81] measured the influence of several color features on visual attention for different image categories. They evidence a different contribution and role of color depending on the image category. Focusing on pictures of tourist attractions, Kim [129] evidenced an effect of color in the task of working memory, but not in visual attention.

Other investigations tackle the problem of measuring color influence in vi-

sual attetnion by focusing on a specific task. In [10, 9], Amano *et al.* aimed at determining in what extend the color properties of natural scenes play a role in search and detection of a target. With such paradigm, they did find a significant ability of color properties in detection performance over scenes.

In this chapter, a study on color impact has been conducted through an eye-tracking experiment performed on color pictures and their greyscale counterpart. The hypothesis behind this experiment is to measure and characterize the contribution of color. Literature about such topic remains unclear as mentioned previously. Our goal is to clarify it to some extend in order to avoid any potential bias for the recording of eye movements with a color harmony task (Chapter 8).

In addition, we foreseen that the emotional strength embedded in stimuli may also create a bias in visual attention deployment, such as recently demonstrated [126, 187, 107, 187]. In the proposed experiment, we also measure the effect of emotion in visual attention deployment by employing pictures annotated along emotional categories.

The work of Hu-Phuoc *et al.* [103] is probably the most related one to this chapter. Adopting the same methodology, they recorded eye fixations for color and greyscale stimuli from natural scenes and analyze the differences on eye statistics. They brighten up their experiment with abnormal color stimuli to reinforce their conclusion about color impact in the visual attention deployment. They end up with the counter-intuitive statement on a low contribution of the color visual feature in driving visual attention. Liu and Heynderickx [152] also evidence the same assessment with a similar experiment aiming at reducing the complexity of a visual attention model by doing only luminance-based modeling. Having the same purpose, Hamel *et al.* [92] set up recently an eye-tracking experiment with colored and grey pictures. They measured a higher variability between observers for eye movement positions of the color stimuli than for those of grey stimuli. Interestingly, they employed unusual metrics: the dispersion and the clustering of eye fixations.

Several papers have investigated the link between the deployment of visual attention and the emotional factor. Nummenmaa *et al.* [136] and Calvo *et al.* [40, 39] have demonstrated the attention biased potentially created by emotional pictures. They presented concurrent pictures or stimuli having pleasant and unpleasant characteristics. Then, they have measured the statistics regarding the first fixations, the saccade amplitude and the proportion of viewing time for each simultaneous stimulus. They found a significant trend to fixate first on emotional pictures. In [105], the authors presented strongly emotional stimuli (high-arousing erotica and mutilation) to observers and recorded ERP (Event-Related Potential or a brain electrophysical response) activity. These measures revealed that the interaction of attention and emotion varied for specific processing stages and then it has to be specified for each stage. The problem is partially solved for threat stimuli since the neural circuitry is now well established, but

it remains complex for positive or reward-related cues [200].

Very recently, a high interest grew up in the visual attention community for studying cognitive processes related to the emotional factor [126, 187, 107, 187]. In addition to emotionally-annotated stimuli, also the current emotional states of observers has been reported as influencing visual attention deployment [127, 126]. However, this aspect is outside of this experiment's scope. Kaspar *et al.* [126] evidenced an impact of emotion on viewing behavior through two experiments involving positive, negative and neutral stimuli. Humphrey *et al.* [107] employed stimuli with emotional objects and salient regions differently located and reported that the first fixation had more probability to fall within the emotional objects in the case of positive and negative pictures than in neutral stimuli. Niu *et al.* [187] concluded the same for the first five fixations. These recent findings are wider discussed in Section 10.1.

Thus, this chapter is a preamble to the central topic of color harmony developed in the next chapter. It proposes two main contributions:

- The study of color factor regarding eye fixations, more precisely fixation duration, saccade amplitude, as well as the prediction of visual attention models [113, 144] and their performances regarding the two conditions.
- The study of emotion factor regarding eye fixations on the same database where each picture has been classified into four emotional categories: positive-passive, positive-active, negative-active or negative-passive.

7.2 Experiment

7.2.1 Protocol

The proposed eye-tracking experiment has been conducted on 200 pictures by means of a SMI RED 50 RED IView X system with a 50 Hertz sampling (Section 3.2). This apparatus has remote and contact-free features for eye movement studies. Two infra-red camera record pupil reflection and extrapolate gaze position at 50 Hertz. Stimuli were presented on a screen resolution of 1280x972 pixels screen at a distance of 60 cm (35x27° of visual angle). Each subject recording began with 9 calibration points. Twenty unpaid subjects have participated in the experiments. All observers had normal or corrected to normal visual acuity and normal color perception. All were inexperienced observers and naive to the experiment. Before each trial, the subject's head was correctly positioned so that their chin pressed on a chin-rest.

Each subject watched randomly the complete dataset of two hundreds pictures either in their color or greyscale version, knowing the repartition between the number of color and greyscale stimuli was equal. Time presentation was 5000 milliseconds for each image and a grey/neutral image with a randomly located cross was presented for one second between each image to minimize the

centered bias. Ten subjects were recorded for each group of pictures. The subjects were instructed to do natural free viewing of stimuli. Participants were also informed that questions can be asked after the presentation of a stimulus. The questions were only asked in order to keep the subject concentrated on the stimuli. The questions were randomized and were about high level content characteristics such as aesthetic, quality, etc. in order to avoid a subject search strategy.

7.2.2 Database relevancy

The color factor is both easy to subjectively apprehend and difficult to quantitatively measure. The database creation remains a key step in this experiment. First, the pictures has been chosen from a database D annotated regarding their emotional category [3, 162]. Second, the database has been refined by removing "too weak" images in terms of color. Images which may be subjectively qualified as dull, pale are considered irrelevant in the global experiment. Indeed, we considered them as being too close to greyscale pictures too elicit a specific behavior between the conditions. Also, atypical values of luminance contrast may distract the observers' attention to the detriment of color factor. Consequently, colorfulness and contrast metrics have been selected as a measure of color quantity or impact in an image. While the colorfulness C_f is a combination of the color variance and the chroma magnitude C in the CIE-Lab space (L, a, b components) [96], the contrast or root mean square contrast C_t is the standard deviation of image luminance L [192]. Equations 7.1 and 7.2 stand for respectively the colorfulness and the contrast metrics used in the determination of final database D_f.

$$C_f = \sigma_{C_p} + \mu_{C_p} \tag{7.1}$$

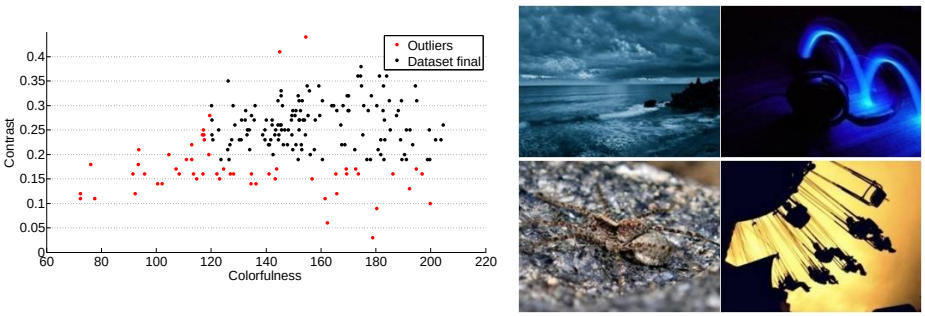
$$C_t = \sqrt{\frac{1}{P-1}\sum_{p=1}^{P}\left(L(p) - \overline{L}\right)} \tag{7.2}$$

where $C_p = \sum_p \left(\sqrt{a(p)^2 + b(p)^2}\right)$, $L(p)$ is the luminance value and $\overline{L}$ is the mean luminance. The p index stands for the current pixel of the considered image having a total number of P pixels.

These two metrics are computed on a image basis for the entire database and form a numerical pair. Thus, images with extreme values of pair would be considered as outliers and discarded from D_f. Empiric thresholds ($T_{lCt} = 0.18, T_{hCt} = 0.4, T_{lCf} = 120$) have been determined to get rid of images I with too low contrast and colorfulness as well as too high contrast values such as depicted in Equation 7.3:

$$\forall I_i \in D,\ D_f = \left\{ I_i \mid T_{lCt} < C_t(i) < T_{hCt} \bigcap T_{lCf} < C_f(i) \right\} \tag{7.3}$$

Figure 7.1 illustrates numerically (a) and qualitatively (b,c) the resulting discarded pictures.



(a) Numeric representation of Contrast/Colorfulness

(b) (92;0.12),(91;0.16)

(c) (178;0.03),(154;0.44)

Figure 7.1: Illustration of dataset selection based on pairs of (Colorfulness;Contrast). (a) For each picture of the original dataset, the contrast is depicted as a function of the colorfulness. Four discarded pictures are depicted in (b) and (c). In (b), colorfulness and contrast are both too low to be considered as a relevant stimulus for the expemeriment. In (c), the blue picture has a very low contrast, while the yellow picture has a too high contrast.

7.3 Characterizing color influence in visual attention deployment

This section proposes to evaluate the similarity of eye fixations recorded on the color pictures and their greyscale counterparts (the two involved conditions). The original assumption is to measure and quantify the influence of color factor in the deployment of visual attention. A first set of results presents a qualitative appreciation of eye fixations maps as well as the statistics reflecting the degree of correlation between the two conditions. Then, a second analysis focuses on saccades and fixations data in order to derive specific cues related to these two conditions.

7.3.1 Visual and statistical similarity

Three metrics, such as the Normalized Scanpath Saliency (NSS), Linear Correlation Coefficient (CC) and the ROC curve, have been computed to estimate the similarity between eye fixations (Section 3.6) measured for color pictures and their greyscale counterparts. The scanpath composed of less than 5 fixations were deleted because they reflect either missing recording or too long fixations associated to visual fatigue. Next, to keep only typical fixations, the fixation duration distribution of each scanpath was computed to discard fixations outside the range [average duration +- 2 x standard deviation]. For each pair of images (color picture versus greyscale counterpart), the computed correlations are averaged over the complete database for each metric. The numerical correlations depicted in Figure 7.2 attest to a high similarity between the two conditions. The higher the three metrics, the higher the correlation. A NSS value above 2 means a pretty high correlation considering a negative value means no corre-

lation. In the same way, a correlation coefficient and the area under the ROC curve above 0.5 means the maps are likely more correlated than taking random samples.

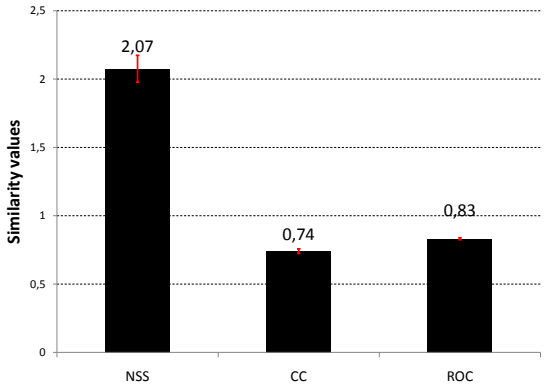


Figure 7.2: Similarity metrics computed between the color pictures and their greyscale counterparts. For each metric, the correlation is computed pair-wise (color versus greyscale counterpart) and averaged over the complete dataset. The standard deviation is illustrated with red bars.

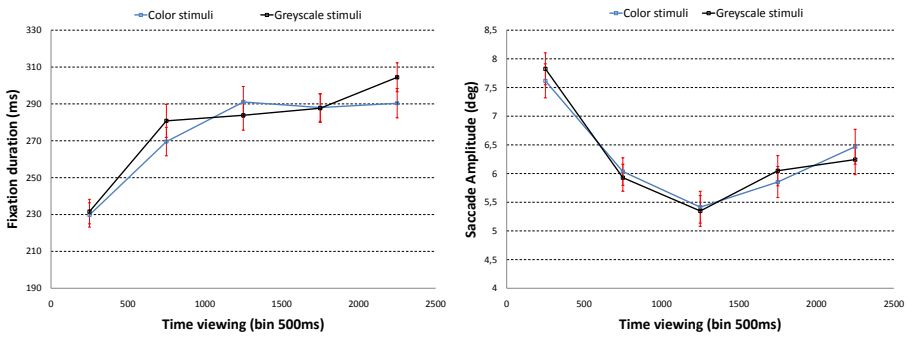
Some oculometric or human saliency maps are depicted in Figure 7.8 and 7.6. The presented examples in Figure 7.8 reflect NSS values from 0.80 to 2.41 computed between the two conditions. One can appreciate qualitatively the high correlation between the maps of the color and greyscale pictures.

For each metric, human saliency maps of color pictures versus grey pictures are highly correlated leading to the conclusion that color does not influence global deployment of visual attention on pictures. Qualitative results on Figure 7.8 demonstrate that semantic meaning and cognitive factors win the "visual attention" competition in front of the low level color factor. Indeed, they are likely responsible for guiding the deployment of visual attention.

7.3.2 Oculometric appreciation

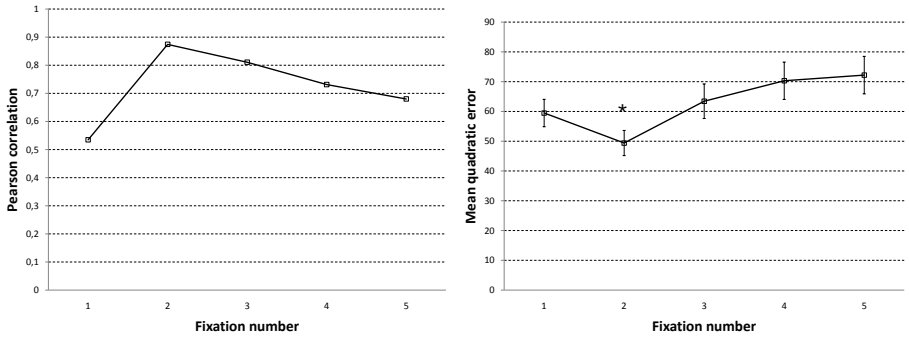
In this section, some accurate properties of eye fixations have been studied: fixation duration (Figure 7.3a), saccade amplitude (Figure 7.3b) and spatial fixation localization (Figure 7.4).

For the two curves in Figure 7.3, we gathered the eye data into bins of 500 milliseconds. Two stages may be identified when appreciating the shape of the curves: early and middle phase, respectively from 0 to 1500 ms and from 1500 to 3000 ms [202]. Similarly to [255, 105], results demonstrate that fixation duration strongly increases just after stimulus onset and then tend to an asymptote around 1280 ms, whatever the condition. The curves of Figure 7.3b indicate that the saccade amplitude decreases over the time course of scene perception. The decrease was stronger at the beginning of the viewing and then becomes stable over time whatever the condition. These curves are consistent



(a) (b)

Figure 7.3: Color stimuli versus greyscale stimuli: (a) fixation duration and (b) saccade amplitude as a function of time viewing.



(a) (b)

Figure 7.4: Color stimuli versus greyscale stimuli: (a) Pearson correlation and (b) Mean Square Error on spatial coordinates. Only the second fixation is significantly inferior to other fixations for the mean square error.

with previous studies [255, 105]. Observing closer the two conditions, there is no significant difference (paired t-tests, $p \leq 0.05$) for fixation duration and saccade amplitude as a function of time viewing.

In addition to the previous indicator, we propose to study the correlation between eye fixations localization for the two categories. Spatial coordinates are potentially an interesting indicator for investigating the differences of visual attention deployment. Pearson correlation and mean square error have been computed for measuring differences between the spatial coordinates over the two conditions (Figure 7.4). The first fixation is the least correlated fixations over the five first fixations (random value close to 0.5). Whatever the context of measure, this result would have been the same; the first fixation location is related to the stimulus with grey cross presented before each image. At the opposite, the second fixation of each scan path provided the highest correlation value between the conditions. Then, the subjects focus on the same area of

interest whatever the presence of color cue. The correlation decreases over fixation number as the competition between bottom-up and top-down mechanisms is increasing. This observation is similar to the inter-observers congruency or consistency [203, 249]. The difference between observers' scanpaths is time-dependent; the consistency in fixation locations between observers decreases with prolonged viewing, assuming that bottom-up or stimulus-dependent mechanisms occurs first. The color component does not seem to influence the eye fixation locations and focal behavior which is usually associated to the beginning of the scanpath [247], but rather the cognitive process associated to later scanning.

In conclusion, the color factor does not influence the visual attention deployment in our experiment with the employed dataset. This result can feed the discussion about the disagreement related to this question in the community as discussed later in Section 10.1.

7.4 Going further

This section brings an additional look on data and proposes to go further than conventionally, looking at the statistics on eye fixations data and maps in the context of saliency models and emotional categorization.

7.4.1 Saliency models

Most of existing computational models, whatever their approach, take advantage of color information when predicting the deployment of visual attention on stimuli. We are interested in this section on the accuracy of their prediction regarding our dataset of color and greyscale stimuli.

The computational model of Itti [113] and the revised version of the LeMeur's one [256] are involved in this study. Their saliency maps have been estimated for the two conditions. Each model has been applied to the color and greyscale stimuli pictures leading to the following notations: *Itti color, Itti grey, LeMeur color, LeMeur grey* for the considering dataset. In Figure 7.5, NSS metric has been computed between the prediction of the considered model and the human fixation maps. As a reference of performances, the NSS between eye fixations from both conditions is also provided. We evidence no significative differences of prediction performances between *Itti color* versus *Itti grey*, as well as *LeMeur color* versus *LeMeur grey*. Only the inter-model prediction (*Itti color* versus *LeMeur color*, and idem for grey condition) brings a significative effect (t-test, $p \leq 0.05$), that translates rather the difference of performances and design between the models.

In order to characterize the contribution of color in the models, it was interesting to cross compare the human fixation maps with prediction maps. In other words, e.g. human fixation maps from greyscale stimuli have been correlated with computational saliency maps computed from color pictures and the other way around (blue bars). This mismatch between predicted and measured data

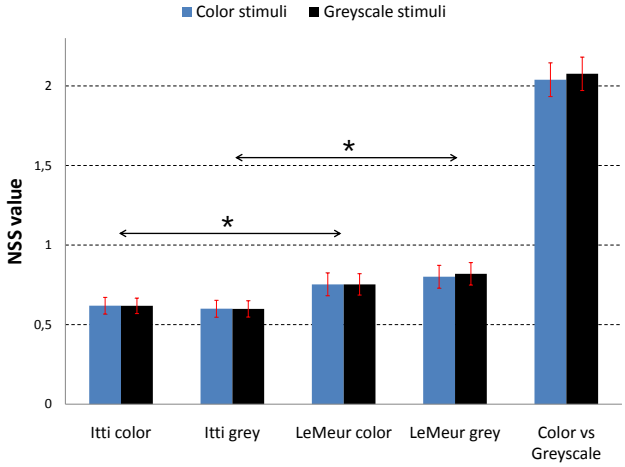


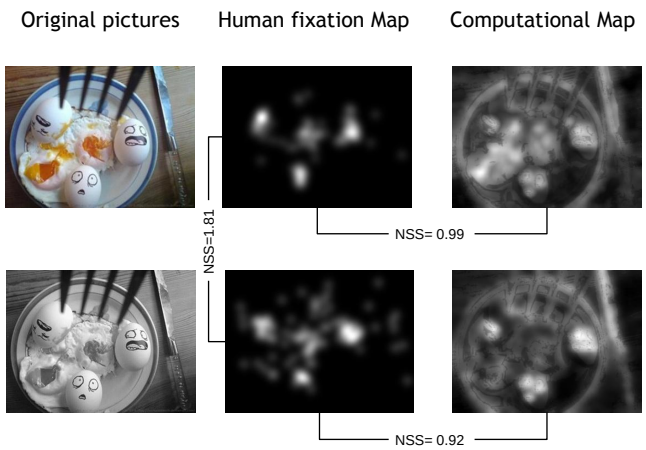
Figure 7.5: Color stimuli versus greyscale stimuli: Saliency models prediction. Itti and Le Meur's models have been computed on the greyscale and color stimuli, then compared to the corresponding human eye fixations. (*) indicates a significant difference (paired t-test, $p \leq 0.05$).

does not change significantly the correlations: the blue bars are at the same level as the black ones in Figure 7.5, advocating again for a non-influence of the color factor.

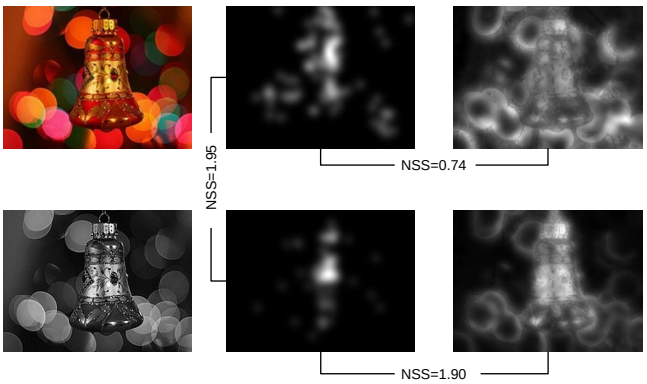
Such observations raise the question about the integration or fusion of color components within the computational model. Probably, it is useless to consider the color factor due to the high correlation of original eye fixations between the two conditionssuch as depicted qualitatively in Figure 7.6). Achieving the same conclusion, Liu and Heynderickx [152] designed a computational model, having originally complexity constraint, that works only on the luminance component. They also evidence the counter-intuitive low contribution of the color factor in the attentional mechanism that do not justify any cost effort in highly constraint environment.

7.4.2 Emotional factor

Originally, the dataset was manually annotated into 8 discrete emotional categories: fear, anger, sadness, disgust, contentment, awe, amusement and excitement. Such representation may be somehow controversial since these categories may induce overlapping of categories. We remapped the image from the discrete categories into 4 more conceptual categories such as positive-active (amusement, excitement), positive-passive (contentment, awe), negative-active (fear, anger) and negative-passive (sadness, disgust). This follows 2-dimensions representation of Arousal-Valence concept. In the literature [200, 105, 40], the visual attention deployment has been studied on emotional stimuli classified in three categories: pleasant/positive, neutral and unpleasant/negative. We propose a slightly finer classification which catches the arousal dimension which has been



(a) Low NSS difference (≈ 0.7) between the prediction in color and greyscale pictures.



(b) High NSS difference (≈ 1.16) between the prediction in color and greyscale pictures.

Figure 7.6: Color stimuli versus greyscale stimuli: Examples of human and computational saliency maps (from [256]) and the associated NSS values.

not intensively studied in the literature.

An evidence of inter-category difference is found regarding NSS values between color and greyscale eye fixations (Figure 7.7): the "positive-passive" category shows up less similarity between the color and greyscale conditions. Nevertheless, this distinction is not found for the other correlation criteria (CC and ROC), then the following findings could be potentially reconsidered in other studies. The "positive-passive" category evidences qualitatively less pictures with clear region-of-interest, but rather "landscape" pictures, leading to less consistency over observers between the two conditions.

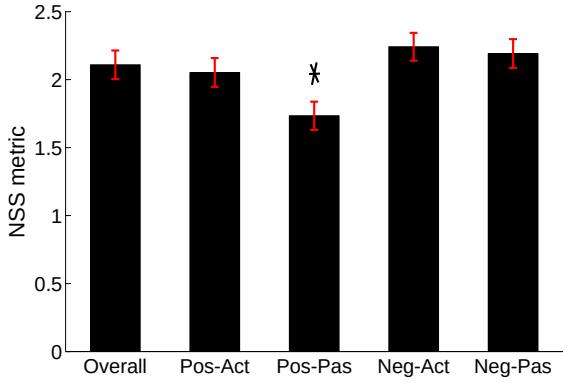


Figure 7.7: Color stimuli versus greyscale stimuli: : Emotional factor. NSS metric computed for the emotional categories between the two conditions. Only the "Positive-Passive" category produces a significant lower (but still high)similarity. (*) indicates a significant difference (paired t-test, $p \leq 0.05$).

7.5 Summary

In this chapter, we propose to study mainly the impact of color feature on the visual attention deployment. An eye-tracking experiment has been conducted on color pictures and their greyscale counterparts. Moreover, the employed database has been previously annotated into emotional categories. Some pictures have been purposely removed from the original dataset based on objective criteria, such as the colorfulness and contrast measures, in order to avoid any bias.

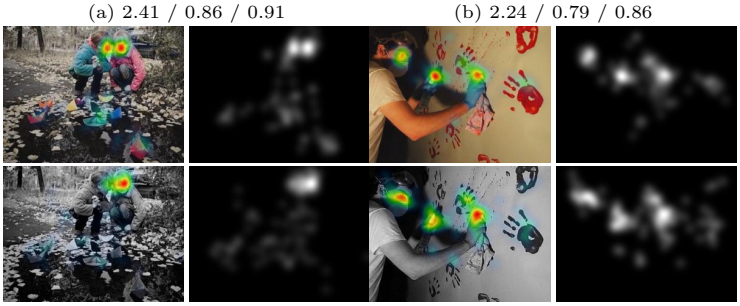
Some correlation measures (NSS, CC and ROC) as well as oculometric statistics have been computed leading to the conclusion that eye fixations recorded for the two conditions were highly similar. In other words, the color information does not impact the visual attention of observers; people keep fixating on the same areas, with the same duration and change to the next area with the same amplitude. These results are counter-intuitive, but can be also attributed to the nature of the dataset.

In the same vein, the computational saliency computed from two state-of-the-art models has demonstrated no significant difference for prediction between the color and greyscale conditions. It raises the question of the integration of such cues in the design of visual attention model. This last point was not a pending issue so far in the literature since it is agreed that color plays a role in visual attention. The compromise is probably the integration of color component within the visual attention models, but weighted by a semantic pre-analysis of image which may determine a priori the potential impact of color within the presented pictures. In this experiment, semantic attributes seem more responsible of the viewing behavior.

The emotional nature of categories present in the proposed dataset was exploited to potentially evidence inter-category difference for the two conditions. Slight category effect has been found for one metric, but not for the two others. Globally, it is difficult to conclude about the impact of the emotion factor.

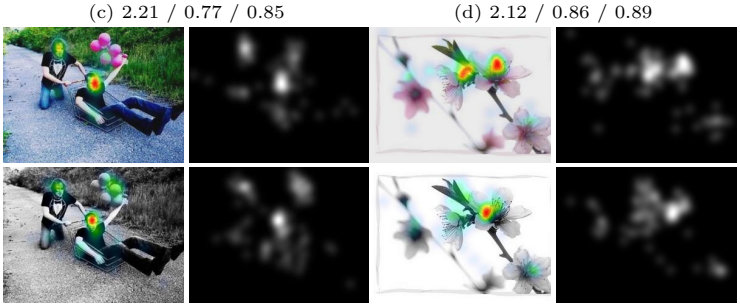
(a) 2.41 / 0.86 / 0.91

(b) 2.24 / 0.79 / 0.86



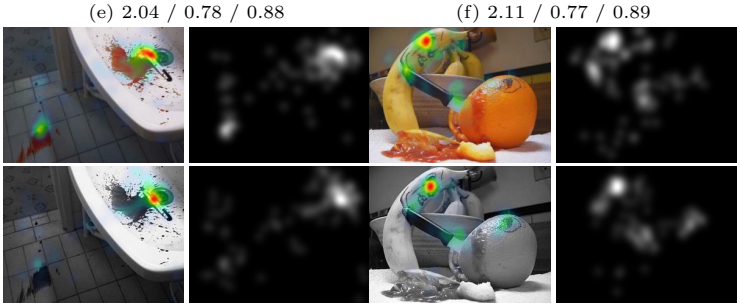
(c) 2.21 / 0.77 / 0.85

(d) 2.12 / 0.86 / 0.89



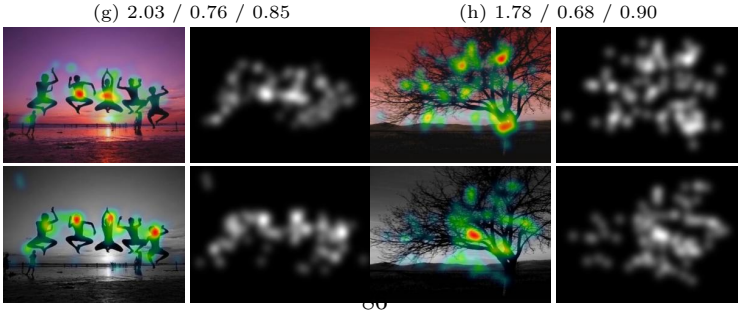
(e) 2.04 / 0.78 / 0.88

(f) 2.11 / 0.77 / 0.89



(g) 2.03 / 0.76 / 0.85

(h) 1.78 / 0.68 / 0.90



(i) 1.39 / 0.69 / 0.93

(j) 0.80 / 0.48 / 0.83

Figure 7.8: Qualitative appreciation of human eye fixation maps: human heat map and saliency map. For each couple of color and greyscale pictures, their similarity metrics are presented in this order (NSS / CC / ROC).

Chapter 8

Experiment 2: Color Harmony

In the previous Chapter, the non-influence of the color factor in the visual attention deployment have been demonstrated during a free-viewing exploration. Regarding the emotion factor, it is more difficult to conclude on its influence on visual attention deployment. We will not conduct further investigations about this aspect. In this Chapter, we investigate the concept of Color Harmony by means of eye fixation recording. As discussed previously, we tackle the assessment of Color Harmony by measuring the visual attention deployment under a particular task. In other words, we define a paradigm that could be the inference of Color Harmony through eye patterns. The experimental setup and the analysis of eye data are depicted along this Chapter.

8.1 Introduction

In this Chapter, we will discuss different issues related to the subjective nature of color harmony in the context of gaze data. As far as we know there is no previous work performing an eye-tracking experiment with a color harmony task. In Section 6.1, we presented some work having implemented a task protocol in the context of eye-tracking. Employed tasks and purposes are vast, thus the experiments show that the prediction of scanpaths is related to the complexity, the own nature of the task. Nonetheless, it is incontestable that a specific strategy is always applied by the visual system for a given task.

As a first contribution in this Chapter, we are investigating the gaze data statistics by proposing a qualitative and quantitative dataset analysis, some statistics on eye fixations and an analysis of the inter-observer agreement. As a second contribution, we carefully selected the involved stimuli in order to study the behavior of specific color distributions. More specifically, we purposely ingested in the dataset some harmonized pictures, processed automatically by a software that adjusts (or remaps) disharmonious hues of a picture, as well as

some pictures having a large color distribution.

Thus, we attempt answering the following questions: Is the harmony task consistent over observers? Is it possible to derive or learn eye movement patterns related to such task? Color is a low-level feature, is harmony related to it or rather inferred at a cognitive stage? In addition, since we controlled the color distribution of different stimuli, we also would like to know if we can characterize a specific behavior depending on the nature of the stimuli.

As mentioned, the employed pictures have specific color distributions that handle complementary, analogous and orthogonal properties of hues. Some of these original stimuli have been post-processed by means of a computational method to generate their harmonized counterpart [24] (Chapter 12). In addition, much complex stimuli with large color distribution are also dropped in the dataset. A two-passes protocol has been designed with such mentioned dataset; these are the two conditions in this experiment: a free-viewing pass serves as a control condition and a second pass measures the eye patterns under a color harmony task.

In the following Sections, we detail the creation of the dataset and provide statistics about it (Section 8.2). Then, we explain the paradigm and the employed protocol of the proposed experiment (Section 8.3). Afterward, we provide three axes for the analysis of results: a focus on fixations and saccades (Section 8.4), a measure of the inter-observer congruency (Section 8.5) and a point about the stimuli with particular color distribution (Section 8.6). Finally, we summarize the findings (Section 8.7) that will be discussed in Section 10.2.

8.2 Dataset Creation

Previous work about color harmony measurement usually restrict their stimuli from having more than three different colors [195, 241, 230]. As the first work on the characterization of the color harmony factor, these studies have purposely limited the number of colors (and their interaction) and controlled some adjacent factors (i.e. luminance, spatial contrast, texture...) that could bias the measurement of color harmony. Thus, the curves and equations fitting the data they collected during their experiment are valid when the interaction between colors and the side features are conform and limited to their experimental setup.

It has motivated our work to evaluate such concept on more complex stimuli. We took care of controlling intrinsic characteristics of our stimuli, while selecting complex images (not patterns). We constituted a dataset of 32 pictures that cannot be qualified as natural pictures or scenes. It is more composed of abstract and artistic pictures.

Our study on Color Harmony directly refers to hue which is the usual component to express color harmony in most of related experiments. In order to have a control and diversity on such parameter, we employed a color search engine available online. The pictures were requested from the search engine TinEye Labs[1] under conditions related to Hue, Saturation, and Value (HSV). Pictures

[1] http://labs.tineye.com/multicolr

are originally uploaded from Flickr©.

In addition, three other criteria motivated our selection: picking only pictures under Creative Commons license[2], balancing the color distribution of pictures and having the most abstract pictures (painting, art effect...) as possible. These two last point are discussed in the next section. Figure 8.1 depicts the pictures selected in the dataset. The two following paragraphs develop how the pictures have been qualitatively selected and provide statistics for characterizing the dataset.

Figure 8.1: Complete dataset created for the Experiment 2 on Color Harmony. It consists of (a) 32 original pictures, having different color distribution on hue wheel that matches the Matsuda's color theory. The seven first rows are related to the 8 Matsuda's harmony templates depicted in (b). Within the 32 pictures, some pictures do not have a straithforward match (unclear row) and others have purposely a large color distribution (last row). The (c) column is the harmonized pictures processed by an image processing tool [24] as well as their resulting color distribution.

[2]http://en.wikipedia.org/wiki/Creative_Commons_license

8.2.1 Distribution: qualitative appreciation

Saturation and value have been requested to be set at 80% in order to favor abstract pictures rather than natural pictures. In a natural environment, the highly saturated pictures are usually over-exposed (leading to clipped areas). However, this is unlikely to reach a very high mean saturation for a natural picture. Otherwise, the pictures is completely over-exposed and unusable.

We expect that abstract pictures (like those in Figure 8.1a) would make the assessment of color harmony possible despite the complexity of the pictures. These pictures are at our sense easier to assess than natural scenes with many different colors and complex elements. In other words, the abstract pictures, such as selected here, are a good intermediary between the simple two-color patterns used so far in color harmony experiments and any natural pictures.

Multiple manual requests have been performed to cover the complete hue wheel, from 0 to 360°. Our attempt is to cover a large and balanced color scale expressed by the hue. The search engine sorts the target results according to a given distance from the color request (manually entered either by Red, Green and Blue components values or a point on the hue wheel) to the color histogram of pictures. Thus, typically the first pertinent pictures have a distribution with a hue peak around the requested hue. This is convenient for our study since we can pick up pictures with limited and controlled number of colors. This makes the measurement and interpretation of potential effect more reliable.

There are still many potential candidate pictures to be selected. Two additional qualitative criteria discriminate the choice of the dataset: the geometric configuration of peaks on the hue wheel and the *semantic presence/meaning* within the picture. The former is related to harmony theory dealing with complementarity, orthogonality and similarity of hues, such as the Matsuda's color templates [168]. As detailed in Section 4.2.1 and illustrated in Figure 8.1b, the templates through their different design, i.e. size and arrangement $(i, L, T \ldots)$, have been qualitatively mapped into the candidate pictures. We selected those that could balance the dataset following the Matsuda's color theory. In other words, the goal is to select a picture set that best balances all template natures for different hue angles.

Additional pictures that could not match Matsuda's templates have been considered within the dataset. In order to investigate the influence of color numbers and more generically the complexity of color distribution, we picked up pictures with either several hue peaks or with continuous hue values all along the hue wheel (last row of Figure 8.1).

Finally, since content itself could influence eye deployment by top-down mechanisms, we qualitatively favor the picture having the fewest objects of interest, being far from natural pictures. Thus, abstract or art pictures would be preferred again.

8.2.2 Statistics: quantitative appreciation

In order to apprehend potential bias of the dataset, we computed first order and second order statistics. Figure 8.2 introduces histogram of hue, luminance and saturation (HLS components) cumulated on all images. The depicted histograms represent the number of pixels (ordinate axis) on the complete dataset as a function of the hue, saturation, log-luminance values. All hues are represented at a certain level, with a slight bias for red values (around 0 and 360°). Due to the constraint on the dataset selection, high values of saturation are more represented. The log-luminance curve is close to a normal distribution, having a skew = -1.02 and a kurtosis = 1.32 which is consistent with other datasets [214].

Second order statistics may be more relevant to characterize a dataset since it may capture spatial relationship between pixels. Spatial frequencies are well-known to be a discriminator in visual attention deployment [131]. We computed the log-power spectrum accumulated on all images as a function of log-spatial frequencies (Figure 8.2). Consistent with literature [223], we found that the curve may be approximated as a line since it follows a power law modeled by $P = 1/f^{\alpha}$ where α is the spectral slope. For the proposed dataset, α was approximated to 3.69 which is higher than usual value, between 1.8 and 2.4, for natural scenes [214]. This difference may be explained by the fact that the dataset does not contain any natural scenes, but rather simpler and artistic pictures. The spectral slope has been identified as a low-level descriptor of texture [25]. The higher the spectral slope, the coarser the texture is. Regarding the nature of our study, this result may benefit our experiments by not biasing eye movement patterns with potential texture attractiveness.

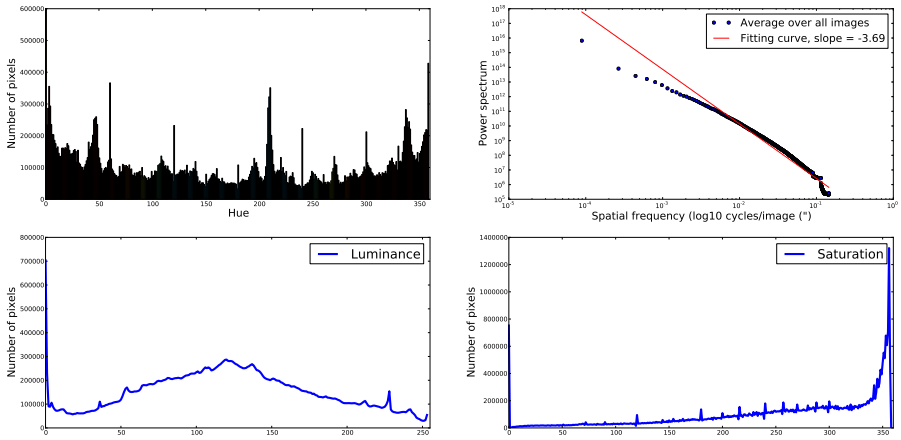


Figure 8.2: Experiment 2: Statistics on the dataset. From top to bottom: luminance histogram, hue histogram, power spectrum and saturation histogram.

8.2.3 Particular stimuli

In addition to previously mentioned stimuli, we added the harmonized counterpart of some original pictures (Figure 8.1c). Those harmonized pictures have been automatically processed by our computational method [24] described in Chapter 12. While in Section 8.2, the pictures are qualitatively associated to one of the eight harmony templates, the computational method quantitatively matches the original color distribution to the *closest* template by minimizing an objective metric. In [24], we proposed the Kullback-Liebler divergence to be minimized between the hue distribution of the original picture and an approximation of the templates distribution. Thus, all hue values that are found outside the sector(s) of the candidate template are remapped into these grey sector(s). For example, in the Figure 8.1, the flower pictures of T template has no blue color after the automatic processing. The T template with a certain angle provided the lowest energy when matched to the original hue distribution and the blue hues are remapped inside the grey sector, just at the border, in green hues. More details about this automatic harmonization processing may be found in Chapter 12.

Also, we intentionally introduced pictures with large color distribution (last row of Figure 8.1a). These harmonized stimuli as well as the large color distribution pictures are addressed in Section 8.6.

8.3 Protocol

Participants

Twenty-seven employees from 20 to 53 years old participated to this experiment with a median age equal to forty-two. There were nine females and eighteen males. The enrollment for participation was organized as a challenge to find the "golden eye" of color harmony, meaning the person that best fits average harmony eye patterns for all images. Participants have normal or corrected to normal vision. They were engaged to participate to an eye-tracking experiments with two different passes on the same set of pictures.

Apparatus

We conducted the experiment with a SR Research Ltd. Eyelink 1000 Hz. The calibration procedure was performed for each pass with an error inferior to 1° of visual angle. A 22-inch calibrated BT709 DELL monitor having a resolution of 1900 x 1200 pixels displayed the stimuli. Participants were located 60 cm far from the screen and positioned on a chinrest to avoid instability of measure due to head movement along time. The experiment setup was located in a dedicated user test room with dark wall and noise free.

Stimuli

Over the thirty-two stimuli composing the dataset, two sets of eleven and one set of ten pictures have been created, noted A, B and C. Each participant was randomly confronted to either set AB, AC or BC, i.e. only twenty-one or twenty-two of them were presented at each participant. This is to constraint the total experiment time under twenty minutes and to limit any effects of fatigue or boredom.

Procedure

Each set of pictures is presented twice to participants, denoted as the free-viewing and task conditions. Before performing the real experiment, participants received a short training to get accustomed to the chinrest, the hue wheel picture and the grey picture (Figure 8.3). For each pass or condition, a grey stimuli with a randomly located cross is presented two seconds preceding the five seconds presentation of the current stimuli in order to avoid any center bias effect of the first fixation. The stimuli are also presented randomly within each pass.

The first pass is a free-viewing exploration of the pictures where participants were asked to look freely at the pictures (FV condition). After a short break, a second pass consisted of watching the same pictures but having a specific task: "watching the most disharmonious colors of the picture" (H condition). The procedure is illustrated in Figure 8.3.

It seemed more relevant to ask for disharmony instead of pointing for harmony areas. Intuitively, harmonious areas are supposed to reflect most of pixels or colors within the picture, while disharmonious area are *inconsistent* colors that do not "go well" in the global picture. Having a free-viewing and a search-task passes allow the study of potential harmony-guided effects under reference eye patterns extracted from the free viewing pass. Moreover, it ensures the feasibility of harmony task that is reasonably difficult in few seconds with unknown stimuli. During the first pass, the observers get more familiar with the stimuli.

The second pass differs slightly from the first pass. An additional question (and then stimulus) was introduced after each picture in order to both estimate the task difficulty and maintain observer's attention to the task. A hue wheel picture (Figure 8.1 or 8.8) was presented following each stimulus where observers were asked for pointing either the hue they identified disharmonious or the center of the wheel if they could not perform the task.

Evaluation methods and preprocessing

For any evidence of significant differences in the next presented results, we computed the Kolmogorov-Smirnov test (KS test) between two sets of data [236]. Since it does not assume a priori knowledge of the samples distribution, it is relevant to gaze data.

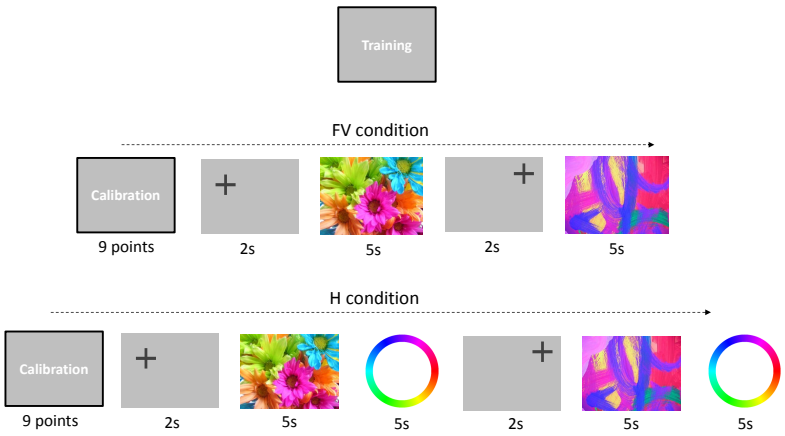


Figure 8.3: Experiment 2: Protocol. Three steps are part of the complete experiment. First, a training is performed to get the participant familiar with the hue wheel and grey stimuli. Second a free viewing pass is recorded. Third, the same stimuli are proposed to the participants with a harmony task.

Due to its ability to measure how closely artificial saliency models match human gaze data, the Normalized Scanpath Saliency (NSS) measure [206] has been employed for measuring the inter-observer consistency and the inter-task similarity. More details are provided in Section 3.6.4.

Since NSS measure does not contain any temporal meaning, another interesting tool is the measure of scanpaths similarity with Eyenalysis tool [167]. More details are provided in Section 3.6.5.

Even if Eyelink 1000 Hz provides efficient sampling rate for data analysis, we chose to extract saccade amplitude and fixation data such as post-processed in the output files of the eye-tracker. Such extracted data are preprocessed in order to remove users with too short scanpaths (under 6 fixations) and the fixations located outside the picture (Section 3.5).

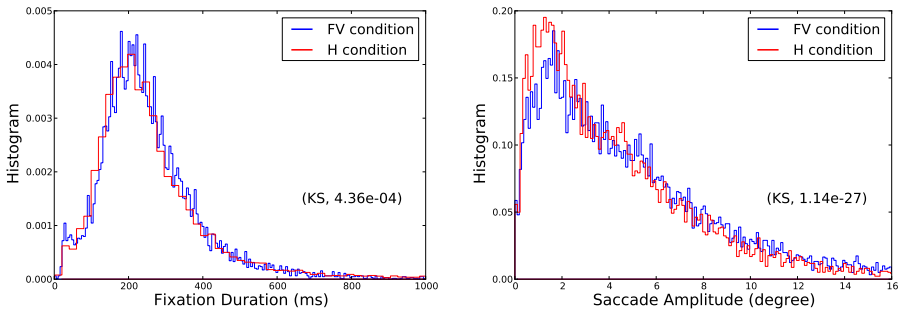
As mentioned in the introduction, we aim at finding eye patterns related to the specific task of color harmony. First, we studied usual statistics on eye gaze related to the two conditions. Secondly, the inter-observer consistency is measured to validate the legitimacy of performing eye-tracking with a color harmony task. Finally, we focus on particular stimuli such as large color distribution and harmonized pictures.

8.4 Analysis of fixations and saccades

8.4.1 Frequency and shape

As usually done in the literature for eye movement, we studied fixation duration and saccade amplitude parameters recorded over the two conditions. Figure 8.4

depicts frequency of fixation duration and saccade amplitude for all observers gathered for all stimuli. Consistent results with literature are found: maximum values of fixation duration are around 200 ms, few fixations are present before 100 ms or after 400 ms [246, 104]. Also saccade amplitude curves depict a typical shape with high number of samples around 2° of visual angle and few of them after 15° of visual angle [14].



(a) (b)

Figure 8.4: Occurrence or frequency of fixation duration and saccade amplitude over the complete dataset. The two distributions have been normalized to achieve an area equal to 1.

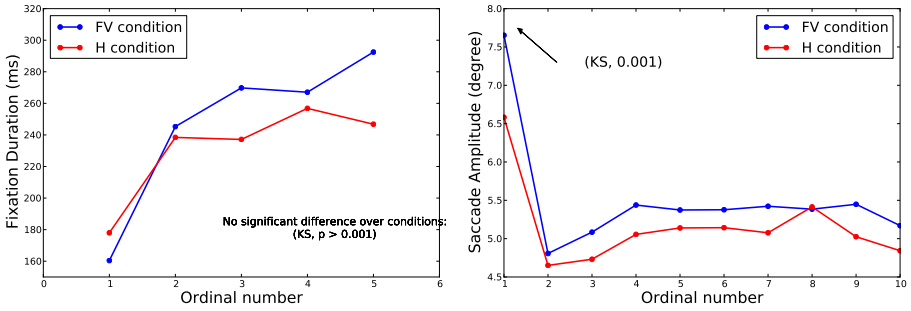
Additionally, we studied these two features as a function of their ordinal number or appearance order (Figure 8.5). We may measure a common mechanism on saccade amplitude and fixation duration: the first element is higher in case of fixation duration and lower in case of saccade amplitude than the followings, whatever the condition. This supports the observations and findings of Castelhano *et al.* [43] for different tasks and those of Ho-Phuoc *et al.* [104] for different processing on stimuli.

8.4.2 Task effect

Having shown the consistency of behavior in Figure 8.4 and 8.5 with literature, we are considering the task condition and more precisely the distinctness between the FV and H conditions. A significant difference was found (KS, $p \leq 0.001$) over conditions for the histogram of fixation duration (Figure 8.4). Dorr *et al.* [70] and Ho-Phuoc *et al.* [104] evidence a different behavior for both fixation duration and saccade amplitude in different conditions. However, there is no consensus on such specific behavior and the differences may be highly related to the task or stimuli nature [70, 104]. Mean fixation duration per observer over all images has been also computed leading to no significant differences over the conditions. The curves in Figure 8.4b depicting the frequency of saccade amplitude show evidence of highly significant difference between the two distributions of respective conditions. More saccade amplitudes are present between 0.5 and 2.5° for the harmony condition. Then, the task pass holds more shorter saccade amplitudes than the free-viewing pass.

While there is a significant effect on fixation duration and saccade amplitude over their global distribution, we did not find any task effect for the first five fixations in ordinal number. Some studies reported increase over the initial viewing of a scene [11, 255], while others found no difference [86] or a decrease of values for a task pass. Once again, the main issue remains the diversity of task and stimuli that prevents any fair conclusion and comparison.

On the contrary, the first saccade amplitude shows a significant difference: around 6.5° versus 7.7° for H pass and FV pass (KS test, $p = 0.001$), respectively. This is also consistent with [43] who found a task effect for saccade amplitude distribution and saccade amplitude in ordinal number. It seems that the strategy associated to harmony task is deployed at the beginning of eye movement patterns.



(a) (b)

Figure 8.5: Fixation duration (a) and saccade amplitude (b) as a function of ordinal number (appearance order). They are depicted for the two conditions.

8.5 Analysis of Inter-observer Consistency

This section aims at quantifying the agreement between subjects in terms of eye movements. More precisely, we are interested in the differences between the free viewing and the harmony task. Is the harmony task as consistent as the free viewing pass (or less or more)? This answer could validate the use of eye movements for the design of human fixation maps in the context of color harmony (Chapter 9). In addition, we also had a look at the agreement for the self-assessment on hue wheel.

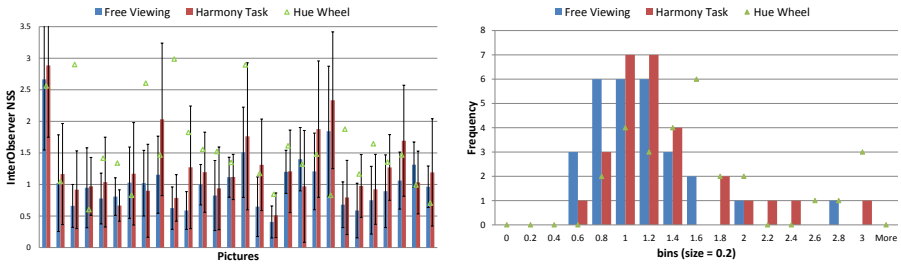
We employed two metrics to evaluate the inter-observers consistency. While the inter-observer NSS (Section 8.5.1) evaluate only the spatial characteristics of eye movements, the second metric of similarity (Section 8.5.2) involves the temporal characteristics of the scanpaths and will allow to confirm or not the conclusion of the spatial metric.

8.5.1 Spatial data: Inter-observer NSS

NSS metric provides quantitative measure of similarity between scanpaths and maps. It is intrinsically a convenient way to handle the inter-observer consistency by competing the scanpaths of each observer against the fixation maps from all other observers (Section 3.6.4).

Figure 8.6a illustrates the mean of inter-observer NSS for each stimulus as well as the associated standard deviation considering the two conditions. All results are positive (better than chance level) and even close or superior to 1 (Figure 8.6b) leading to the conclusion that a high consistency exists between observers for each condition independently. Since the distributions of inter-observer NSS over the dataset can be assimilated to a normal distribution (Figure 8.6b), we evaluate by means of a t-test whether the distributions are significantly different. Specifically, we applied a paired t-test since our observations deal with the same set of subjects, but for two different conditions. We found that the consistency values from H condition are significantly higher than those for FV condition (paired t-test, $p \leq 0.001$). These values are consistent with previous work on variability of inter-observers eye movement for various stimuli [70].

In addition to the consistency over the two conditions, we computed the average inter-observer NSS for the hue wheel stimulus asked for cross-checking the harmony task. The agreement is sometimes much higher than for the two conditions since basically observers do agree on their harmony interpretation. Otherwise, it is comparable or lower than those of the two conditions. Its dynamic level is more remarkable.



(a) (b)

Figure 8.6: Inter-observer Consistency measured with Inter-Observer NSS. (a) Mean values per stimulus for the two conditions as well as for the hue wheel agreement. (b) Histogram of mean values for all pictures and for the two conditions as well as for the hue wheel.

8.5.2 Temporal data: Eyenalysis© similarity

We employed another measure of inter-observer consistency that takes into account temporal features of gaze data. By means of Eyenalysis© tool [167] (Section 3.6.5), we computed the mean similarity of scanpaths per stimuli over the two conditions having as input data the fixation coordinates, starting time (or

timestamp) and fixation duration. Such input vector enables the introduction of temporal information in the similarity measure.

An interesting behavior can be appreciated in Figure 8.7. While FV and H conditions have comparable values of scanpaths similarity (KS, $p \geq 0.001$), the set containing all scanpaths of the two passes leads to significantly (KS, $p \leq 0.001$) lower similarity values. The degree of similarity between the two conditions is comparable, leading to the conclusion that the search-task related to harmony is as reliable as a free-viewing task.

This conclusion does not confirm the previous findings, when using only the spatial characteristics of the eye movements (inter-observers NSS). It seems that the observers agree more on the spatial characteristics of the fixations for the harmony task than for the free viewing task. However, the introduction of temporal characteristics in addition to the spatial locations of fixations set the similarity between the two tasks at the same level. Nonetheless, it is important to notice (in relation to the work developed in Chapter 9) that the eye movements data for the harmony tasks are at least as similar as the one of the free viewing task.

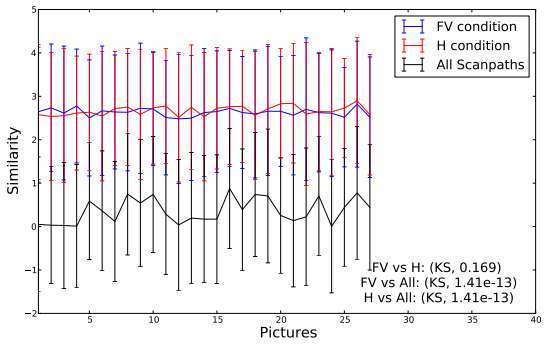


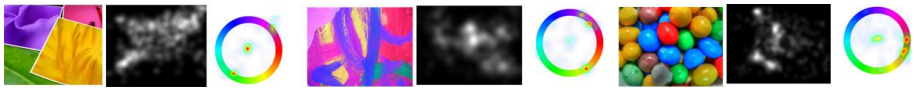
Figure 8.7: Scanpaths similarity including temporal information per stimuli from Eyenalysis tool [167]. The similarity is computed for three sets of scanpaths: from the free viewing, the harmony conditions and all aggregated scanpaths.

In addition, we check out whether eye movement patterns can be classified into two conditions. In other words, is there a specificity carried by the scanpath that can characterize a task? Greene *et al.* [88] make the evidence that task cannot be predicted only from eye movement pattern. We tested the same assumption by applying a k-means clustering (2 clusters and 10 iterations) on all scanpaths. Unfortunately, we ended up with the same conclusion. The classification results in free-viewing or harmony task was purely random. This aspect is discussed in Section 10.2 with recent insights using sophisticated machine learning techniques.

8.5.3 Qualitative appreciation of fixation maps

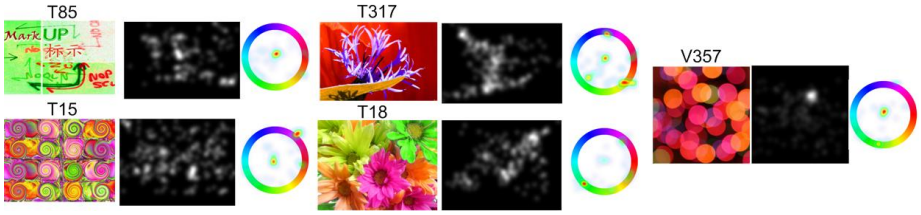
In addition to these statistics, Figure 8.8 depicts a visual assessment of fixations maps for several stimuli recorded during the Harmony condition. In this section, we focus on the three top sets of pictures depicting the original dataset without the large color distribution pictures. Even if the statistical tests are satisfying to validate the consistency of observers for the complete dataset, there are some stimuli that are not following the global tendency. We identified qualitatively three sets of pictures based on results observed from the self-assessment of observers on hue wheel: observers do agree on pointing the center of hue wheel (they do not know which color is disharmonious), obervers disagree on the disharmonious colors, they either point out the center and a color or two different colors, and finally observers do agree on the disharmonious color (most of pictures).

(a) Observers do not know which color to point out

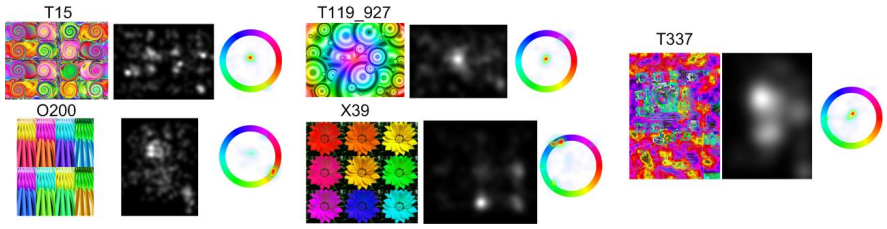


(b) Observers do not agree on disharmonious color

(c) Observers do agree on disharmonious color



(d) Harmonized pictures



(e) Pictures with large color distribution

Figure 8.8: Harmony pass: Qualitative appreciation of fixation maps for stimuli classified in different pictures subsets. The three top sets (a-c) depicts the fixation maps of original pictures into three groups: (a) the observers highly agreed on center of hue wheel, (b) the observers disagrees between colors or/and center and (c) the observers agreed on disharmonious colors. The two last sets are the (d) harmonized and (e) large color distribution pictures.

8.6 Analysis of Particular stimuli

8.6.1 Harmonized set

Figure 8.9 depicts the inter-observer and inter-condition NSS measures specifically for the five harmonized stimuli and their counterparts. We make the following assumptions for eye movement pattern of such stimuli:

1) in the case of free viewing pass (dark and light blue bars), the visual attention deployment should be more consistent over observers for the harmonized stimuli due to the reduction of involved colors (see Figure 8.1 and the histogram depicted on the hue wheel);

2) in the case of harmony pass (dark and light red bars), it should be the other way around, since the specific task of pointing out disharmonious areas should be somehow disconcerting for observers on harmonious pictures.

3) Considering assumptions 1) and 2) verified, we tentatively assume the original pictures would be more consistent over conditions and then have a higher inter-condition NSS measure (dark green bars). This could be related to a very high disagreement during harmony pass for the harmonized pictures (involving noisy maps with no specific areas being fixated) that avoids any appropriate matches between the scanpaths of the FV and H conditions (light green bars). It is actually difficult to predict.

Figure 8.9 depicts the data related to these three assumptions. The second assumption about the H condition (red bars) is partially validated, all NSS measures of the original pictures are slightly higher to those of the harmonized pictures and significantly superior for two pictures (as depicted with a star on the Figure). However, we observed a counter-intuitive behavior when focusing on the FV condition: the inter-observer NSS (blue bars) provided a slightly lower consistency on harmonized pictures versus their original counterparts, significantly only for the T85 picture. This is actually not obvious that observing a harmonized picture in free viewing leads to more consistency between observers than observing the same picture with more colors (not harmonized picture).

Focusing on the inter-conditions NSS (dark and light green bars), we did not find any significant differences between each condition. These correlation between conditions for such specific stimuli are more tricky to analyze.

We qualitatively observed spatial fixation maps recorded on the hue wheel for the harmonized pictures (Figure 8.8d) and their original counterparts (Figure 8.8a and Figure 8.8c). Four pictures (T85, T317, V357, T18) over the five harmonized pictures involved in the experiment lead us to qualitatively classify them into a category where observers do agree on disharmonious colors (Figure 8.8c). In other words, the disharmonious color of those pictures was easy to point out. When presenting the harmonized version of these four pictures, the observers get disconcerted by the harmony task. The wheel fixation maps evidence several colors pointed out or the center of the hue wheel (Figure 8.8d). At the opposite, the fifth stimulus (T15) evidence a different behavior. While

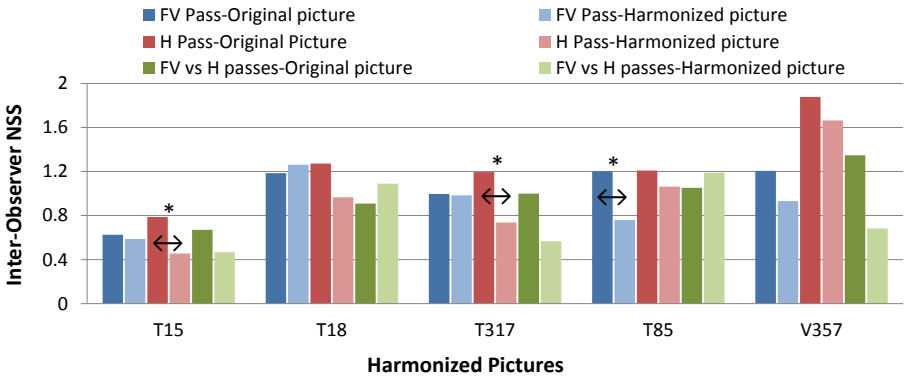


Figure 8.9: Statistics of the harmonized pictures: the data of the five harmonized pictures (light bars) as well as their original counterpart (dark bars) are used to compute the inter-observer NSS measures of the free-viewing pass (blue bars), the harmony pass (red bars) and the inter-condition (green bars) NSS measures. (*) means the difference is significant (t-test, $p \leq 0.05$).

the observer did not know which color is disharmonious on the original stimulus, they tentatively pointed out several areas on the harmonized counterpart. This behavior is probably related to the large color distribution of this particular stimuli: reducing the involved colors supported the harmony task, nut not enough to reach a consensus between observers. This particular kind of stimuli is discussed in the next section.

In summary, corresponding hue wheels highlight in Figure 8.8d that harmonious pictures disconcert observers. They either don't know the color to point out or they highly disagree on candidate, although the disharmonious area was clear on the original version of picture.

8.6.2 Large distribution set

Regarding the pictures with a large variety of color modes, we assumed that they would be difficult to be assessed by observers (following our eye-tracking protocol) during the Harmony condition and that it would be confirmed by the hue wheel assessment. Previous work [195, 230] have employed reduced number of color combination (two or three) to avoid any ambiguity in their experiment of harmony rating. NSS measure reveals for inter-observer (blue and red bars) and inter-condition (green bar) cases a lower agreement than the average on the complete dataset except for X39 stimuli (Figure 8.10).

When comparing qualitatively the characteristics of the concerned stimuli (Figure 8.1), it seems that a large color distribution is not a differentiable factor in the task of harmony assessment, but rather the way the color are spatially grouped. Indeed, the two stimuli where an agreement on disharmonious color is reached (the O200 and X39 pictures in Figure 8.8e) have clear uniform color areas, while for the "dont' know" stimuli (people point out the center of hue wheel in Figure 8.8), colors are more spatially mixed. Additionally, one can de-

duce that distribution with several peaks (or clear modes) (Figure 8.1) is easier to assess than continuous large distribution of hues. We assumed that there

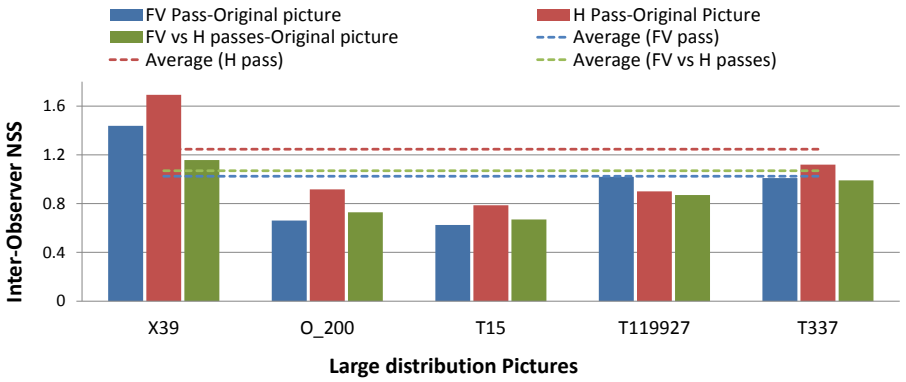


Figure 8.10: Large color distribution pictures: inter-observer and inter-condition NSS. Dotted lines are the average NSS over all pictures for each condition and inter-condition case.

is a relationship between the color distribution and the inter-observer agreement: the larger the color distribution, the lower the agreement due to more choices in fixation locations. We expressed the color distribution by computing the histogram of hue values and by thresholding it according to the number of pixels having a certain hue. Thus, we computed the linear correlation between the hue width (the number of hue values in the histogram having more than T pixels (T=500, representing 0.1 % of pixels for a 800x600 picture) and the inter-observer NSS measure for the H condition. As expected, there is a slight link between these two parameters ($R^2 = 0.12$) that verifies the mentioned assumption. Note for comparison that the same correlation computation between the inter-observer NSS measures for the FV condition and the hue width provided a very low score ($R^2 = 0.01$).

8.7 Summary

In this chapter, we propose to study mainly the eye movement recorded under a task related to harmony assessment. An eye-tracking experiment has been conducted on color pictures, their harmonized counterparts and pictures with large color distribution. Two different conditions are confronted: a free viewing protocol and a search task protocol are recorded using the same dataset.

Some correlation measures (NSS, CC and ROC) as well as eye movements statistics have been computed leading to the conclusion that eye fixations recorded for the two conditions differ, being consistent with task protocols in state-of-the-art studies. More concretely, the distributions of saccade amplitude and fixation duration are significantly different for the two conditions. Also, the first saccade amplitude is significantly higher for the harmony condition than for the free viewing condition.

To ensure the task of harmony assessment is consistent over observers, two metrics have been employed for that purpose: the NSS and the scanpath similarity computed from Eyenalysis tool. These metrics demonstrated a high correlation between scanpaths and similarly for each condition. These results are more deeply discussed in Section 10.2.

Chapter 9

Ground Truth Creation

Contribution: *Chamaret, C.; Urban, F., Lepinel, J., Creating experimental color harmony map, Proc. SPIE 9014, Human Vision and Electronic Imaging XIX, 901410 (25 February 2014).*

This chapter proposes to exploit pragmatically the collected data from the eye-tracking experiment described in Chapter 8. In this latter, we statistically demonstrated a high agreement between observers during a task of harmony assessment. Thus, we aims at creating a dataset by post-processing these human and experimental fixation maps. Such dataset serves to the validation of computational methods developed in the next part of the manuscript.

9.1 Introduction

As exposed previously, the exploration of Color Harmony topic deals either with simple controlled patterns constituted of two or three different colors [195, 241] or with natural pictures, but then assessing the concept of Color Harmony globally [239]. In other words, there is no work that **locally** both predicts and assesses the Color Harmony. One cause could be the lack of ground truth and the difficulty to create it. This is the main contribution of this Chapter.

In the previous experiment (Chapter 8), the self-assessment of observers on hue wheel provides a cross-validation of the harmony task. In the harmony condition, they were asked for pointing out the most disharmonious colors (on the hue wheel) that they previously identified for the current stimulus. Thus, they should confirm the areas they have fixated. If they did not identify any colors, they could point out the center of the hue wheel picture. Some fixation maps on hue wheel are depicted in Figure 9.1c. Before starting this protocol, we did not have a clear view if the task of "pointing out the most disharmonious colors" was understandable and doable in few seconds for our specific stimuli set.

In Section 8.5 and more specifically on the Figure 8.6, we appreciated the

high agreement between observers when computing the NSS measure over scanpaths of all observers per image. This latter is pretty high (> 1) and demonstrates by itself a high consistency between observers. The way is opened to exploit these data in order to obtain an exploitable ground truth.

In this Chapter, we discuss the employed methodology (Section 9.2), then we describe the different steps to split the original dataset into different clusters (Section 9.3) and finally we explain the applied processing (Section 9.4).

9.2 Methodology

A deeper analyze of the data collected during the previous experiment lead us to suspect two different behaviors especially when confronting the Inter-Observer NSS coming from the hue wheel and from the human fixation maps (Figure 8.8): 1) the agreement is partly due to visual features different from color harmony, 2) there are different stimuli classes where the agreement may be balanced and interpreted.

In Chapter 7, we discussed the unexpected non-influence of color components in visual attention deployment for a free viewing exploration. These conclusions have been relayed by other references in the literature [104, 152]. However, even if we can make the assumption that the color factor does not influence the visual attention deployment, at least other low-level features (luminance component, texture, spatial frequencies...) do. In which extent, the agreement is related to color harmony factor and not to these mentioned low-level features? This is difficult to quantify. In other words, it is difficult to decorrelate the bottom-up mechanisms from higher level mechanisms; both are competing after several milliseconds [71].

When having a look on particular stimuli (Figure 8.6a), we appreciate different values range, e.g. some stimulus have similar NSS values for human fixation maps (whatever the condition), while not for the hue wheel. The other way around for others, their NSS measures on hue wheel are similar, but not those for human fixation maps on stimulus. Clearly, these observations could be more investigated in order to identify the cause of such differences. It seems that these different behaviors could lead to categorize stimuli.

In Figure 8.8c, we introduce two complementary representations for apprehending qualitatively the human fixation maps: the heat map and the transparency map. While this former clearly point out the majority areas of fixations (red spots), this latter gives a feedback on the spatial dispersion of fixations. For the T85 picture, observers pointed out mostly the pink colors on hue wheel, but also a bit of green such as revealed by the transparency map. Also, their choice is split between the blue and green colors for the L355 picture.

More concretely, we have qualitatively observed three different cases (also mentioned in Figure 8.8a, 8.8b, 8.8c):

1. observers do agree on disharmonious color,

2. observers do agree they cannot identify any disharmonious colors,

3. observers do not agree, they point either different colors or a color and the center of the hue wheel.

In addition to this statement, it seems tricky to directly derive some color (dis)harmony maps from the human fixation for playing the role of ground truth. These latter are noisy; even if they contain the areas that observers wanted to point out, there are *parasite* fixations that result from a more global exploration or analysis of the picture. From all these observations, the main idea developed

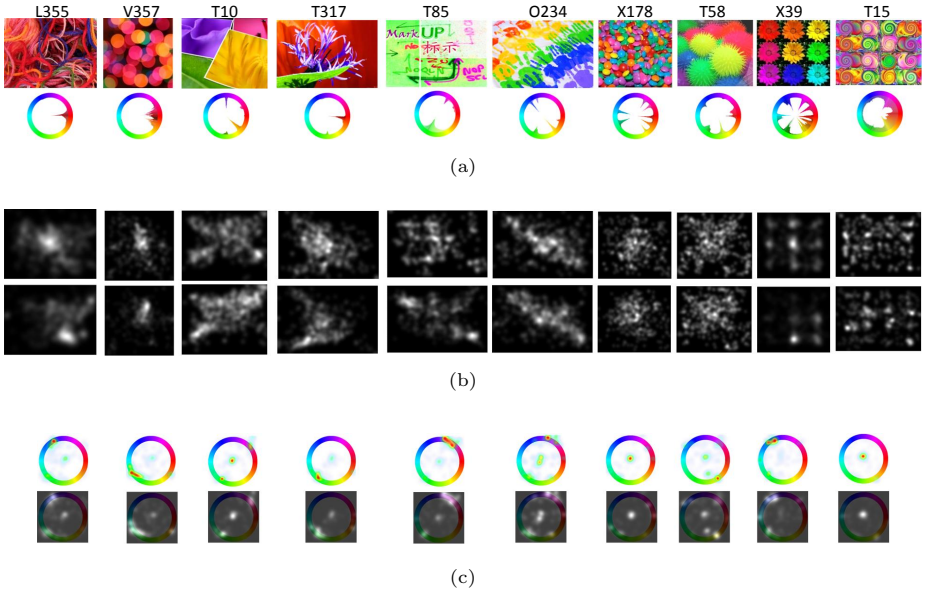
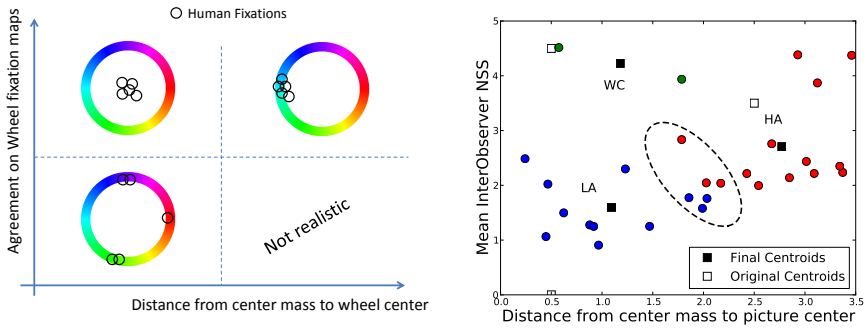


Figure 9.1: Qualitative appreciation of observers' fixation maps: (a) Original stimulus and hue distribution of original stimulus, (b) free-viewing human fixation maps (top row) and human fixation maps for harmony task (bottom row), (c) observers' hue wheel assessment: heat map (top row) and transparent luminance representation (bottom row).

in this Chapter is 1) to computationally find three classes of stimuli that would report from the agreement uncertainty, 2) to design a specific post-processing in order to treat each data subset according to the reliability associated to data. In other words, we propose in the following section to cluster objectively the pictures into three different classes and then to apply a post-processing specific to each class.

9.3 Pictures Clustering based on Agreement

A high agreement of fixation locations on the hue wheel (measure by Inter-Observer NSS) does not mean that observers have clearly and easily identified disharmonious colors. Indeed, the agreement may be high if everybody point



(a) (b)

Figure 9.2: Dataset clustering into three categories reflecting the degree of agreement for color harmony assessment. (a) Expected categories from the employed dimensions. (b) Result of the dataset clustering for color harmony agreement into three categories: low agreement (LA), high agreement (HA) and high agreement on hue wheel center (WC). Note that the axis of clustering correspond to whitened data and not to the original values of each axis. The dotted ellipse depicts some ambiguous classification case.

out the center of hue wheel, although it indicates that observers did not find out an answer to the task. Also, more than one hue could be potentially not harmonious at different degree but each of them could be pointed independently by different observers. Three categories defining the level of agreement will be quantitatively identified by a clustering method.

For each fixation map from the hue wheel, we computed the center of mass that indicates how far the fixations are from the picture center. It provides a first input to know if observers tend to agree on center of hue wheel or on one or several colors of hue wheel. We coupled this information with the mean of inter-observer NSS computed from the hue wheel stimuli, since it gives an information on global agreement for a given picture. Figure 9.2a illustrates the three expected cases that we expect to highlight with the clustering method. If the distance from the center of mass to the wheel center is low, there are two possible configurations: either most of eye fixations are located on the wheel center (meaning the agreement between observers is also high) or there are several colors pointed out on the hue wheel and spread around that lead to a center mass located in the center of the wheel (meaning also the agreement between observers is low). The last ideal case is encountered when most of eye fixations are located on one area of the hue wheel (meaning the agreement between observers is also high).

Following these expectations, a k-means clustering function has been applied on these two dimensions for each stimulus: the number of requested clusters has been set to 3, the number of iterations to 50 and finally the centroids have been initialized with a value located in one quadrant (Figure 9.2a). The result allowed us to separate the images into high agreement case (HA), low agreement case (LA) and high agreement on the hue wheel center (WC), such as depicted in Figure 9.2b.

We found fourteen pictures with high agreement (HA cluster) on disharmonious color, twelve pictures for low agreement (LA) on disharmonious color and two pictures having a high agreement for no identified color (WC). As expected, there is no pictures located in the right lower corner. This result is pretty close to our qualitative appreciation performed on hue wheel fixation maps (Section 8.5.3). However, there are pictures at the cluster's intersection that could be misclassified, such as circled on Figure 9.2b.

Having found such clustering, we aim at creating some experimental color harmony maps out of 1/ human fixation maps on stimuli and 2/ human fixation maps on hue wheel, recorded during the harmony condition of Experiment 2.

9.4 Designing Color DisHarmony Maps

We propose in this section to post-process data collected from two different sources of information: 1/ the fixation maps recorded on the original stimuli and 2/ the fixation maps recorded on the hue wheel picture during the harmony condition of the Experiment 2 (people were asked to point out the most disharmonious colors in the picture). We aim at designing consistent color disharmony maps that would be reliable as a ground truth.

As discussed in Section 10.2, the stimuli have different color distributions (similar hues, opposite hues, orthogonal hues or even large distribution) that impact the confidence in the observers assessment. Following this fact, the stimuli have been clustered depending on the inter-observer agreement (Figure 9.2b). Thus, we keep employing the clusters determined in the previous section for which a dedicated method is empirically determined in order to end up with color disharmony maps from eye fixations.

Figure 9.1 depicts some fixation maps recorded during the harmony pass. Color disharmony maps could have been the direct use of such map, since the observers usually point out accurately the color in relation with the task. However, there are clear limitations to employ such simple protocol:

1. All eye fixations are not be related to the task, some low-level processing are probably involved and parasite a purely eye fixations map created from a *Harmony* task.

2. The *Harmony* task is not straightforward and observers may hesitate or look for target colors during the beginning of stimuli scanning.

3. Even if the disharmonious color is easily identified by observers, these latter can not fixate all pixels having the target color. Moreover, a hue is not involved alone in the disharmony perception but it is rather a range of similar hues.

Based on this statement, we define three ways in accordance to the three picture clusters to design the experimental disharmony maps.

9.4.1 Method

Fixation maps Distribution

First, we derived the two 1D distributions of fixation map values as a function of hues.

The first distribution is computed from the fixation map recorded on the stimulus FM_S, where the value at each site $u = (x, y)$ of the fixation map are accumulated for each hue h:

$$S(h) = \sum_{u|h(u)=h} FM_S(u)) \tag{9.1}$$

The second distribution is computed from the fixation map recorded on the hue wheel FM_W, by accumulating the value of the fixation map also for each hue:

$$W(h) = \sum_{u|h(\theta,r)=h} r(u) * FM_W(u), \tag{9.2}$$

where $\theta(u)$ and $r(u)$ are respectively the angle and the radius of the pixel location in polar coordinates. Note that the contribution of FM is weighted by r, i.e. the distance between the hues crown and the wheel center.

Analysis of Distributions

We are interested in the match between these two kind of information. Are they consistent, complementary, antagonist for some hue values?

Figure 9.3 provides qualitative appreciation for these two distributions. We can observe that the distribution are not well aligned. Even if the hue peaks are located in the same regions, they may not be aligned enough to evidence a good consistency. We argue that it is due to the inaccuracy on hue wheel assessment. Indeed, several observers reported after the experimentation, a discomfort to match the "right" blue or green on hue wheel. The saturation or luminance values influence the hue perception. They were annoyed by the static hue wheel stimuli (medium luminance and saturation) that did not match exactly the stimuli perception.

Additionally, the inaccuracy of fixation maps on original stimuli may be also reported. Observers do not realistically fixate all the time on disharmonious colors. Thus, when a background colors is very important with regards to the disharmonious colors, its hue may parasite the 1D distribution computed on original stimuli. This is typically the case for the i211 picture on Figure 9.3; the blue sky report fixations that creates a peak around 200.

A method for each cluster

The HA set clustered in Section 9.3 is basically the one with the highest confidence measured on observers. Thus, we simply decided to reuse such information alone. The W distribution is remapped in the 2D stimuli space to design the final

disharmony map D_{HA} of such category of stimuli, meaning $D_{HA}(u(h)) = W(h)$. Some examples are provided in Figure 9.3.

The LA set of stimuli needs more information to get reliable disharmony map. We propose to merge information coming from the two available source of information: fixation maps on original stimuli and on hue wheel. More specifically, the final disharmony map is computed from W distribution and refined with the S distribution from fixation map on original stimuli:

$$D_{LA}(h) = \left(W(h) + \frac{W(h)}{|W(h) - S(h)| .G} \right) . \frac{255}{[W(h)]^+} \tag{9.3}$$

G is a gain empirically set to 3000. $[.]^+$ represents the maximal value of the distribution. If the information between W and S coincide at a given hue, the ratio $\frac{W(h)}{|W(h)-S(h)|.G}$) tends to infinite, D_{LA} is very high. On contrary, if W and S are not aligned for a given hue, meaning S or W are close to 0 but not the other quantity, then the ratio $\frac{W(h)}{|W(h)-S(h)|.G}$ is close to 0 and D_{LA} is equal to $W(h)$ as in the HA case.

The obtained 1D distributions $D(h)$ are remapped into 2D image space $DM(u(h))$. Several qualitative disharmony maps for the different category are depicted in Figure 9.3.

The disharmony maps of WC dataset are more tricky to design. Even if a majority of observers pointed out the center of hue wheel for such stimuli, few of them may have discriminated a color over the distribution. However, it is not reliable enough to design a way to create color harmony maps for such stimuli.

9.4.2 Maps Appreciation

The examples of Figure 9.3 depict the final harmony maps for the two categories of stimuli as well as one example of maps for the WC dataset. As appreciated for the HA dataset, the hue wheel map provides more reliable results than fixation map from original stimuli. Indeed, the 1D distribution is not confused by the fixations of the dominant color (pictures i211, V16, V357). Moreover, the hue wheel provides more focused and relevant colors; few colors are selected. On contrary of fixation maps from original stimuli, few colors may interfere with disharmonious colors (pictures T18, T85).

For the LA dataset, where the agreement is less obvious between observers, both fixation maps bring interesting inputs for computing the final disharmony map. In picture L355, the blue and green lines are intuitively the disharmonious colors to be identified. However, the green color is not identified from the hue wheel, but the use of fixation map from the original stimulus allows enhancing this color in the final disharmony map.

The last row evidences the high uncertainty in the establishment of rules for creating disharmony map for the WC dataset from fixation maps of hue wheel and of original stimuli.

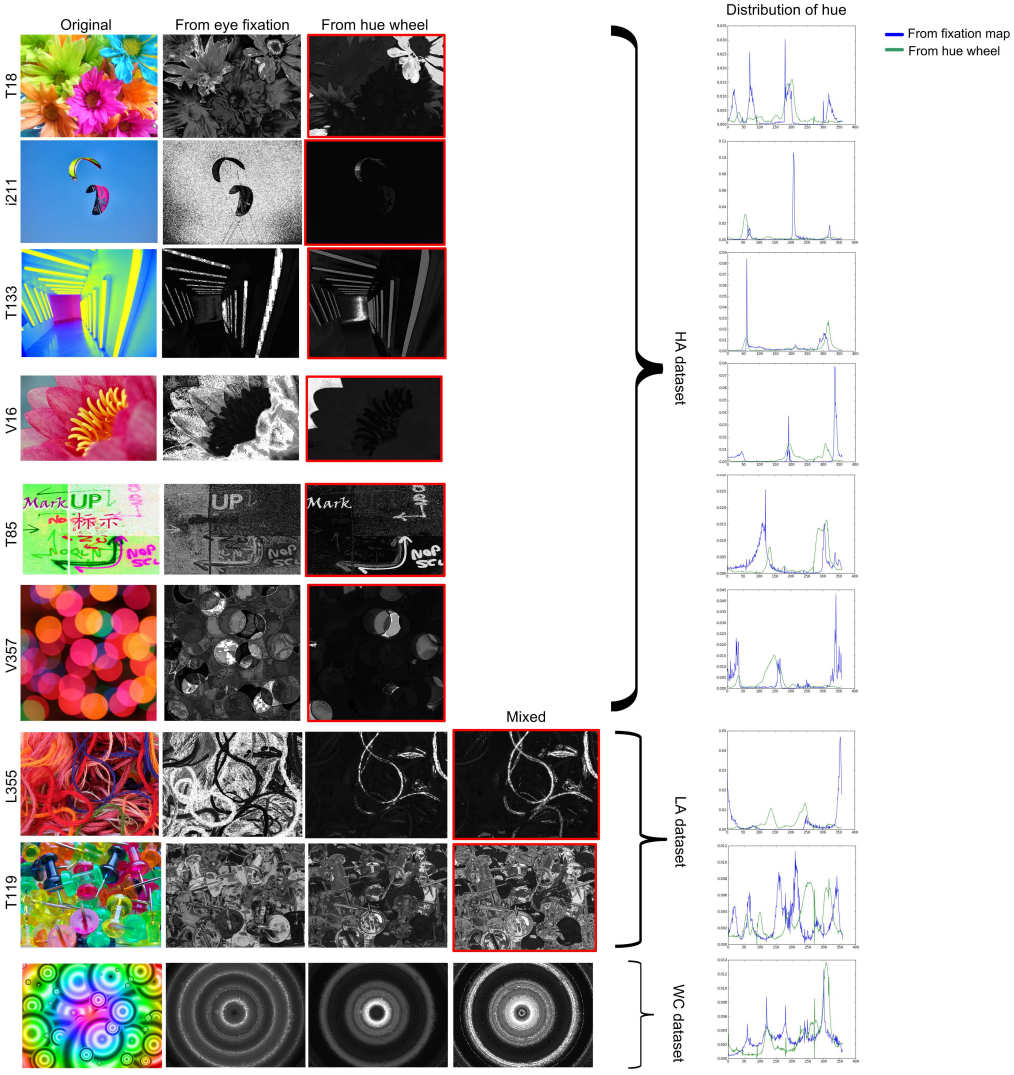


Figure 9.3: Disharmony maps designed for the HA and LA categories from the eye fixation maps recorded on both hue wheel and original stimuli. On the right-hand side, the two 1D distributions in hue domain are depicted to appreciate the potential match. Red squares are the final disharmony maps: the witter, the more disharmonious the pixels.

9.5 Summary

The main idea developed in this Chapter is 1) to computationally find three classes of stimuli that would report from the agreement uncertainty, 2) to design a specific post-processing in order to treat each data subset according to the reliability associated to data.

We analyzed the inter-observer agreement on such data and derived three categories of pictures: the "high agreement", "low agreement" and "do not know" cases. Based on such clustering, different strategies are proposed to design the final disharmony maps. The fixation maps recorded on the original stimuli as well as those on hue wheel are combined to catch different colors and to remap the disharmony values of all pixels. This is a first attempt to provide a ground-truth in the context of color harmony in order to help people applying this principle. This partially answers the need and current limitations of color harmony field mentioned in Section 5.4.

Chapter 10

Discussion

This chapter proposes to discuss the different findings of the three preceding chapters. This is the place to come back to the different formulated hypothesis in the introduction (Section 6.4).

10.1 Experiment 1

This experiment (Chapter 7) introduces the second experiment (Chapter 8) dealing with the specific problem of color harmony. While color is a low level factor, that can be objectively measured (colorfulness, 3D histogram of color components...), the emotion associated to a picture are likely related to high level cognitive factors that are difficult to quantify and to apprehend. The link between emotion and color is not straightforward to characterize.

In Section 6.4, we formulate two hypothesis:

- Hypothesis 1: The color factor influences the visual attention deployment.
- Hypothesis 2: The emotion factor plays a role in visual attention deployment. We refute these two assumptions, sometimes in accordance with state-of-the-art or in discordance with recent studies.

We demonstrated that color has no influence on visual attention deployment as a low level factor for our dataset. In addition, emotional categorization can not be differentiated from eye movement data for the two color conditions, except for one category. Nonetheless, this result is not supported by all employed metrics.

Color factor

We could split the studies around color factor into two balanced teams (Section 6.3): pros and cons of color influence in visual attention deployment. Fortunately, we obtain the same conclusion as Ho-Phuoc *et al.* [104] who used a similar protocol. Other work did report an influence of color on visual attention, but to a lesser extent [249, 10, 9]. Tatler *et al.* [249] reported that the luminance

and chromaticity features discriminate 57% of fixated areas, meaning that the remaining percentage is induced by other factors, unrelated to image statistics. Amano *et al.* [10, 9] also found a comparable influence of color properties during a search and detection task.

More globally than color, the low level features are being reconsidered in the literature as a relevant factor of visual attention. Their role in visual attention deployment is controversial and more generally the role of stimulus-driven attention seems more and more minimized to the favor of task-driven mechanisms. This leads to the questionable validity of low level cues in the design of visual attention models.

On the other hand, some studies reported a concrete influence of chromaticity on visual attention [81, 82, 238]. In these studies, the authors carefully controlled their stimuli. Either they employed isoluminant patterns with the color being the only changing factor [238] or they have strict image categories with similar statistics [81]. We did not control so much the potential factor of bias in our dataset, since the goal was also to measure the influence for the emotional factor in parallel. Thus, our stimuli are semantically very rich and emotionally laden (faces, knifes, children, blood...). This aspect could have taken the lead as a mechanism in the selection of pointed areas. If verified, this is also an indicator of the low implication of the color feature in visual attention deployment.

Emotional factor

Kaspar *et al.* [126] set up two experiments to measure the impact of emotional-laden stimuli on viewing behavior under natural conditions. They employed pictures from the IAPS database [140] at the extremity of the valence scale. In other words, very high and very low valence ratings guided their selection for positive and negative pictures. They also introduced in their two experiments some neutral emotional stimuli (fractal images) as well as target images with a negative, neutral and positive context. In the second experiment, they had a look also on scene categories (nature, urban and fractal types). Globally, they found an influence of emotion on visual attention deployment. More specifically, the negative pictures evidenced shorter fixation durations, longer saccades and more spatially spread fixation distribution than positive stimuli. This was encountered when having a emotionally-randomized presentation of stimuli, but not when the stimuli were grouped emotionally during the presentation. This suggests that watching such strong emotional stimuli influences the inner emotional states of the observer leading to a different attention strategy. Finally, they also evidenced an impact of image category on viewing behavior, e.g. urban scenes did not provided the same evidence as nature scene.

In our experiment, we did find an impact for one emotional category. However, when using different metrics we could not met again the same effect between the two conditions. The experimental conditions of Kaspar *et al.* differ from our pioneer work on emotional pictures. We described the emotion along four categories involving also the arousal notion. The finer categorization we

employed may explained why we did not reach the same conclusion. In addition, there is likely an effect due to the employed database. While Kaspar *et al.*'s experiment relies on very emotional-laden stimuli, we employed a database potentially less controlled in terms of annotation and more questionable regarding the categorization.

10.2 Experiment 2

In the previous experiment, we found that color should not interfere as a standalone factor on eye fixation statistics, when designing a task-related protocol. This is also the same conclusion for the emotional content of pictures. Assuming color harmony stands at a higher level of brain mechanisms, it makes sense to design an eye-tracking experiment with a color harmony task or instruction without exposing ourself to any bias from the color choice or arrangement.

In Section 6.4, we formulate two hypothesis:

- Hypothesis 3: during a task protocol with eye-tracking for measuring color harmony, the inter-observer consistency is high.
- Hypothesis 4: there is an influence of color distribution and color diversity in the assessment of color harmony.

We attempt to confirm these hypothesis through the analysis of data recorded from the experiment described in Chapter 8. Below is a discussion related to it.

Is the color harmony task realistic?

The inter-observer agreement makes the evidence that people can reasonably point out the color (dis)harmony within a picture at a spatial and local level. The inter-observer NSS measured in the experiment is significantly superior for the harmony condition compared to the free-viewing condition. Observers agreed on the assessment of disharmonious areas. The slight increase in harmony condition may be also related to the fact that observers have already seen the stimuli during the free-viewing condition. However, directly assessing the color harmony appeared to us a too difficult task, leading to the introduction of the first free-viewing condition. Consistent with the previous metric, the similarity of scanpaths is in the same range for both free-viewing and harmony conditions.

In addition, we investigated qualitatively and individually the different stimuli. We toned down the previous evidence, computed globally, by studying the disharmonious color pointed out on the hue wheel stimuli. We found three sets of agreement: when observers agreed on the hue center (they do not know which color to point out), when they disagree by pointing out two colors or also the center of hue wheel, when they agreed on a disharmonious color. Fortunately, most of pictures are in the last category, but we investigated an objective method to automatically classify the pictures into one of the categories (Chapter 9). Assessing the color harmony locally is reasonable with eye-tracking.

Any evidence of eye patterns related to color harmony?

Eye features such as saccade amplitude and fixation duration are usually relevant for characterizing eye movement patterns in specific conditions. Processing such data allows including temporal effects of visual attention deployment that cannot be caught with purely NSS and spatial statistics. Multiple studies highlighted a task effect on fixation duration factor [137, 246, 43]. Over the complete dataset, fixation duration and saccade amplitude provided evidence of significant differences for their global distribution (KS, $p \leq 0.001$) between the two conditions (free viewing and harmony conditions). In the same vein, task-effect is reported for the first saccade amplitude in ordinal number, but neither for the first five fixation durations nor the nine following saccade amplitudes in ordinal number.

If the task was carried by eye movement patterns, it would have been possible to classify blindly all scanpaths from the free-viewing and harmony conditions into two corresponding distinct categories. Clearly, it does not work that way for our stimuli and conditions, such as found also by Greene *et al.* [88]. Scanpaths born of harmony task are not statistically separable from those of free-viewing pass, meaning that low level processing (typically involved in free viewing scan) is still involved and competes with any cognitive processing that could interact for such task. Rather than splitting the global set of scanpaths into two balanced categories, we handle with widely unbalanced clusters (using k-means clustering) that seem to point out potential outliers for both conditions. This result is consistent with Greene *et al.* [88] and demonstrates that the color harmony task is not differentiable by only eye movement patterns, at least by using a simple clustering method. Recently, Borji and Itti [31] evidence the power of using more sophisticated machine learning techniques (e.g. k-nearest-neighbor and boosting) in order to predict a task from eye movements. This promising approach has not been investigated with our data.

Is the color distribution playing a role?

Studying color distribution of involved stimuli also brings new understanding. We assumed that the color distribution, defined by either its hue variety or its hue width, could be a factor influencing the reliability and perception of harmony within a picture [230]. We evidence a correlation between the inter-observer NSS and the color distribution width computed per stimulus ($R^2 = 0.12$). This conclusion has been refined by showing that many peaks in the hue distribution are not responsible for low agreement, but rather moderate hue levels spread around the hue wheel. In addition, the close spatial repartition of hue allows a better and faster understanding of the harmony.

The particular behavior of observers on post-processed pictures [24], i.e. harmonized stimuli related to color harmony theory [168], has been carefully studied. The agreement measured on hue wheel, where observers pointed out the disharmonious color, drops when confronted to harmonious pictures. The specific task of assessing the disharmonious color on harmonious pictures discon-

cert the observer, while he successfully agreed on pointed out one disharmonious color on the original stimuli. This is a first step toward the validation of the image processing algorithm performing the automatic color harmonization of picture (developed in Chapter 12).

In conclusion, we strongly believe there is an agreement for spatial assessment of color harmony, but its degree of validity should be carefully related to the nature of the stimuli [246] and more particularly to its color distribution and spatial arrangement.

10.3 Ground truth creation

The definition and the characterization of color harmony started to be investigated some centuries ago. Today, even if no consensus has been reached in this field, some image processing tools (recoloring, automatic harmonization, aesthetic metric...) based on color harmony principles are clearly emerging. Thus, the need for a ground-truth grows up at the same time.

We proposed an experimental protocol to create color harmony maps for complex stimuli knowing there is no relevant previous work on this topic. We chose to take advantage of the Experiment 2 (Chapter 8) using the eye fixations collected with a color harmony task.

Since we introduced a computational post-processing of eye fixations, we are facing the issue of validating the ground truth. This is an egg and chicken problem: how to validate what will be used for the validation later of computational methods? We could not provide any other way than a visual appreciation of final disharmony maps based on empirical assumptions and observations.

10.4 Conclusion

In this part, we tackled the topic of Color Harmony from an experimental perspective. We assumed a strong relationship between visual attention and color harmony. Therefore, we investigated the measurement of such factor through an eye-tracking protocol having a search task. We evidenced a high inter-observer agreement that demonstrated the universal aspect of the color harmony notion. Such way of experiencing Color Harmony is pioneer and may suffer from deeper analysis. However, it paves the way to further investigations.

We proposed some hypothesis in Section 6.4 that we attempted to validate along this part.

Experiment 1

Hypothesis 1: The color factor influences the visual attention deployment.

Findings 1: Measured through our dataset, it is not the case. The human fixation maps compared between color stimuli and their grey counterparts are not significantly different.

Hypothesis 2: The emotion factor plays a role in visual attention deployment.

Findings 2: It does only for a specific category of emotion. However, such finding was not confirmed with other metric applied on human fixation maps between categories.

Experiment 2

Hypothesis 3: During a task protocol with eye-tracking for measuring color harmony, the inter-observer consistency is high.

Findings 3: This is confirmed by several employed metrics.

Hypothesis 4: There is an influence of color distribution and color diversity in the assessment of color harmony.

Findings 4: This is confirmed. The color distribution width does not seem to be the main factor, rather the spatial arrangement of color variety.

Ground truth

Hypothesis 5: The eye movements and individuals' assessment of color harmony are relevant enough to create a ground truth.

Findings 5: This is confirmed when using both information. Two aspects prevent from using directly the eye fixation maps recorded with a color harmony task. First, they are too noisy due to parasite fixations related to search task and low level attentional mechanisms. Second, the complexity of stimuli measured by their color distribution varies and lead us to clustered them according to observers' consistency in their assessment.

Global hypothesis: the concept of color harmony is well understood by observers, homogeneous between anyone and close to universality.

From our experiments, we concluded the concept is homogeneous enough to design any computational models that would predict in some extend a common human behavior and assessment.

Part III

Models and Applications

Chapter 11

Introduction

This chapter introduces the Part III, named *Models and Applications*. This part deals with two computational methods which exploit the findings of the Part II as well as two editing tools that implement the proposed methods. The part is composed of four chapters: a global Introduction (Section 11), the Saliency-guided Consistent Color Harmonization algorithm (Section 12), the Harmony-guided Quality Assessment algorithm (Section 13) and Color Harmony for Image Editing (Section 14).

In this Chapter, we explain our approach by depicting the Computational Perspective (Section 11.1), such as done for the experimental part. Then, we draw a picture of the existing algorithms (Section 11.2). Finally, we describe the formulation used along the three Chapters (Section 11.3).

11.1 The computational perspective

On one hand, a color science community conducted some work to characterize the color harmony concept. For example, the right association from two or three color patches controlled in terms of saturation, lightness has been annotated by naive experimenters. Out of these experiments, fitting curves have been extrapolated to model the relationship between hue, saturation, lightness and the degree of perceived harmony [195, 241]. From collected data during years, other color scientists designed a geometric representation of Color Harmony through hue wheel, saturation and lightness polygons [168, 252]. This community did not investigate a formulation or an algorithm for an automatic processing, even if some effort have been devoted to the integration of such rules into editing tools, but still with a manual intervention of an experienced user [106, 50, 161].

On the other hand, the image processing community found in the Cohen-Or *et al.* implementation [54] a first innovative recoloring algorithm which relies on fuzzy color science findings. This implementation opened the door to computational work on Color Harmony by providing a formulation of the Matsuda's harmony templates [168], i.e. the sector angles values and the histogram match-

ing between picture distribution and templates.

However, the experimental and computational work remain specific to each community, which did not exchange more than these experimental templates and mathematical formulation. Still the experimental community suffers from the generalization of the experimentally-designed models or representation (to natural pictures, large variety of colors, cultural specificity...) and the image processing community encounters the lack of available dataset, then the difficulty of an objective validation of their algorithm. Nonetheless, by means of large scale techniques, recent approaches have emerged which consist in gathering and inferring the opinion of a large community of designers on color patterns [194]. These work could fill in the gap between experimenters and algorithm scientists.

Having investigated the experimental and computational perspectives, we propose in this thesis to bring closer these two approaches. The experimental work serves the computational approaches described in this part. We did not perform any experiences involving any color science protocols, but rather focus on our experimental and computational expertise related to visual attention. As far as we know, the notion of color harmony has not been investigated under the view of visual attention.

Concretely, we propose in Chapter 12 to use a visual attention model into a color harmonization algorithm, inspired from Cohen-Or *et al.* work, in order to improve the choice of the best Color Harmony template. Other contributions are also introduced and their benefits are demonstrated individually.

In Chapter 13, a harmony-guided quality metric is designed. This is the first attempt to use the color harmony theory into a perceptual quality metric. Once again, we believe the perception of low-level features and the masking effects may influence the local and overall harmony perceived and interpreted by the human brain.

In Chapter 14, two editing tools have been designed. They rely with the two proposed computational methods about color harmony. These conclude and finalize this thesis by bringing the Color Harmony principles to the end user.

11.2 Color Harmony and Algorithms

A complete overview of existing methods involving the theory of color harmony is described in the literature review (Chapter 4): from models (Section 4.2) to applications in image processing (Section 4.3). Additionally, we also provide a review of editing tools in the Chapter related to this aspect (Section 14.1).

Briefly, there are three families of color harmony model:

- the **geometric models** which build on a wheel to represent the color spatial relationship and the associated harmony;
- the **numerical models** which fit curves on user annotations of pairs/triplets of controlled patterns;

- the **contingent model** which proposes a framework to integrates other modalities than only the stimuli properties, such as cultural heritage, age, mood, context, effect of time...

The numerical models remain largely unconsidered by the image processing community. Only the work of Solli and Lenz [239] did employ the model of Ou and Luo [195] to get a local measure of the color harmony. On the contrary, the geometric models, from Moon and Spencer's metric [172] to Matsuda templates [168], have been largely used in most of the image processing algorithms. The contingent models keep acting at a conceptual stage without any concrete implementation.

Following the trend in Image Processing, we keep working on the Matsuda's templates for both proposed methods. We reviewed partially the Cohen-Or formulation such as described in the next Section. In the context of an innovative perceptual quality metric, we propose a new framework to employ Matsuda's templates. In addition, we propose also two new editing tools which ease the user's manipulation of pictures. They offer a personal and homogeneous rendering of users' pictures.

As reported for the prior art, the two proposed algorithms may suffer from robustness due to lack of testing, even if we made an effort in this sense. Their fully automatic property as well as the uncertainty related to Matsuda's templates (originally built on design and fashion dataset) may be responsible for failure in specific cases. Consequently, we propose also the semi-automatic tools presented in Chapter 14 that could be a good compromise to encompass the subjectivity, cultural differences, taste and expertise of the user.

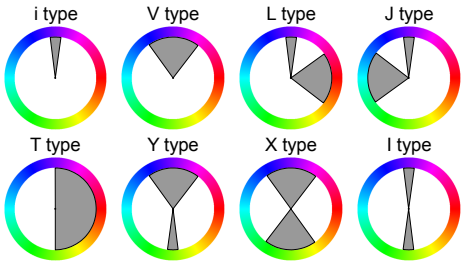


Figure 11.1: Effective color harmony templates used in our computational methods. Compared to Cohen-Or, we removed the O template and added the J one, symmetrical design of L.

11.3 Harmony Formulation

We are inspired from the formulation of Cohen-Or *et al.* in [54]. However, we try to improve its generalization by the introduction of sectors in the formulation, instead of using the template border as they did.

The eight harmonious templates $T_m, m \in \{i, I, L, T, V, X, Y, J\}$ have different sector layout and size to handle color complementary, color orthogonality

and color similarity (Figure 11.1). They can be turned around the hue wheel. Note that we added the template J as being the symmetrical version of L template.

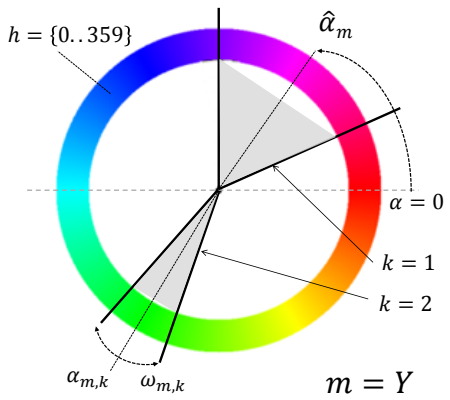


Figure 11.2: Color Harmony Formulation: Template notations. For ease of understanding, the notations have been implemented on a Template Y.

A template m is depicted in Figure 11.2 and defined as:

$$T_m : \{(\alpha_{m,k}, w_{m,k}); k = 1, ..., K_m\} \tag{11.1}$$

where $K_m \in \{1, 2\}$ is the number of sectors, $\alpha_{m,k}$ is the angle of the k-th sector's center on the hue wheel and $w_{m,k}$ its width in degrees. For notation simplicity, α_m denotes the angle of the first sector, which is also referred to as the template angle. For a given picture, an appropriate rotation angle $\hat{\alpha}_m$ is computed to align at best T_m with the hue distribution of the image. It is the template angle that minimizes the Kullback-Liebler divergence between the normalized hue distribution $M(h)$ of the picture, typically in the form of an histogram with $L = 360$ bins, and the normalized hue distribution of the template:

$$\hat{\alpha}_m = \arg\min_{\alpha} \sum_{h} M(h) \ln \left(\frac{M(h)}{P_m(h - \alpha)} \right), \tag{11.2}$$

where $P_m(h)$ is the hue distribution of the template T_m with angle 0, such as:

$$P_m(h) = \frac{1}{J} \sum_{k=1}^{K_m} P_{m,k}(h), \quad J = \sum_{h} \sum_{k=1}^{K_m} P_{m,k}(h), \tag{11.3}$$

$$P_{m,k}(h) = e^{\frac{-1}{1 - \left(\frac{2|h - \alpha_{m,k}|}{w_{m,k}} \right)^{10}}} 1_{\left\{ |h - \alpha_{m,k}| < \frac{w_{m,k}}{2} \right\}} \tag{11.4}$$

Each template's sector has a form of sigmoid such as expressed by the previous formula. Indeed, the distribution of a sector is characterized by a bump function,

under the form $P(x) = e^{\frac{-1}{1-(x)^t}}$ if $|x| < 1$. If we perform the following change of variables $x = \frac{2|h-\alpha_{m,k}|}{w_{m,k}}$, we reach the same formula. Note that t controls the curvature of the bump function. We empirically set it to 10 in order to get a *hat* border at the side of the distribution, as illustrated in Figure 11.3. This seems more appropriate and a good compromise compared to the use of a triangle or a square to represent a sector. The value $E_m = \sum_h M(h) \ln\left(\frac{M(h)}{P_m(h-\hat{\alpha}_m)}\right)$ represents the residual energy of template m for image at hand.

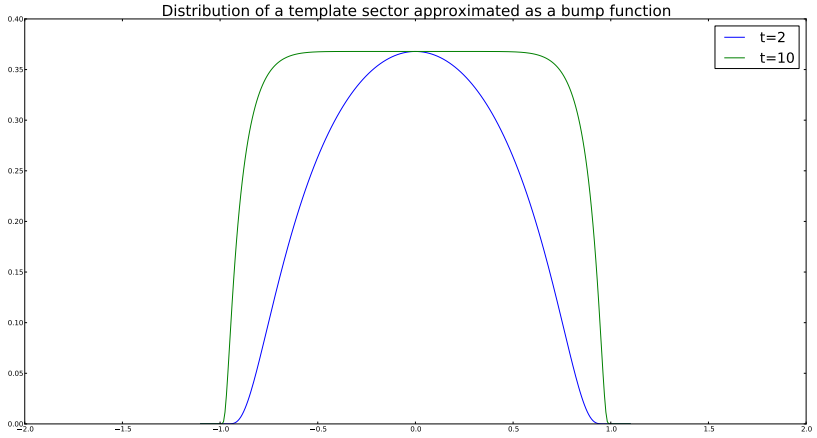


Figure 11.3: Color Harmony Formulation: Template distribution as a bump function. Two values of t are depicted.

Chapter 12

Saliency-guided Consistent Color Harmonization

Contribution: *Y. Baveye, F. Urban, C. Chamaret, V. Demoulin, P. Hellier, Saliency-Guided Consistent Color Harmonization, Computational Color Imaging Worshop, Lecture Notes in Computer Science, Springer p.105-118, 2012.*

In this chapter, we investigated a new computational method for Color Harmonization. Following the same intent as previous work (Section 4.3.1), the idea is to automatically recolorize the pixels that are estimated disharmonious in a picture. The obtained color palette of the processed picture is objectively more harmonious for human observers.

12.1 Introduction

The proposed automatic algorithm builds on the pioneering work of Cohen-Or [54] which takes advantage of Matsuda's color harmony model [168]. Thus it proposes the same algorithm architecture that tackles the main issues related to this processing: 1) through an energy minimization, the template that best matches the picture distribution is found, 2) color shifting is performed while preserving spatial consistency. However, we identified several limitations in the previous work that we carefully describe in Section 12.2: the choice of template is sometimes inadequate with regard to the picture distribution and the fidelity to the original content is not always preserved. Targeting also professional audience for such application, this last argument is somehow unacceptable. Colorists and professional designers hardly tolerate major alteration of their creative intent.

Having this in mind, we bring three contributions for such method: first, a saliency model is used to predict the most attractive visual areas and estimate a consistent harmonious template. Second, an efficient color segmentation algo-

rithm is proposed to perform consistent color mapping. Third, a new mapping function substitutes usual color shifting method. Results show that the method limits the visual artifacts of state-of-the-art methods and leads to a visually consistent harmonization.

Before describing the proposed method (Section 12.3), we first go deeper in the analysis of limitations of previous methods (Section 12.2). As a substantial contribution, several aspects of the validation are developed: from the gain introduced by each algorithm step (Section 12.4.1) to the global benefit of applying this harmonization method (Section 12.4.2).

12.2 Limitations of previous work

In addition to fixing the limitations of current state-of-the-art, we aim at providing a high fidelity of the original picture by changing as less as possible the colors of the original content. Hereby, we briefly provide a feeling about potential improvements for existing methods.

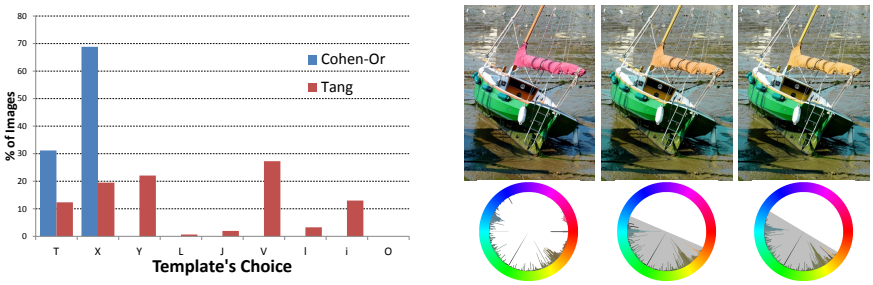
12.2.1 Template determination

We created a small dataset of 154 images with various content type. We compare the selection of templates for the method of Cohen-Or *et al.* [54] and Tang *et al.* [243] on this dataset. In [54] the cost function is the sum of the distance between each pixel hue and the template sectors, weighted by its saturation. In [243], the template is determined using the ratio between the number of pixels contained in the template and the maximum number of pixels that could be contained in the sectors. As expressed in Figure 12.1a, [54] tends to favor only the T and X templates, while [243] has a more balanced repartition of selected templates. Since Cohen-Or 's averages the contribution of each pixel, this is not surprising that larger templates gain the selection. However, such large templates do not subtly enclose the color picture distribution.

We also show the obtained template on a particular image on Figure 12.1b. Templates estimated using the two methods are shown at the bottom, and the harmonized images at the top. [54] and [243] have estimated the T template, leading to highly changes in the final color rendering. The distribution is pretty large around the hue wheel and the T template favors the hues that are similar within the grey sector to the detriment of the pink color. A X template could have been a better solution such as illustrated in Figure 12.6a. The choice of a suitable energy remains the challenge for fixing such issue.

12.2.2 Alteration of original content

The loss of meaningfully minor colors in the global distribution is a challenge for automatic color harmonization. Modifying skin or sky colors in an unnatural way may have an annoying impact on the image semantic. A manually defined



(a) (b)

Figure 12.1: Limitations of current methods: the issue about template determination. (a) Distribution of template selection for a database of 154 images regarding the Cohen-Or [54] and Tang [243] methods. (b) Illustration of a large template choice and the resulting color mapping (From left to right: Original pictures, Cohen-Or and Tang implementations).

mask is usually used to harmonize only selected pixel and thus preserve other areas [54]. Salient areas are visually attractive and then important in a picture. However, they have by definition a low representation in the color distribution. Previous methods did not highlight such problem illustrated in Figure 12.2a.

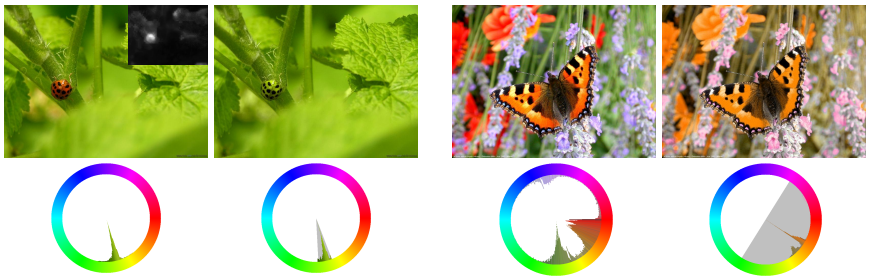
Another source of alteration of the original content is the color mapping or shifting which is applied after finding the candidate template. In state-of-the-art solutions, color mapping functions tend to highly contract hue even for colors that originally lie in the middle of harmonious sectors. Thus, the harmonization might over-modify original pictures. This effect may be sought in some applications, but may be disruptive when the intent is to stick to the original image. Color dynamics are not preserved even for the pixels whose hue is near the middle of the template in the original image (Figure 12.2b).

12.2.3 Spatial inconsistency

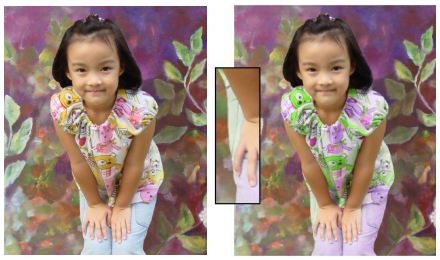
The need for color segmentation has been discussed previously in [54]. Some illustrations are provided in Figure 4.8b. However, depending on the color segmentation approach, a higher level of inconsistency may still be encountered. Using a spatial segmentation might lead to split a disconnected object into two different components such as illustrated in Figure 12.2c. Using a non spatial algorithm allows to treat all pixels having the same colors without a priori on their position.

12.3 Our Approach

Having identified the previous limitations, we tried to bring a solution while still performing a minimal recoloring of the picture. A better determination of template is achieved by using the Kullback-Leibler divergence as an energy for distribution matching. Indeed, it is a non-symmetric measure of the difference between two probability distributions P and Q. Specifically, the Kull-



(a) (b)



(c) from Coehn-Or [54]

Figure 12.2: Limitations of current methods: alteration of the original content and spatial unconsistency. (a) Lost of colors for small and salient areas. (b) Color mapping function is too "strong". (c) Object color inconsistency computed from Cohen-Or [54]

back–Leibler divergence of Q from P is a measure of the information lost when Q is used to approximate P. This seems fitted better the problem of matching templates and picture distribution.

We applied this measure on a weighted histogram of the picture that take into consideration the most salient areas of the pictures. Thus, more weight is applied to the colors of these pixels allowing a bias (in favor of these salient colors) in the template selection. To determine the salient areas, a saliency map is computed from a visual attention model. In order to avoid too much changes or alteration of the original colors, we also introduce a new color shifting function which softly tightens the original color distribution around the center of the harmonious sectors.

Regarding the issue of spatial consistency, we prefered a non-spatial color segmentation algorithm that clusters the image on an histogram basis without spatial a priori. An overview of the method is shown in Figure 12.3. We then describe the different steps of the algorithm.

12.3.1 Template determination

This step consists in choosing the best template amongst one of the nine harmonious templates T_m ($m \in \{i, I, L, T, V, X, Y, J, O\}$). The first eight templates are depicted in Figure 11.1. Note that template O has been added since we

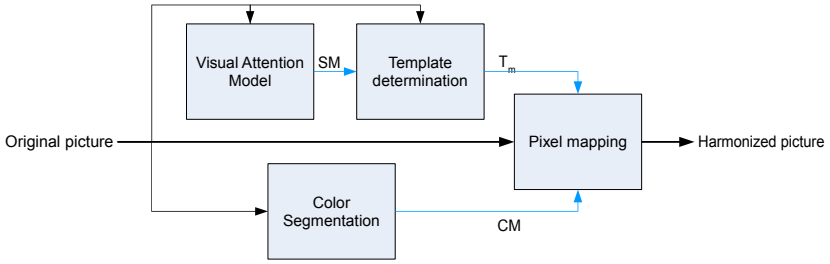


Figure 12.3: Overview of the proposed Color Harmonization algorithm. The first step (top part) consists in estimating the best harmonic template. Based on a measure of the salient areas, a template determination is performed so as to minimize a statistical distance between the histogram and a harmonious template. Prior to the color mapping, a color segmentation technique is performed to ensure that consistent colors are finally mapped into the same harmonious template sector.

want to leave unchanged (i.e., no harmonization) the pictures containing a large distribution of hue values, e.g. rainbow-like pictures. We followed the same formulation as the previous computational method (Section 11.3).

The color histogram of the original image is computed in HSV space using L bins (typically, $L = 360$). However, to determine the template matching best the color distribution, a modified hue histogram M is weighted by the saturation and value. At each pixel $\mathbf{u} = (x, y)$ with associated hue $h(\mathbf{u})$:

$$\forall i \in [1, L], M_i = \frac{1}{\sum_{\mathbf{u}} S(\mathbf{u}) \times V(\mathbf{u})} \times \sum_{\mathbf{u}|h(u)=i} S(\mathbf{u}) \times V(\mathbf{u}) \tag{12.1}$$

Saliency map

The saliency map SM is computed for the considered picture from the visual attention model of [170]. The saliency map provides a representation of the most visually attractive pixels (white pixels stand for the most attractive one.). The basic idea is that the estimation of the template should be driven mainly by visually representative areas of the image to obtain a visually consistent harmonization. As for the entire image, a weighted hue histogram S is computed for the T most salient pixels, such as done in Equation (12.1). The threshold T has been empirically set to 1% of the total number of pixels in all experiments.

Cost function or energy

A cost function is required to determine the most appropriate or the closest template for the distributions M (and S). The appropriate template shape T_m and the associated orientation α that best fits the hue distributions M is chosen by minimizing the Kullback-Leibler divergence computed for each template and each orientation, such already expressed in Section 11.3:

$$\hat{\alpha}_m = \arg\min_{\alpha} \sum_{h} M(h) \ln \left(\frac{M(h)}{P_m(h - \alpha)} \right), \tag{12.2}$$

where $P_m(h)$ is the hue distribution of the template T_m with angle 0, already specified and justified in Section 11.3.

The value $E_m = \sum_h M(h) \ln\left(\frac{M(h)}{P_m(h-\hat{\alpha}_m)}\right)$ represents the residual energy of template m for image at hand.

The cost function is applied independently on both M and S distributions, leading to two optimal templates and their associated angles $\hat{\alpha}_{m,M}$ and $\hat{\alpha}_{m,S}$, following these two steps minimization:

$$T^* = \arg\min_m \left\{ \arg\min_{\alpha_m} \sum_h M(h) \ln\left(\frac{M(h)}{P_m(h-\alpha_m)}\right) \right\}. \tag{12.3}$$

Note that $\hat{\alpha}_{m,M}$ is equivalent to (12.2) and that:

$$\hat{\alpha}_{m,S} = \arg\min_\alpha \sum_h S(h) \ln\left(\frac{S(h)}{P_m(h-\alpha)}\right), \tag{12.4}$$

Merging the two templates

Both templates are then combined to form a new distribution or model P'. The combination consists in taking the maximum value between the optimal templates of $P_{m,M}(h-\hat{\alpha}_{m,M})$ and $P_{m,S}(h-\hat{\alpha}_{m,S})$, relative to the whole image and to the salient pixels:

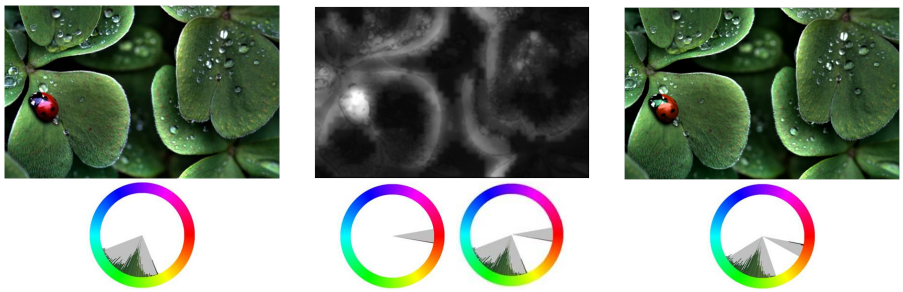
$$\forall h \in [0 \dots 359], P'(h) = \max\left(P_{m,M}(h-\hat{\alpha}_{m,M}), P_{m,S}(h-\hat{\alpha}_{m,S})\right) \tag{12.5}$$

Once again, the most similar template to this new combined distribution P', among the nine harmonious templates, is found by minimizing the Kullback-Leibler divergence under α and the residual energy under the different templates as expressed in equation (12.3).

The different steps are illustrated in Figure 12.4. Since minor hues in color histogram have to be taken into account due to their semantic or attentional relevance, two templates based on two different hue distributions (the entire image and only most salient pixels) are separately computed and then combined. The main advantage of using two hue distributions is that the method is parameter-free, compared to a weighting of salient pixels where the weights would have to be tuned carefully.

12.3.2 Pixel color mapping

Once an harmonious template has been estimated, colors that lie outside the template will be mapped into the template. This step is described here and the main purpose is to perform a larger modification for pixels whose color is outside the harmonious template and a minor (or unexisting) change for those inside the template. Some artifacts may be created during this step because two neighboring pixels that have similar colors can be mapped into two different sectors of a template. We discussed this problem earlier in Figure 12.2c. To



(a) (b) (c)

Figure 12.4: Illustration of the use of saliency for finding the final template. (a) Original picture and the best template computed on its color distribution M, (b) Saliency map, the best template corresponding to its 1% highest values S (left template) and the unified template P' (right template) (c) the final template (from Matsuda's set), the closest one to P'.

remove these artifacts, all pixels that are in the same segmented area are mapped in the same sector. To do so, a dedicated color segmentation will be used and is described in the next section.

A hue-mapping map is thus created where each pixel is assigned with the direction of mapping of its hue value:

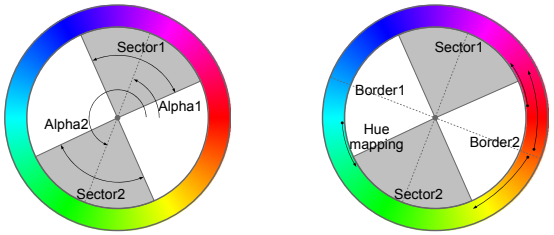
- For each pixel p, its original hue quadrant between a sector angle alpha and a border is determined,
- each segmented area is sub-divided depending on the pixel hue where all pixels having their hues between a sector angle and the next are in the same sub-segment (this is because segmentation problems appear only at sector borders - each segment can be divided in two parts (see Figure 12.5),
- for each sub-segment, the majority quadrant (between the 2 around a given border) is selected as the hue mapping direction for each pixel p, so that all pixels belonging to the same segment and having their hues around a sector border are harmonized in the same direction to avoid mapping artifacts.

A sigmoid function is then used to transform the hue of each pixel at site $\mathbf{u} = (x, y)$:

$$h^{'}(\mathbf{u}) = \alpha_{m,k} + Sign \times \frac{w_{m,k}}{2} \times \tanh\left(\frac{2\times \parallel H(\mathbf{u}) - \alpha_{m,k} \parallel}{w_{m,k}}\right) \qquad (12.6)$$

where $C(\mathbf{u})$ is the central hue of the sector associated with $\mathbf{u}$, w is the arc-width of the template sector, $\parallel . \parallel$ refers again to the arc-length distance on the hue wheel and $Sign$ is the sign associated to the direction of mapping.

This sigmoid function has interesting properties for pixel color mapping: its asymptote in extreme values auto-clamps pixels in the template and its middle



(a) (b)

Figure 12.5: Harmonic mapping of hue depending on the initial hue value. (a) Chosen harmonic template with two sectors and their angles, (b) hue mapping examples

section (normal behavior) is nearly linear so at the center of a sector, hues are not changed. Original image and dynamics are less modified compared to Cohen-Or [54] where the mapping function contracts hues even near the center of sectors. Since dominant colors are likely to be chosen near the middle section of harmonic sectors, they are less subject to hue change, thus the original color intent is maintained. Therefore automatic harmonization can be performed without any user interaction.

12.3.3 Color segmentation

As mentioned previously, a color segmentation step is necessary to ensure the mapping consistency, in other words to maximize the probability that pixels belonging to the same semantic object are mapped to the same harmonious sector.

State-of-the-art techniques [54] and [108] apply conventional graph cut algorithm [32] for segmentation, while Tang *et al.* [244] perform a two-steps graph cut processing both at the region and pixel levels in order to handle issues regarding spatial inconsistency of graph-cut algorithm. Sawant *et al.* [229] do not use any image segmentation but implements histogram splitting on the hue histogram. Such a method does not account for spatial arrangement of the pixels and can be seen as color clusterization of the outlier pixels in two clusters.

For color harmonization, the spatial aspect of the color segmentation may not be compulsory and might even introduce artifacts as it can be seen on Figure 12.2c. Therefore, we think that a histogram segmentation technique is adequate here, such as the popular K-means method. However, such a histogram segmentation should obey the following constraints:

- It should be unsupervised, meaning that the final number of color clusters should not be a parameter. As a matter of fact, the color harmonization would be very sensitive to a incorrect number of meaningful colors.
- The histogram segmentation technique should be capable of segmenting *small* modes of the histogram. In other words, small regions that could be seen as color outliers, such as the ladybug of Figure 12.2a, should be detected as separate modes.

In order to meet these requirements, we propose here a color segmentation method that build on the work of Delon *et al.* [65] referred to as ACoPa (Automatic Color Palette). This color segmentation technique is based on *a contrario* analysis of the color histogram modes. A statistical estimation of meaningful histogram modes is performed. Instead of the hierarchical estimation of modes in the H, then S, then V space, we propose to perform a histogram decomposition of each component independently. The obtained modes are combined from all modes obtained, and segments with a very limited group of pixels are discarded. Finally, based on these histograms modes, a K-means post-processing is used to group the modes that are perceptually similar using a dictionary expressed in the Lab color space.
We have obtained a segmentation technique that is approximatively 10 times faster than the original version, and deals more efficiently with achromatic pixels.

12.4 Validation

In this section, we attempt providing a large qualitative appreciation of the method compared mainly to Cohen-Or method. In a second part, we have perform an user experiment to validate the method. The basic idea is to demonstrate that observers prefer or better rate harmonized pictures compared to non-harmonized pictures. We set up a pair-wise protocol where observers were asked for choosing the most harmonious pictures between a pair. Pictures may be completely different (not necessary the pair: original versus its harmonized counterpart) and are reorder at the end on a global scale, from the least to the highest harmony rating.

First, we will demonstrate that we have fixed the state-of-the-art limitations mentioned in Section 12.2, then we will provide a bench of visual results and finally we introduce the experiment and its results.

12.4.1 Fixing state-of-the-art limitations

In Section 12.2, we introduced the limitations of previous work and developed our approach for avoiding the mentioned effects. Below is a bench of results related to the different issues.

Searching for a better fidelity

Figure 12.6 provides the results of our approach for all previously illustrated limitations. We outperform Cohen-Or method for all of them. A better fidelity is reached with our approach. In Figure 12.6a, the choice of X template better fit the original color distribution. While in Figure 12.6b we keep the color of the ladybug by means of saliency component, in Figure 12.6c the soft color mapping within the template allows keeping most of the original colors.

(a) see Figure 12.1b (b) see Figure 12.2a (c) see Figure 12.2b

Figure 12.6: Results of our method for the limitations of the state-of-the-art. For comparison, the reader can refer to the mentioned figures in caption.

Figure 12.7 illustrates the interest of having changed the energy and used the KL divergence. The choice of final template is different and lead to drastic differences in final color remapping. In 12.7a, Cohen-Or method drastically change the original color intent. With our method, we arrange only the color of the tee-shirt, shifted to blue color, to fit well with shoes color. The example in Figure 12.7b is clearly tricky. This is a typical case where the nature meets man-made colors; this is difficult to harmonize and Matsuda's template potentially does not handle this case. Our method slightly changes the rendering, while Cohen-Or completely alterates the original scene.

Figure 12.8 illustrates the interest of using the saliency map within the selection of template. The choice of final template is different and leads to drastical differences in final color remapping. Even if we do not use the saliency (third case), our method maintains the background color while Cohen-Or contracts too much the blue within the template.

Figure 12.9 illustrates the interest of having a soft color mapping function. Most of original color distribution is maintained. The other colors are softly mapped into the template's borders. The skin is preserved for our method and gets yellowish due to the high contraction of Cohen-Or method. Also for the second picture, the spirit around blue colors is maintained in our result, while not for Cohen-Or.

No Spatial artefacts

By means of our histogram-based color segmentation, we fixed the problem of unconsistency for non-convex areas having the same color. Since we do the segmentation on the histogram, we do not have any a priori on the spatial consistency which leads to better global results, such as illustrated with Figure 12.10. The trouser has a homogeneous color for our method.

(a)

(b)

Figure 12.7: Results of our method to illustrate the interest of our energy formulation.

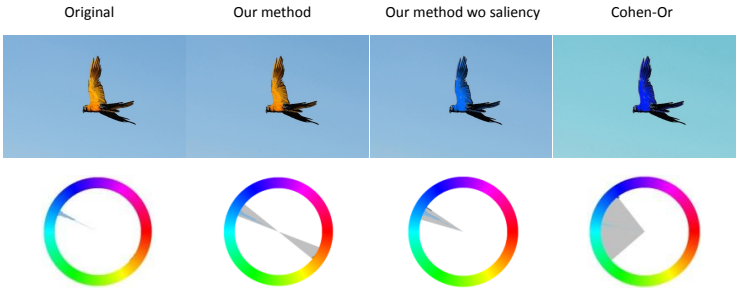


Figure 12.8: Results of our method to illustrate the interest of the saliency map.

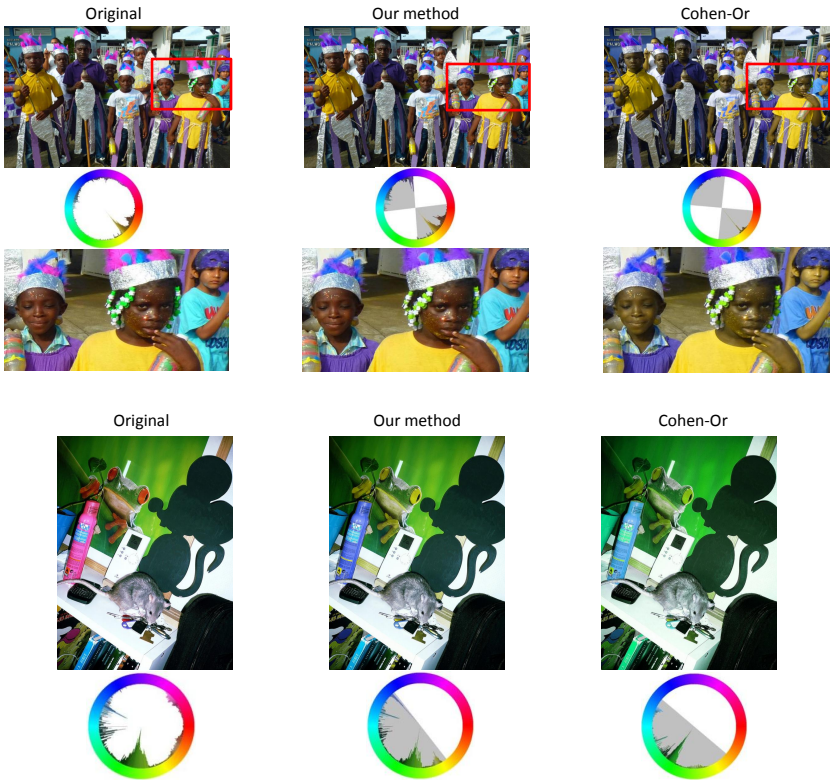


Figure 12.9: Results of our method to illustrate the soft color mapping function.

(a) (b)

Figure 12.10: Results of our method to illustrate the non-spatial segmentation. (a) The original picture and the harmonized picture with spatial unconsistency of object from [54], (b) our harmonized result without artefact.

12.4.2 User test

This section relates an experiment of pair-wise comparison where we asked for the "most harmonious pictures" to be picked up during a side-by-side presentation. This experiment leads us a bit further in the understanding of color harmony. We answered the following questions: Knowing there is no previous work on the pair-wise assessment of harmony for color pictures, is there any high agreement between observers? Can we reasonably classified pictures with different color content on the same scale from the least to the most harmonious pictures? Have the large color distribution pictures a particular behavior? Is the template theory from Matsuda and consequently the derived color processing methods reliable for automatic harmonization task?

Participants

Twenty-one employees from 20 to 53 years old participated to this experiment, with a median age of forty-two years old. The repartition between male and females is unbalanced (four females and seventeen males). The enrollment for participation was organized as a challenge to find the "golden eye" of color harmony.

Material

We conducted the experiment on a 22-inch DELL monitor having a resolution of 1900 x 1200 pixels. Participants were located around 60 cm far from the screen. The experiment setup was located in a dedicated user test room free of noise with dark wall. The pair-wise comparison was implemented on a python server that manages the experiment itself as well as the results visualization after each run. It allows the monitoring of ranking after each participant. Moreover, it has a console for the participant to enter its own characteristics (name, age, expertise, gender, affiliation) and a training page to get familiar to the side by side assessment.

Stimuli

Over the thirty-two stimuli composing the dataset of the Experiment 2 (Chapter 8) depicted in Figure 8.1, twenty-three of them and their harmonized counterpart computed from our color harmonization method [24] were proposed to the participant. In addition, three "large distribution" or complex stimuli of the original dataset have been dropped into the involved stimuli but without their harmonized counterpart. A total of forty-nine stimuli were involved in the experiment.

Procedure

Participants were asked first to fill in the form about their personal information. After performing the training session, they started the pair-wise experiment. It

consists in choosing "the most harmonious pictures" over the side-by-side display of two stimuli extracted from the forty-nine involved stimuli. They had eight seconds of visualization time to make their choice. After that, the pictures disappear and they had an unlimited time to finally choose between one over the two pictures. The choice was forced and could be made also during the first eight seconds of visualization.

The paired pictures presented to the observers were not set randomly. The specification document from ITU-T [114] for subjective quality assessment recommends presenting all possible pairs even the same pairs being unversed on screen in order to derive accurate final ranking of all n involved pictures. Following such constraint leads to a number of pair comparison equal to $n \cdot (n-1)$. Having an average visualization time of at least six seconds per pair (4 seconds of visualization and 2 seconds for choice) and forty-nine pictures (2352 pairs) causes a total experiment time of 235 minutes (about four hours). This is obviously unrealistic. Twenty minutes of experiment with previously mentioned conditions leads to using no more than 25 stimuli. Since we already reduced the number of involved original pictures (only twenty-three over the thirty-two original stimuli) to also include their harmonized counterpart, we have focused on an optimal solution to reduce the number of presented pairs while guaranteeing a good accuracy of the final picture ranking.

The Bradley-Terry model [34] is usually the approach recommended for subjective quality assessment. This linear model analyses participants' choices during pair comparisons in order to map the probabilities of preferences to scales (see probability results for each ranked picture in Figure 12.12). Our proposed experiment used the Matlab implementation from Wickelmaier and Schmid [269]. Li *et al.* [150] proposed an adaptive square design method that decreases the number of pairs comparison to $n \cdot (\sqrt{n}-1)$, leading in our case to 294 pairs and about 30 minutes. The good convergence for ranking is controlled for each iteration(after each participant) and the algorithm is initialized with the previously computed ranking scale such as performed by Li *et al.* [150]. Thus, the convergence of ranking has been reached after recording around fifteen participants. However, we conducted twenty-one iterations to ensure the convergence.

Inter-observer agreement

We first want to answer the question related to the reliability of pair-wise comparison for color harmony task. Thus, we need to measure the global inter-observer agreement. Regarding the pair-wise aspect combined with a dynamic creation of pairs (that may be different for each observer), we needed an agreement metric that could handle ordinal annotation and a number of annotation that may vary between pairs and finally for each stimulus.

Several measures of inter-annotator agreement are used in the literature such as percent agreement, Fleiss' kappa [79] and Krippendorff's alpha [133]. Percent

agreement is widely used and intuitive but overestimates inter-annotator reliability since it does not take into account the agreement expected by chance. However, this parameter is even presented as a high boundary in Table 12.1. Fleiss' kappa and Krippendorff's alpha both take into account observed disagreement and expected disagreement but suffers from prevalence: they consider that annotators know a priori the quantity of cases that should be distributed into each category [216]. It results that, especially using binary response which is the case here, if a value is very rare the reliability is low even if there are few mistakes in the annotations. Randolph's multirater kappa free [216] is not subject to prevalence because it does not depends on how many values are in each category.

In our annotation process, prevalence is not an issue. All possible pairs are not presented to the observers; a stimulus is presented at most four times over the 294 total numbers of pairs. Moreover, the most harmonious stimulus (regarding the ranking of the previous iteration) is presented either on left or right side of the screen. We do not induce an unbalanced effect on the binary rating related to the stimuli presence or position. Both kappa values need a fixed number of annotators; we used four annotations to compute all these measures, because it is the best compromise between the number of discarded pairs and a high value of annotators. The inter-annotator reliabilities for these sub-samples are displayed in Table 12.1. Their values can range from 0 to 1 for percent agreement and from -1 to 1 for the other measures. For Fleiss' kappa, Krippendorff's alpha and Randolph's kappa, a value below 0 indicates that disagreements are systematic and exceed what can be expected by chance, a value equal to 0 indicates the absence of reliability and a value equal to 1 indicates a perfect agreement between annotators.

Agreement Metric	Complete Dataset	Original Pictures	Harmonized Pictures	Orig. vs Harm. pictures
Percent agreement	0.73	0.72	0.76	0.74
Fleiss' κ	0.12	0.13	0.19	0.16
Krippendorff's α	0.12	0.07	0.17	0.14
Randolph's κ	0.55	0.57	0.62	0.58

Table 12.1: Comparison of inter-observer agreement metrics for all stimuli (second column) and several subsets: only pairs involving original pictures (third column), harmonized pictures (fourth column) and an original and its harmonized counterpart (fifth column)

In Table 12.1, all values are positive which means that agreement is slightly better than what would have been expected by chance and have similar range of values to other subjective tasks, such as emotion annotation studies [139]. Randolph's kappa which is robust against prevalence gives the highest reliability value compared to Fleiss' kappa and Krippendorff's alpha. Landis and Koch [139] suggest that a score of 0.375 indicates a fair agreement and a score of 0.452 corresponds to a moderate agreement. Since the experiment achieves a Randolph's kappa of 0.55 on the complete dataset, we conclude that the task reaches a high rating agreement despite its subjectivity.

Since Krippendorff's alpha takes into account all pairs even if they are not rated with the same number of annotators. As explained before, the prevalence notion is not an issue for our experiment then the Randolph's kappa has no interest in such context. Consequently, we will discuss only the results related to Krippendorff's alpha in the following section.

12.4.3 Ranking Results

Having said inter-observer agreement is high enough, we focus now on the two questions mentioned at the beginning of this section: 1) the classification of different stimuli on a scale and 2) the validity of Matsuda's templates through the processing tool of pictures harmonization. There are two distinct studies: the ranking of different stimuli categories (original versus harmonized sets) and the ranking of each stimulus versus its harmonized counterpart, mentioned as the inter-stimuli and the intra-stimuli conditions, respectively. Thus, we created subset of annotated pairs by gathering all concerned pairs for the observed condition.

Inter-stimuli condition

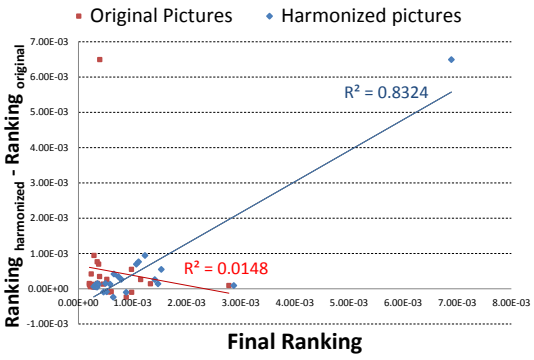
Focusing on inter-observer agreement, Table 12.1 illustrates difference of performances regarding the set of observed pictures. While original pictures reaches the lowest agreement (Krippendorff' $\alpha = 0.07$), the set of harmonized pictures have a higher agreement score (Krippendorff' $\alpha = 0.17$). Figure 12.11a introduces the correlation between final rankings of a category as a function of ranking differences for each stimulus. Clearly, the harmonized ranking is highly correlated to the difference, meaning the higher the ranking, the higher the difference is. However, this is not the case for the original pictures. Their ranking is not correlated to the difference of ranking ($R^2 = 0.0148$, Figure 12.11a) which is more difficult to interpret.

Figure 12.12 points out qualitatively that harmonized pictures are rather at the top of harmony scale and original pictures at the bottom of it. Previous numerical proof seems to make the evidence that classifying original pictures on a harmony scale is a difficult task, while comparing harmonized pictures is more reliable and easier for observers.

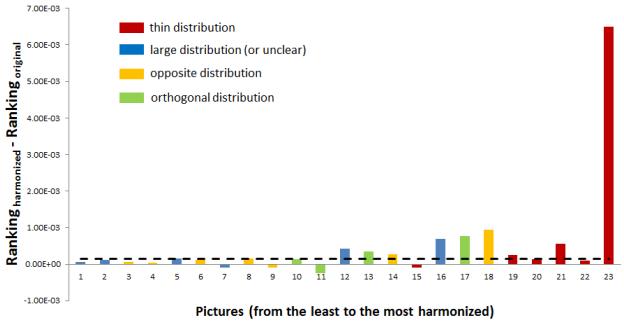
Intra-stimuli condition

Working with the harmonized aspect of the experiment, over the twenty-three pairs (original and harmonized counterpart), only four stimuli had a ranking where the harmonized picture has a lower ranking (blue squares in Figure 12.12). Our color harmonization method and the color harmony model [168] associated to it are valid. As observed in Figure 12.12, those particular four stimuli are located in the middle of the ranking scale indicating potential neutral aspect of

the harmony and thus difficulty for harmony assessment. Additionally one can notice that these differences are quite low in Figure 12.11b (negative bars).



(a)



(b)

Figure 12.11: Ranking differences and statistics. (a) Linear correlation between Ranking differences and final ranking for each set: original versus harmonized pictures. (b) Ranking differences between original pictures and their harmonized counterpart as a function of harmonized pictures order.

12.4.4 Color Distribution Role

Since we introduced complex stimuli with large color distribution or several hue peaks in their distribution (Figure 8.1), we are interested in any potential correlation between the hue histogram or number of hue values larger than a minimal threshold ($T = 500$) and the final ranking. We found a correlation between those parameters ($R^2 = 0.26$ with a linear model and $R^2 = 0.48$ with a logarithmic model), that confirms the intuitive assumption: the larger the color distribution, the lower the rank.

In the same vein, the correlation between inter-observer agreement and color distribution does exist at different degree according to the considered sets ($R^2_{all} = 0.16$, $R^2_{original} = 0.07$, $R^2_{harmonized} = 0.20$). It is more important for the harmonized dataset.

Focusing on the ranking in Figure 12.11b and color distribution of the dataset depicted in Figure 8.1, we can observe that large, unclear and opposite color distributions are located in the first third of the low harmony rate. Thin color distributions (the two first rows of Figure 8.1) are rather at the top of the harmony scale. Orthogonal distributions are rather in the middle of the scale.

We can formulate two conclusions out of these observations. First, thin distributions are probably easier to analyze by observers and also quite intrinsically harmonious due to their arrangement with similar hues. Large distributions are easy also to process for harmony task, but rather lead to disharmony opinion. Orthogonal distributions are located in the middle of the scale and depict either a high ambiguity for assessment or a neutral harmony score. However, there are some outliers to these conclusions (pictures 16, 17 and 18 in Figure 12.11b). When focusing on differences of ranking, it appears that these harmonized pictures are quite far on the scale from their original counterpart (ranking difference above the average). The harmonization processing of such stimuli highly improves their harmony feeling as annotated by observers.

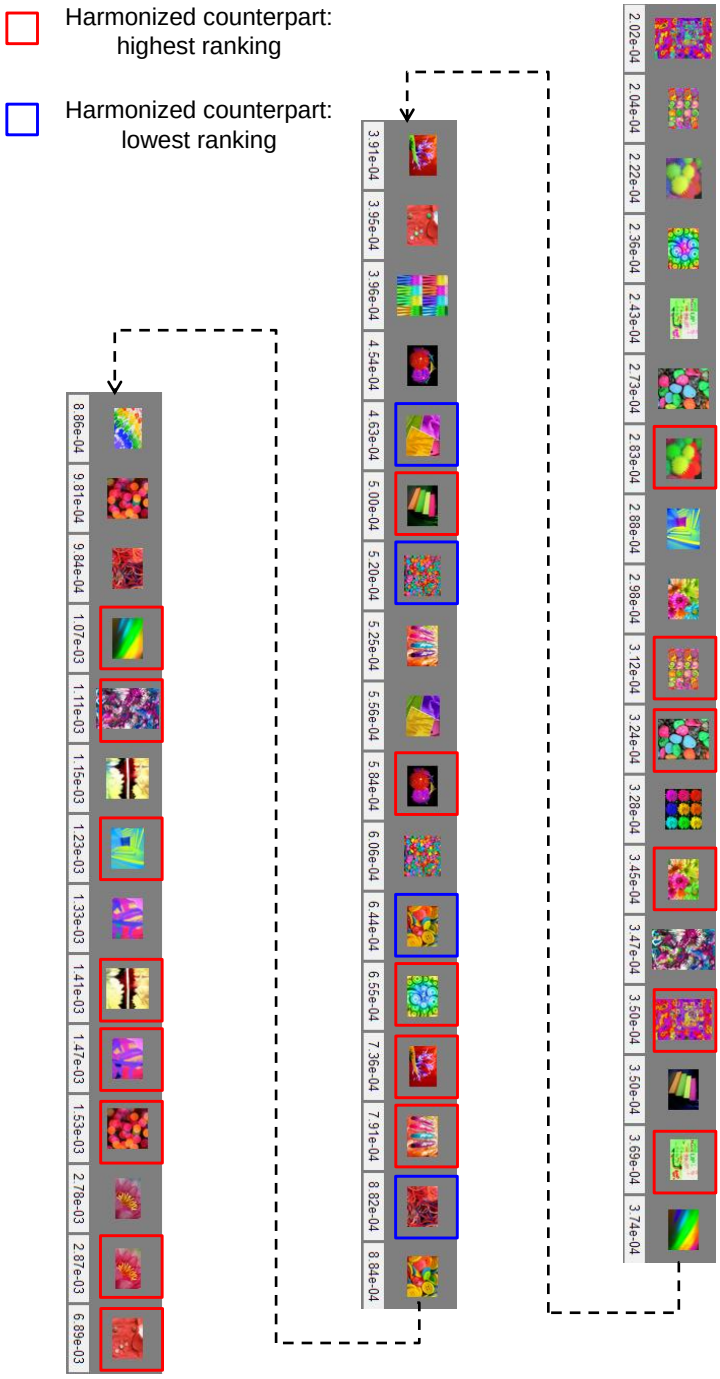


Figure 12.12: Qualitative observation of the harmony ranking for all stimuli. First row is the lowest ranking for harmony score and so on. Red squares points out the harmonized counterpart when this one is ranked above the original one.

12.4.5 Discussion

Fidelity criteria

Regarding the computational method exposed in this chapter, we can discuss one limitation of having used the fidelity criteria. As originally motivated, the aim of preserving original content is dictated by professional and potential audience of such algorithm that would argue for the respect of their creative intent. However, this fidelity may lead to few changes in the harmonization process. Thus, having more changes and/or contraction of the color distribution may sometimes bring more pleasant results because more consistency in terms of colors.

Figure 12.13: Discussion on fidelity criteria. (a) Original picture and its associated hue distribution, (b) the harmonized result from our method (only the green part of cushion is harmonized), (c) Cohen-Or harmonization with our shifting function (less contraction), (d) Cohen-Or harmonization with his original shifting function (highly contracted).

One example is illustrated in Figure 12.13. The first row depicts from left

to right the original picture and our harmonized result. Few changes have been performed, only the green pattern of cushion has been recolored in blue. Two versions of Cohen-Or's algorithm are depicted on the second row: the left-hand side version uses Cohen-Or's cost function and our shifting function, while the right-hand side version implement both Cohen-Or's cost function and shifting function (high contraction). One can appreciate the harmonious rendering of Cohen-Or implementation, while our result preserve maybe too much the original intent. The different choice of templates for our method has clearly conducted to a better preservation of blue color, itself driven by the saliency feature that we have introduced in our method. This is a typical case, where the saliency does not help much. However, we can always appreciate the interest of using a slight shifting function for the contraction of colors within the template, as depicted in Figure 12.13c.

Assessment methods

A tricky point about such image processing method is indisputably the validation aspect. There are neither objective result, nor ground truth to be compared to. Quantitative evaluation is somehow impossible to provide. A visual appreciation remains usually the only way to provide to the community a benchmark about the performances of the different methods.

Following an usual validation process, we provided qualitative results of our method, but also we carefully proposed evidence of improvements for each identified limitations of current state-of-the-art (Section 12.4.1). It seems proven that our algorithm fix all limitations by applying suitable strategies on the identified components of the algorithm.

As a new contribution we proposed to set up an user experiment where participants directly assessed the most harmonious pictures from a presented pair. Figure 12.12 attests of the right classification on a global scale (from low to high harmony rank), where harmonized pictures locate mostly at a higher rank than their original counterpart. This experiment expresses the improvements performed by the computational method. However, we can discuss the nature of stimuli. All of them received a substantial level of harmonization, despite the criteria on fidelity. In other words, many pixels have been recolored to get the picture globally consistent. We believe the ranking would have been completely different and maybe not in favor of harmonized stimuli, if few pixels would have required a recolorization (such as the picture in Figure 12.13a).

12.5 Summary

In this Chapter, a color harmonization method has been presented. Building on the work of Cohen-Or [54], the method proposes three main contributions to reduce possible artefacts and obtain a visually consistent result. First, a cost function measuring the statistical distance between a weighted hue histogram and the possible templates was used, jointly with the information of saliency.

In Cohen-Or *et al.* [54] the cost penalizes only pixels that have a hue outside the template, and favors large templates. In Tang *et al.*[243], the cost penalizes large sectors when not necessary. Our approach considers hue histogram and templates as statistical distributions. The Kullback-Liebler divergence is used to quantify the distribution similarity and emphasizes histogram differences. In addition, the weighted hue histogram is closer to the perception of colors. Consequently, the selection of templates better fits the original image color distribution. Second, a new color mapping function was used. Results have shown that this mapping function contracts less the mapping within the desired template and leads to a more consistent mapping. Third, a dedicated color segmentation was used and was able to segment small histogram modes, while leading to satisfactory result.

In Section 12.4, we attempted to convince that the method is achieving good performances by tackle different tracks for validation. We provided usual visual results, but also we carefully answered the questions raised during the exhaustive exposition of limitations (Section 12.2). In addition to the computational method of color harmonization, a major contribution deals with the introduction of a pairwise protocol in the context of color harmonization. We aimed at showing that harmonized pictures are objectively rated more harmonious than their original counterpart. Indeed, 19 over 23 stimuli achieved this expected conclusion.

Chapter 13

Harmony-guided Quality Assessment

Contribution: *C. Chamaret, F. Urban, No-reference Harmony-Guided Quality Assessment, Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on , vol., no., pp.961,967, 23-28 June 2013*

In Section 4.3, the different employments of color harmony rules are reviewed. From automatic color harmonization to the use as a low-level feature, the field of application in image processing is large. The models of color harmony serve to derive image aesthetic scores, infer emotion or to re-colorize pictures. However, no work attempts to explore the field of quality assessment (at a larger extend, the visual perception) with a color harmony perspective.

Extensive research has been done in the context of quality assessment to define what is visible or not in images and videos. Techniques based on human visual system models use signal masking to define visibility thresholds in order to rate the impact of dedicated artifacts. What can the human eye perceive disharmonious?

Based on the theory in both fields, we have designed a quality assessment method which measures what is harmonious or not in an image. Color harmony rules from Matsuda's color coordination are used to detect which parts of images are disharmonious, and visual masking is applied to estimate to what extent an image area can be perceived disharmonious.

As an interesting tool for content creator and targeting a maximization of the artistic effect, the proposed computational method outputs a no-reference perceptual harmony-guided quality map as well as a score of disharmony. These outputs are integrated into a photo editing framework to guide the user getting the best artistic effects (Chapter 14).

13.1 Introduction

When manipulating, editing, improving images, the best quality as well as a certain artistic intent are usually the purpose. Nevertheless, although the issues related to objective quality assessment have been largely studied in the context of low level artifacts (blur, blockiness, jitter...), the artistic intent remains difficult to model or learn. As an intermediary indicator, aesthetic quality metrics based on high-level features intuitively related to beauty (colorfulness, line orientation, shape...) and rules of thumb (composition, rules-of-third, skyline...) are showing up recently in the community [174, 63, 274]. Depending on the context, some approaches take advantage of a reference source or do the best effort without any reference when providing an absolute quality measurement.

Such as developed in Section 4.3.2, color harmony model is used in [186, 149] as a global image cue for the assessment of aesthetic quality. To our knowledge, the use of color harmony concept in the context of perceptual quality assessment has been limited to the estimation of global image cues. Thus, the main contribution of the proposed method is to take advantage of human vision knowledge in the context of color harmony assessment. In other words, we have introduced a model of color harmony in the context of a perceptual quality metric.

Perceptual quality metrics [270, 166, 183] take into account the human visual system properties to exploit local and global masking effects in evaluating image or video quality with full reference. They provide perceptual quality maps that mimic the human perception of degradations by highlighting visible degradations. In the same vein, the structural similarity index [263] is largely used due to its fair correlation with subjective judgment and its simplicity of implementation.

Previous work of Wang [263] has been extended to color by introducing hue similarity into the SSIM index [233]. Thakur and Devi [250] proposed a color quality index that performs in the spatial domain and takes advantage of the Human Visual System (HVS) properties to assess color quality with reference. Without any reference, Ouni *et al.* [199] have proposed different color statistics analysis (distribution of hue histogram, proportion and dispersion of the dominant color...) to derive a quality score, but they do not propose any local information such as a color quality map.

Nishiyama *et al.* [186] focus on color harmony theory to compute an aesthetic estimation of the input image. As it is usually done in aesthetic quality classification, features are extracted and a model is learned from an annotated ground truth. Although features are based on color harmony, there is no use of perceptual precept to derive any quality maps.

The proposed computational method relies also on Matsuda's model or templates (Figure 4.6a) that we reinterpreted in the context of quality assessment. To clarify the proposed computational method, we start by introducing some outputs computed from the quality metric in Figure 13.1.

In this chapter, we will develop the approach for bringing the Color Har-

(a) (b) (c)

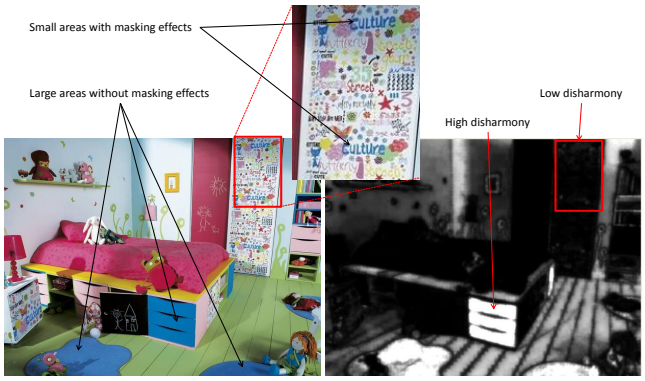
Figure 13.1: Is it possible to quantify the color harmony of a pixel in a picture? The original picture and its harmony-guided perceptual quality map are depicted. The whiter the pixels, the more disharmonious the pixels relatively to the global picture. In (c), balls of color can be sorted by disharmony level: from blue, green to yellow; red and orange are the most harmonious colors.

mony theory into a quality metric (Section 13.2), then we described the proposed method (Section 13.3). Finally, we assess the performances of the method (Section 13.4) and discuss the main findings (Section 13.4.3).

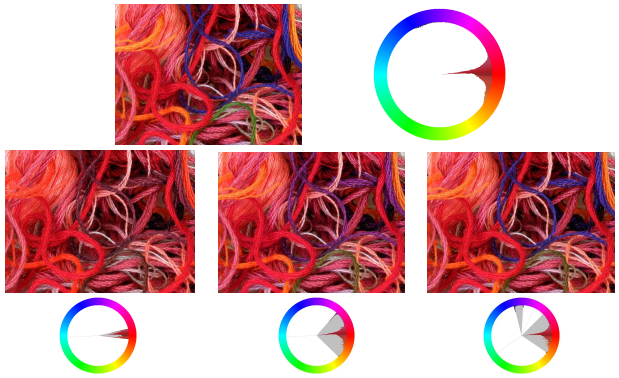
13.2 Paradigm

Such as elaborated previously, the proposed method links two distinct topics: a no-reference perceptual quality metric and the harmonious templates. The basic idea is to compute a harmony distance for each pixel regarding its corresponding hue. Thus, the harmony distance, computed on a hue basis, of each pixel is balanced by its degree of visibility in the picture. In other words, we apply spatial masking rules to weight the harmony perceived at a spatial location. Figure 13.2a illustrates the concept: blue pixels are considered disharmonious by the metric; while the large blue areas are then rated with high value (meaning high disharmony), the disharmony levels of small blue areas in the wall are tempered due to their low level of visibility or due to the high level of masking in the considered area. We assume that the spatial frequencies play a role in the perceived disharmony.

In the proposed method, we are applying Matsuda's templates as it is done usually in the image processing literature. Cohen-Or *et al.* [54] and previously Tokumaru *et al.* [252] assumed that one template at a time can be matched to a picture. Thus, Cohen-Or's method searched for the optimal template, minimizing an energy according to each template design and each potential angle $(0 - 360°)$. We rather interpret Matsuda's models as providing us an information about harmony. We are considering that each template delivers an information of harmony which is more or less relevant depending on how well they fit the picture distribution. One pragmatic consideration is that some templates are included within others, e.g. V may be inside T. Consequently, both templates may reflect to some extent (e.g. depending on the energy used



(a) On the right-hand side, the corresponding harmony quality map is depicted. The blue colors are the most disharmonious but at different levels depending on visibility threshold.



(b) Top row illustrates the original picture and its hue distribution. Bottom row depicts different harmonization algorithms having different energy for determining the candidate template. From left to right, the employed energy is the average of hue weighted by saturation (Cohen-Or-like), Kullback-Liebler divergence and Kullback-Liebler divergence combined with saliency information (computed by our method [24]). All results are more or less harmonious, but at the expense of the fidelity.

Figure 13.2: Illustration of the concept behind the harmony-guided quality assessment: (a) Contrast masking, (b) Sensitivity of the energy used for matching.

to match their distribution and the picture distribution) the degree of harmony of a picture. Such consideration has also the advantage to minimize the impact of the energy metric used to match the templates and picture distribution. This point is illustrated in Figure 13.2b. As can be observed, the original picture has a hue peak in red values and few pixels around blue and green values. Depending on the strategy of matching, those minor colors have been taken into account or not leading to a complete different candidate template. However, these three results may be considered individually harmonious. The issue around these three implementations is rather related to their fidelity to the original picture.

In the proposed metric, we consider the contribution of each template and their minimal energy with respect to the optimal angle (providing the best match to the picture distribution) in the computation of harmony distance.

13.3 Proposed method

Figure 13.3 depicts the different steps that are detailed in the next sections. Based on the previous precept, we define a harmony distance relative to the hue distance from the current pixels to all optimal templates. This distance is applied on each pixel to derive a disharmony map. Afterward, the perceptual quality map is obtained by applying masking functions that provide information regarding the local visibility. Those mentioned processing are done at different levels of resolution to mimic the human visual system. A pooling step is then required to aggregate all resolution level maps leading to a perceptual harmony quality map (Figure 13.1). A final score is then computed from the map. The reader can notice that the map as well as the score are expressed regarding a disharmony level and not relatively to the harmony. We considered such representation more intuitive since white pixels (high values) conventionally depicts outliers or a minor behavior (e.g. saliency map). The description of the

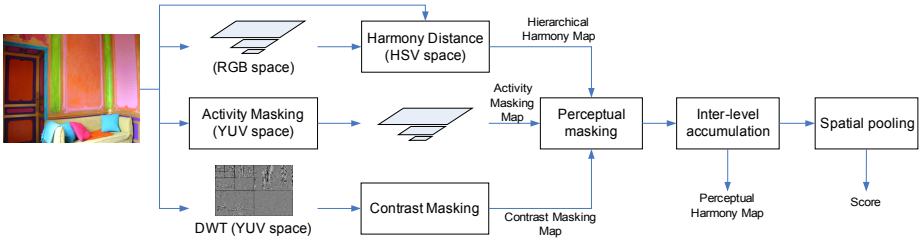


Figure 13.3: Overview of the Harmony-guided Quality Assessment: the different steps of the complete system are depicted.

following steps relies on the harmony formulation introduced in Section 11.3.

13.3.1 Harmony distance

The definition of harmonious templates reveals to be convenient information that may be arranged to provide a spatial harmony map. For a given template T_m, a hue h is considered harmonious if it is enclosed by a sector (meaning its harmonious distance is 0), while a hue outside the sector is not harmonious regarding a certain proportion defined by the hue distance $d_m(h)$. It is evaluated by computing the arc-length distance on the hue wheel (measured in degrees) to the closest sector:

$$d_m(h) = \min_{k=1\ldots K_m} \left[|h - \alpha_{m,k}| - \frac{w_{m,k}}{2} \right]^+, \tag{13.1}$$

where $|.|$ is the arc-length distance and $[.]^+ = \max(0, .)$. Then, assuming that each template (associated with its optimal angle) provides harmony information about the picture, the d_m maps are computed for all templates and combined at the pixel level. At each pixel $\mathbf{u} = (x, y)$ with associated hue $h(\mathbf{u})$, the harmony distance map $\mathcal{G}(\mathbf{u})$ accumulates the harmony distances $d_m(h(\mathbf{u}))$ as follows. The contribution of each template is weighted according to its respective energy E_m (referenced in Section 11.3 and related to equation (11.2)), to give more importance to well suited templates (having low energy);

$$\mathcal{G}(\mathbf{u}) = \left(\sum_m \left(1 - \frac{E_m}{\sum_{m'} E_{m'}} \right) d_m\left(h(\mathbf{u})\right) \right) \cdot s(\mathbf{u}) \cdot v(\mathbf{u}) \tag{13.2}$$

where s and v are saturation and value of the image. Weighting the harmony distance by saturation and value gives a more perceptual result because the more saturated the color or the higher its value, the stronger it is perceived. This point is independent of the following perceptual masking, but it already introduces perceptual concept into the metric. Some qualitative results are depicted in the second column of Figure 13.4.

13.3.2 Perceptual masking

At this stage, the harmony distances obtained at the previous step are converted into perceptual harmony maps. Spatial masking refers to the alteration of the perception of a signal by surrounding background, i.e, visibility increase (pedestal effect) or decrease (masking effect) due to the surrounding signal. As recommended by Watson *et al.* [265], both contrast masking and entropy masking are incorporated in the proposed quality metric. Contrast masking models the visibility change of the signal due to contrast values created by edges or color gradation. Entropy masking reflects the uncertainty of the masking signal, due to texture complexity. Entropy masking is also known as activity masking or local texture masking [184]; in the following, activity masking is used. The masking values are computed on the luminance channel of the YUV color space, because a perceptual masking is already employed on saturation and value in the previous step. The multi-channel behavior of the HVS is commonly simulated

by multi-resolution analysis [183, 264]. Discrete Wavelet Transform (DWT) has proven to be efficient both in term of prediction and computation performances [183].

A CDF 9/7 (Cohen-Daubechies-Feauveau) kernel is used in our implementation with L decomposition levels. Each decomposition level comprises 3 orientation sub-bands (horizontal, vertical, and oblique frequencies). The spatial frequency range of a decomposition level $l \in [1; L]$ is $\left[2^{-l} \cdot f_{max}; 2^{-l+1} \cdot f_{max}\right]$ where f_{max} is the maximum spatial frequency of the image. The number of decomposition levels is set so that the lowest resolution level L contains the frequency 1 cycle/degree: $2^{-L} \cdot f_{max} < 1c/d < 2^{-L+1} \cdot f_{max}$.

Contrast masking $C_{l,o}(\mathbf{u})$ at level l and orientation o is defined as the wavelet transformed value at site $\mathbf{u}$, weighted by the CSF (Contrast Sensitivity Function) that describes the variations in visual sensitivity to the spatial frequency. The CSF value $N_{l,o}$ is the mean value of the 2D CSF [62] over the spatial frequencies covered by the sub-band $w_{l,o}$ at level l and orientation o:

$$C_{l,o}(\mathbf{u}) = w_{l,o}(\mathbf{u}) \cdot N_{l,o}, o \in \{1, 2, 3\}. \tag{13.3}$$

The activity $A(\mathbf{u})$ is usually evaluated by the computation of entropy on a n-by-n neighborhood. Unfortunately, this tends to give high values on areas with a color step even between uniform areas, while we aim at detecting complexity of texture and not edges. Thus, it leads to overestimate the masking effects, such as pointed out by Ninassi *et al.* [183]. Instead, the spatial gradient of the image luminance Y for different directions (horizontal, vertical, diagonal) are computed for each pixel at the full resolution and the minimum value is retained in the activity map. A high value would mean a contour is present in several directions, highlighting potentially a texture:

$$A(x, y) = \min\left(\min\left(g_x, g_{d_1}\right), \min\left(g_y, g_{d_2}\right)\right) \tag{13.4}$$

with $g_x = \frac{|Y(x+1,y)-Y(x-1,y)|}{2}$, $g_y = \frac{|Y(x,y+1)-Y(x,y-1)|}{2}$, $g_{d_1} = \frac{|Y(x+1,y+1)-Y(x-1,y-1)|}{2}$ and $g_{d_2} = \frac{|Y(x-1,y+1)-Y(x+1,y-1)|}{2}$

This map A is computed at different resolutions, yielding maps A_l's, to match the HVS simulation and the resolution of contrast masking maps.

In quality assessment models, contrast values are used to find a visibility threshold for the contrast difference between two images [184]. Here the studied signal being the color values, and not contrast and structure such as in quality assessment, the masking functions thus apply to the harmony map only. To apply masking functions, a multi-resolution analysis of the harmonious distance map is adopted to match the multi-channel behavior of the HVS. A multi-resolution image pyramid is first computed in the RGB space and converted into an HSV pyramid. Harmonious distance maps G_l, $l = 1...L$ are then computed as explained above, along the pyramid. Note that no map is computed at full resolution because the added information is mostly contrast information and is poorly related to perceived color.

Then, perceptual masking is applied to harmonious distance maps at each resolution level. This step consists in applying contrast masking $C_{l,o}$ and activ-

ity masking A_l on the harmony distance map at each resolution l. Daly [62] and Nadenau [179] have proposed different complex intra-channel models for integrating the different kinds of masking effects. However this involves the tuning of parameters to adjust masking strength. Instead a simple masking function is used:

$$\mathcal{H}_l^*(\mathbf{u}) = \frac{\mathcal{G}_l(\mathbf{u})}{1 + \frac{1}{2}\left(\sum_o C_{l,o}(\mathbf{u}) + A_l(\mathbf{u})\right)}. \tag{13.5}$$

The contribution of disharmonious pixels is reduced when the masking effect is high.

13.3.3 Pooling and rating

This step consists in accumulating the perceptual harmony maps $\mathcal{H}_l^*$ for the different resolutions $l \in [1...L]$ to build the final perceptual harmony map $\mathcal{H}$. Finally, a score or rating is derived from this aggregated map. The perceptual harmony map is obtained with successive upscaling and combination of the map at each resolution level:

$$\mathcal{L}_{l-1}(\mathbf{u}) = \mathcal{L}_l(2^{-1}\mathbf{u}) + \mathcal{H}_l^*(2^{-1}\mathbf{u}), \quad \mathcal{L}_L(\mathbf{u}) = 0. \tag{13.6}$$

The perceptual harmony map is the accumulation over all the resolution levels:

$$\mathcal{H}(\mathbf{u}) = \mathcal{L}_0(\mathbf{u}). \tag{13.7}$$

Note that $\mathcal{H}$ is visually close to the harmony distance map, but integrates masking effects by decreasing impact of colors in textured areas (Figure 13.4). The final image rating is defined as:

$$\mathcal{R} = \left(\frac{1}{W \cdot H}\sum_{\mathbf{u}} \mathcal{H}(\mathbf{u})^\beta\right)^{\frac{1}{\beta}}, \tag{13.8}$$

where W and H are respectively the width and height of the original picture and β is a parameter empirically set to 2.

13.4 Validation

Due to inherent issues related to color harmony field (Section 5.4), the no-reference harmony-guided quality assessment initially suffered from a lack of ground truth for the validation. Ideally, a dataset with perceptual harmony maps as well as the associated score would be useful for the extensive validation of the proposed algorithm. These aspect has motivated our work in Chapter 9.

In this section, we divided the validation question into two usual approaches: the qualitative and quantitative appreciations. In the qualitative section, we are going to answer the following question: are the harmony-guided maps visually consistent with user perception of disharmony? In the quantitative approach, we

confront 1/ the harmony perceptual quality maps to the ground truth elaborated in Chapter 9 from the eye-tracking experiment (Chapter 8) and 2/ the harmony score to experimental ranking elaborated in the experiment described in Section 12.4.2.

13.4.1 Qualitative appreciation

The role of masking maps

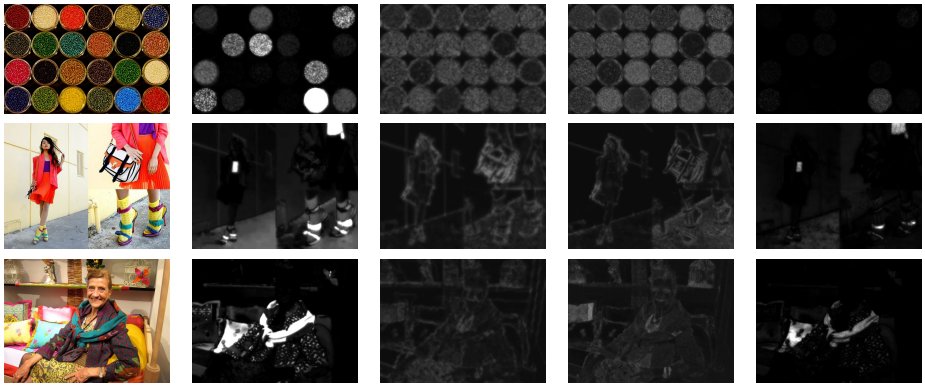


Figure 13.4: Visual appreciation of perceptual harmony-guided quality maps. The first column is the original picture, the second column is the harmony distance map, the third column is the contrast masking map and the fourth one is the activity map. Final maps are perceptual and harmony-guided in the sense they are close to harmony distance map but with applied masking effects.

As mentioned previously, the masking effect is a key concept in perceptual quality assessment. In this section, we demonstrate the interest of the two masking maps introduced in the context of harmony guidance. First row in figure 13.4 illustrates the main advantages of masking modeling. The harmony map at the first row highlights blue and green areas as being disharmonious with orange/red major hues of the picture. The corresponding masking maps are not related to hues, but to visible spatial frequencies. Regions with complex texture are detected in the two masking maps: typically, black and clear areas with low contrast regarding the background are not masking areas, while red and green areas have complicated texture with high contrast. High masking contrast areas have a reduced contribution on the final perceptual harmony map.

In the second example, the gray/blue hues of the ground are detected as being disharmonious in the harmony distance map. However, the two masking maps clearly distinguished from the two areas of ground. Indeed, the ground of the right hand side of the picture is more textured leading to high contrast and potentially no clear analysis of HVS in this kind of area. The perception of the disharmonious hues in this area is thus masked by the image content. Consequently, the contribution of such area is minimized in the final perceptual map.

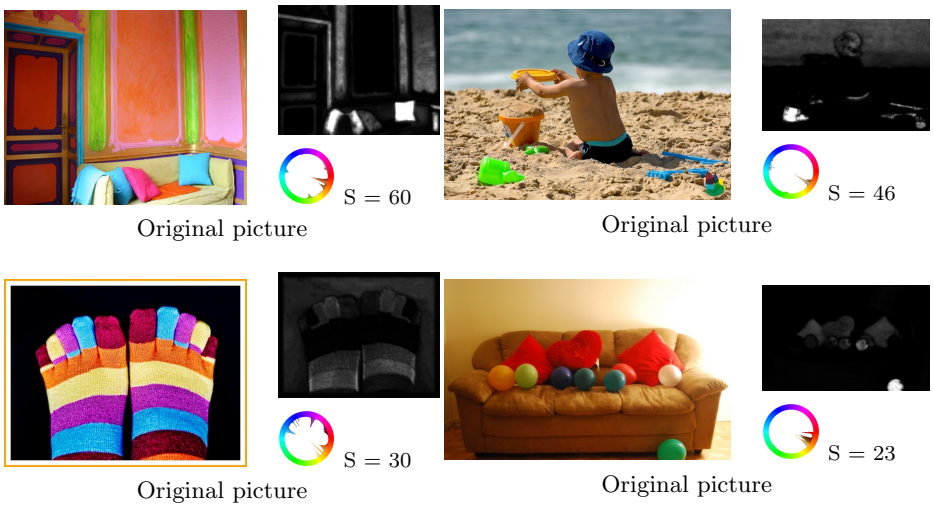


Figure 13.5: Qualitative appreciation of perceptual harmony maps and its associated score S. Also, the hue histogram of each picture is depicted.

Finally, on the third row, the complementary roles of the two masking maps can be appreciated. The blue color is perceived as a disharmonious hue in this picture. Nevertheless, all the blue pixels can not be treated the same way, since masking effects alter the perception of some small group of blue pixels (jacket of the woman). High activity masking effects (4^{th} column) appear in the cushion (above the yellow one) leading to non-detection of non-harmonious blue hues by the HVS. Although green colors (particularly on the shelf) are also mentioned as having an average level of disharmony in the harmony distance map, they are detected in the contrast masking and then masked in the final perceptual map.

The two masking maps are complementary and allow the attenuation of disharmonious regions that are not perceived by the HVS due to their complex neighboring environment.

Harmony score and map

Figure 13.5 depicts four results for a qualitative appreciation of the perceptual harmony map and its associated score: from the least harmonized pictures (top left) to the most harmonized picture (bottom right). The blue colors in left-hand side pictures have been identified as being not harmonious with regard to the other dominant colors. For the right-hand side pictures, it is rather the green colors that do not match well with the yellowish dominant colors.

13.4.2 Quantitative appreciation

In this section, we propose to appreciate quantitatively the performances of the quality metric in terms of harmony score and spatial map.

Harmony Score

Relying on the pairwise experiment described in Section 12.4.2, we exploited the ranking score estimated by this experiment on the dataset. The main idea is to estimate a potential correlation between the experimental ranking and the score provided by the metric.

In Section 12.4.4, a high correlation between the color distribution and the harmony ranking has been established ($R^2 = 0.26$ with a linear model and $R^2 = 0.48$ with a logarithmic model). Also, in Chapter 9, the creation of a ground truth evidenced the need for classifying the considered dataset into three distinct categories depending on the complexity of the stimuli, measured by the inter-observer agreement. Thus, we considered the two reliable categories estimated within the dataset: the Low Agreement (LA) and the High Agreement (HA) sets. For each of them, we computed the linear correlation R^2 between the ranking score and the estimated score from the metric s_{metric}, as well as the linear correlation between the ranking score s_{exp} and the color distribution (Table 13.1). Note that the color distribution (n_h) is expressed as the number of hue values in the histogram having more than T pixels (T=500, representing 0.1 % of pixels for a 800x600 picture).

	Linear Correlation R^2, Ranking score s_{exp} versus ...		
Dataset	Color Distribution: n_h	Metric Score: s_{metric}	f(Metric Score, Color Distribution): f(s_{metric}, n_h)
LA	0.0092	**0.48**	
HA	0.35	0.02	**0.79**

Table 13.1: Validation of the score computed from the metric. The linear correlation has been estimated between different factors: the color distribution (or hue width) of the considered stimuli, the experimental ranking score and the metric score for the two datasets. Note that in the last column a function has been determined regarding the HA dataset.

As observed in Table 13.1, the color distribution is correlated to the experimental ranking score for the HA dataset: the larger the color distribution, the lower the experimental score. For the LA dataset, where there is less consensus between observers, there is no correlation between these two factors. However, the correlation between the experimental score and the metric score is pretty high. This confirms that the harmony score computed from the proposed model is relevant for the LA dataset.

Nonetheless, this is not the case for the HA dataset. Since the color distribution influences the experimental ranking score, it seems suitable to include such feature into the metric score. We have empirically fitted a logarithmic function including the color distribution as a variable, such as described in Equation

(13.9):

$$\hat{f}(s_{metric}, n_h) = \frac{a}{n_h} \cdot log\left(\frac{b}{s_{metric}}\right), \tag{13.9}$$

with $a = 360$, $b = 8$.

The logarithmic curve is weighted by the color distribution ($a = 360$, the maximal number of hues). Since the metric score provides a score of disharmony, it is inverted in the formula to match the harmony score provided experimentally. Figure 13.6 illustrates in 2 dimensions the correlation between the fitting curve and the experimental scores.

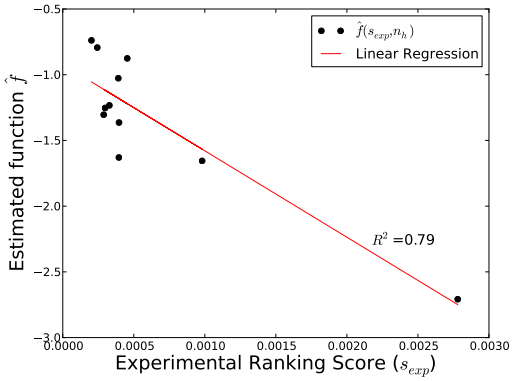
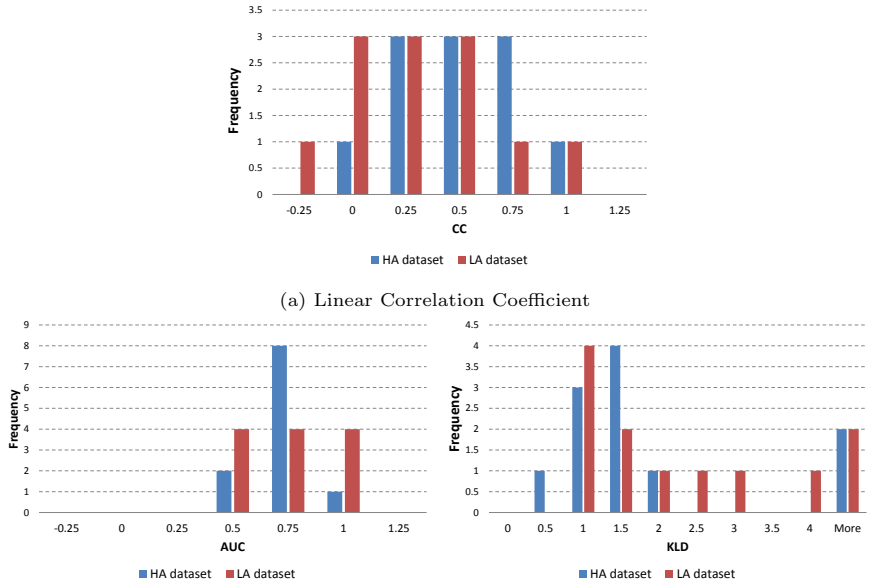


Figure 13.6: Correlation between the fitting curve and the experimental data for the harmony score.

Globally, we found a high correlation between the experimental score of harmony and those computed from the computational metric.

Perceptual Map

In this section, the maps designed as a ground truth in Chapter 9 are employed in order to confront the computational method of this chapter with experimental results. Following the same methodology as in the previous section, we distinguished the LA and HA dataset. In Chapter 9, the experimental maps for those two datasets were built differently. Since the eye fixations have been post-processed, we could not employed the recorded scanpaths and consequently, the NSS measure. Thus, the similarity in terms of Correlation Coefficient (CC), Area Under the Curve (AUC) and Kullback-Leibler Divergence (KLD) is depicted in Figure 13.7 under the form of an histogram. The inter-map similarity is limited, but does exist. Only two or three pictures per dataset come up with a high divergence. As expected, the similarity is higher for the HA dataset than for the LA one, whose the inter-observer agreement was lower.



(a) Linear Correlation Coefficient

(b) Area Under the Curve (ROC)

(c) Kullback-Leibler Divergence

Figure 13.7: Validation of the computational and perceptual harmony maps. The histograms of CC, KLD and AUC values for the two datasets are depicted.

13.4.3 Discussion

We demonstrated the role of masking maps and their interest in a metric of quality assessment dedicated to harmony concept. As a first attempt in the design of such metric, only luminance-based functions of perceptual masking were employed and tested. A significant improvement could be brought by taking advantage of masking functions dedicated to color, such as e.g. in [141].

For the validation of the harmony score, there are several unexplored paths that can be mentioned. There are two prior works that aim at predicting the harmony from content. Solli and Lenz [239] extended the Ou's model for estimating the perceived harmony of multi-colored images in a context of image indexing (Section 4.2.2). Also, Moon and Spencer extrapolated a score of harmony from their geometric formulation of color harmonies [173]. A comparison to these two models would be useful to benchmark the performances of each approach.

Following adjacent work on aesthetic assessment, we performed a comparison of recorded aesthetic scores (AVA dataset [178]) and harmony scores computed from the proposed model. Consistent with Nishiyama *et al.* [186], no correlation were found between the aesthetic and harmony scores. Aesthetic assessment is not directly related to harmony, even if this latter certainly participates to the

global aesthetic appraisal.

13.5 Summary

In this chapter, a perceptual and no-reference harmony-guided quality metric has been introduced as a new way to assess picture quality. It relies on perceptual masking effects that are important properties of Human Visual System, and on harmony templates that have previously been designed during psychological experiments. The integration of these two concepts leads to the computation of perceptual harmony-guided map and an associated score. Both types of information may be useful in the context of image content creation, edition and retouching for guiding expert and non-expert content creators such as proposed in Chapter 14.

Qualitative results show that the harmony-guided map reflects the perception of color harmony by a human eye. Perceptual maps closely mimic the HVS to take into account masking effects and to discard disharmonious regions that are not perceived. The computation of score is consistent with potential changes of colors that can be done to improve picture harmony.

The quantitative analysis reveals a correlation with two experiments serving respectively for the validation of the harmony score and the perceptual harmony maps.

Chapter 14

Color Harmony for Editing

Contribution: *Christel Chamaret, Fabrice Urban, and Lionel Oisel, "Harmony-guided image editing", in IEEE International Conference on Image Processing 2014 (ICIP 2014), Paris, France, Oct. 2014, pp. 2176-2178.*

This chapter proposes two applications that directly implement the previously proposed computational methods to guide the end user in a task of image editing. In Section 14.2, the perceptual harmony guided map provides an useful information for selecting locally the non-harmonious colors to be retouched and an algorithm extrapolates a harmonious color palette to serve the user editing task. In Section 14.3, the color harmonization process is biased by a dataset previously uploaded by the user.

First, we briefly present an overview of the available tools related to Color Harmony.

14.1 User-assisted color design: state-of-the-art

Related to color harmony, extensive work has been carried out to produce comprehensive and user-friendly tools for performing color design. Once again, main difficulties remind 1) the translation of empirical color harmony rules, 2) the usability and accessibility for non-expert audience. More specifically, colors are quantitatively enormous, color spaces or representations are also varied depicting different characteristics and color interacts with each others. Considering all these problems, designing color schemes or rendering tools have been explored by researchers. This set of tools showed up for guiding user in his design choice and more particularly in combining colors together.

In 2004, Lyons and Moretti surveyed the current state-of-the-art of tools that generates harmonious color schemes [159]. At this time, most of computer interfaces suffer from proposing strong and reliable guidance to achieve harmony schemes. Most of them relied on the color wheel representation which may limit

the user's choice: one (hue) or maybe two components are represented at a time, but they do not display the full three-dimensional structure of the color universe. Also, most of tools do not force user's choice by adjusting its previous color choices to guarantee respect of geometric harmony rules.

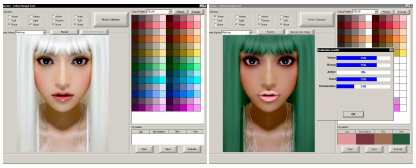
In such context, Lyons and Moretti were undertaking a long-term investigation into mathematical solutions for introducing color harmony within computer interfaces [161, 159, 160, 176]. The main idea behind their substantial work is to mathematically formalized and conceptualized the color harmony notion as *an abstract color scheme which can be represented as a rigid shape, a color molecule that is free to move within a 3D color-space*. As a missing element of previous attempt, they proposed a consistent mapping between the distances in the color space and perceived color differences. Thus, they have built the *Color Harmonizer* tool which basically operates in two steps: 1) it generates the abstract color scheme by depicting essential color relationships between interface components (button, texts...); 2) using a direct manipulation interface, the user locates the targeted sample in color space. The chosen colors are harmonious and ensure also a visual distinction between the different components. However, they specify a restrictive applicative domain related to web design. They did not demonstrated the power of their tool on pictures or more sophisticated graphics.

Having less sophisticated harmony rules, but displaying appealing results, the work of Hascoët [95] also investigated the design of web page from the color of an original picture (Figure 14.1a). Following their experience and experiments funding, Ou *et al.* have also created an editing tool [198] that tackles color harmony principles (Figure 14.1b).

Recently, a color planning and visualization system were proposed by Chen and Wang [50]. Their system includes four interesting functions: 1) the model component grouping, 2) color arrangement, 3) visualization and 4) an online evaluation. The color arrangement function directly applies Moon and Spencer's harmony as well as Munsell's law by computing contrasts for two or three-colors combinations. In addition, they applied specific rendering effects, such as Phong shading model, Cook-Torrance shading model and environment reflection effect. The user selects a group of characteristics (color, design mode...) and then the system proposes a computational visualization based on the harmony rules and rendering effects. If desired by the user, an online evaluation may be requested and allows him reiterating regarding such feedback.

In the same vein, Hu *et al.* [106] also proposed an interactive visualization tool for generating harmonious color schemes. It relies on two color harmony principles: *familial factors* (the ability for sharing attributes among colors) and *rythmic spans* (the property of repeated distances in the color space). Based on these two principles, they fix one or two components and generate schemes within their harmonic color generator tool. Visualization on patterns may be proposed to the user that can immediately appreciate the combinations he designed.

(a) Extracted from [95]



(b) Extracted from [198]

Figure 14.1: Examples of color design tools. All of them use color harmony concepts for guiding the user in his editing task. (a) Hascoët and the design of webpage [95], (b) Ou *et al.*'s editing tool [198] based on their harmony model.

Finally, we can cite two specialized design tools which have integrated the color harmony concepts in order to guide the color design of specific material.

In [51], Cheng and Liu proposed an smartphone-based application that behaves as a fashion advisor. They imagine an user taking a picture when shopping and submitting it to the known personal garment database of the user. Thus, the tool evaluates potential match, based on color harmony rules derived from Matsuda's color coordination.

Focusing on another material and usecase, Jahanian *et al.* [116, 117] have implemented an automatic tool to design magazine covers. In addition to color harmony theory, they also introduced notion of color semantics and more particularly mood-based color descriptors. From Itten's theory, they defined the harmonious color palette they can employ for text with regard to the background picture.

For the two following sections, we propose the same descriptive organization. First, the tool is introduced and its motivation is developed. Second, its design and use are described because they participate to its added value. Third, the link and the algorithm allowing to make the link with the core technology described in the previous chapter are depicted. Finally, a last subsection provides an example of usage.

14.2 Picture Retouching

14.2.1 Introduction and motivations

This tool [49] is based on the harmony-guided quality metric developed in Chapter 13. Based on the disharmonious areas identified by the metric on a picture, the end user picks up a target area and color to be retouched. Then, the tool proposes to him a harmonious color palette that guarantees the improvement of global harmony score, also computed from the harmony-guided quality metric.

When manipulating, editing, improving images, a high artistic quality is somehow purposely maximized. However, even if modern tools propose user-friendly image editing, there is no guarantee that anyone has good taste by essence for retouching color images. Improving color rendering of pictures requires an expertize. Hopefully, the scientific community, very active in the design of rules for understanding the user experience when confronting to color harmony, is also driven the transfer of such notion into comprehensive engine or tool (Section 14.1). However, their use is restricted to a color science community and they are far from being exploitable by any naive person. On the other hand, automatic solution for color harmonization appeared recently by means of new connection between color scientists and image processing community. Nonetheless, such kind of automatic solution may alter too much the original content and/or change too much the creative intent. There is a need for supervised solutions, where the user could select the area that he really wants to retouch.

In such context, we propose in this section a tool that could be integrated as a plugin in any image editing software. It aims at facilitating the color retouch of images by 1) depicting an informative representation of non-harmonious areas, 2) proposing a color palette that guarantees the increase of global color harmony. The end-user iteratively selects a color to be retouched (typically the one of an object) based on the provided guide (a spatial map of disharmonious pixels), pick up a color in the proposed harmonious color palette and validate its choice for the current area. Thus, in few clicks and without color expertize, he has maximized the color rendering of the pictures while his artistic intent is respected.

14.2.2 Implementation

Design

The User Interface (UI) has been designed in C language using QT. Having segmented the picture in uniform color areas [55] allows an intuitive and friendly way to select the disharmonious pixels. Such as usually done for image editing, the main environment is composed of several frames: the harmony-related information, i.e. harmony quality map and color palette, are displayed at the bottom of the interface, while the top area depicts the high level information, such as the original, retouched and segmented pictures. An overview of the UI is provided in Figure 14.2.

Use

This section describes the different steps and interactional aspects of the UI. The following numbering follows the steps depicted in Figure 14.2.

1. First, the user uploads a picture within the interface, then while opening the selected image, two libraries run in parallel: the color segmentation

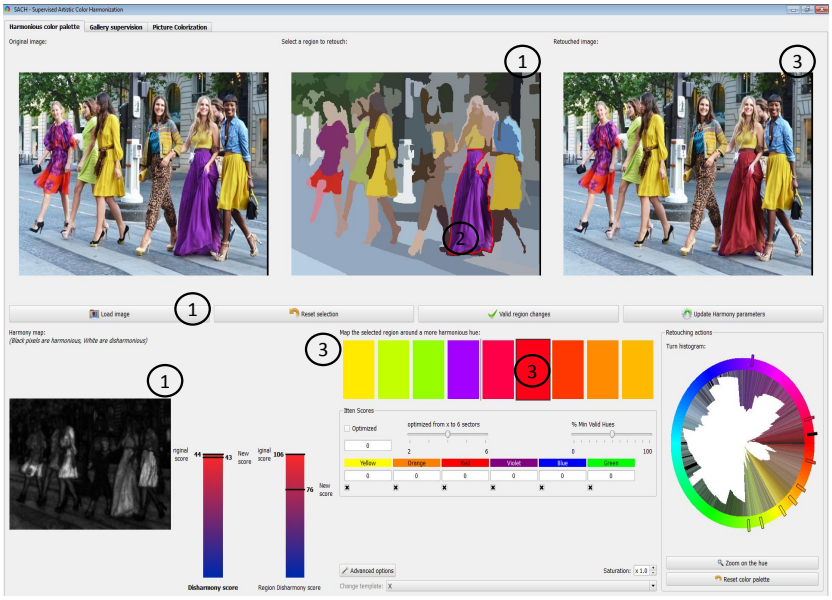


Figure 14.2: Screenshot of the User interface. The numbers refer to Section 14.2.2

and the harmony quality metric. In few seconds, the user interface outputs the results of these two algorithms. Then, the user may simply observe the harmony quality map to immediately identify disharmonious colors that he might retouch.

2. Once he takes his decision, he can pick up an uniform area on the segmentation map. Note that several uniform areas can be selected as soon as they reflect hues similar to the original area in average.

3. Then, a color palette able to improve the global color harmony is proposed to the user. He can choose the color that best fits his intent and control the improvement by checking the harmony quality map and the associated score that are instantaneously updated at each color selection.

Several options allow performing a finer color editing. Two color segmentation algorithms are plugged into the interface: the well-known meanshift implementation segments coarse uniform areas [55] while the histogram-based approach inspired from [65] provides an accurate selection of areas. Also, the user may zoom in the color palette proposal to access intermediary hues.

14.2.3 Method

Core technology

The tool is built on a core technology that is purposely introduced as a guidance for the end-user. It consists in creating computational harmony map and score that predicts or assess the degree of color harmony, as described in Chapter 13.

Color Palette Creation

This step consists in computing the optimal color palette C associated to the selected color h_s picked up by the user. This color palette is supposed to guarantee the increase of color harmony. The color palette C is computed through four main steps, where each variable is illustrated in Figure 14.3:

1. Determine the central hue h_s of the selected area,
2. Find the distribution G which is the harmony contribution of each template for each hue,
3. Find the N first maximum values in $G(h)$, such as $H = \{h_n; n = 1..N\}$ verifying that the distance between two maxima is high enough to obtain distinguishable colors:
   $\max_h G(h)_{|h_i - h_j|_{i \neq j = 1..N} > Th}$
4. Create C by selecting the hues from H that have a harmony contribution superior to the one of the hue h_s picked up by the user, $C = H - h_{n, G(h_n) < G(h_s)}$.

Th is empirically set to 10 degrees. Note that in Chapter 12, each template is preliminary rotated to map the hue distribution of the picture, such as described in Equation (11.2). Then, the harmony contribution G consists for each hue in aggregating the distance of the hue to the template border, such as done in Equation (13.2). $G(h)$ corresponds to $\mathcal{G}(\mathbf{u})$, where $\mathbf{u}$ are the spatial coordinates of the image.

14.2.4 Use Case

In this section, we provide intermediary inputs to understand the interest and results of the tool. We iterate on several colors/areas to be retouched such as depicted in Figure 14.4.

14.3 Biased Color Harmonization

14.3.1 Introduction and motivations

This tool is derived from the method of automatic color harmonization proposed in Chapter 12. Targeting the harmonization of a picture, the end user biases

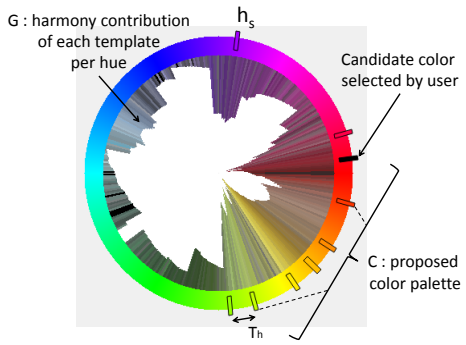


Figure 14.3: Color Palette Creation. It illustrates the case depicted in Figure 14.2. The user wants to retouch the purple skirt and choose the red color proposal.

Figure 14.4: Picture retouching Usecase. (a) Original picture and hue distribution, (b) Harmonized picture from Chapter 12, (c) Picture retouched locally with the proposed tool and the method from Chapter 13

the original processing of automatic harmonization with the colors present in a set of pictures.

This tool answers the need for harmonizing pictures based on examples. It follows the same philosophy as performing *Color Transfer* processing [220, 208, 273, 213], but less drastically. Indeed, we do not apply the colors of a reference picture, but rather maintain the original colors of the source picture while shifting the original color distribution to the colors of the reference pictures. Thus, we guarantee that the final picture contain both harmonized set of colors and potentially colors from the reference pictures.

It may be used for harmonizing different rooms of a building (Section 14.3.4) or the objects of a room or even the clothes of a wardrobe.

14.3.2 Implementation

Design

The User Interface (UI) has been designed in C language using QT. It intuitively displays the original picture (top left-hand side), the processed image (top right-hand side) and at the bottom a thumbnail version of all reference pictures. Some buttons allows changing of strategy for the bias application. The user is free to play with options and can go back to previous result with one click since the results are stored locally. An overview of the UI is provided in Figure 14.5.

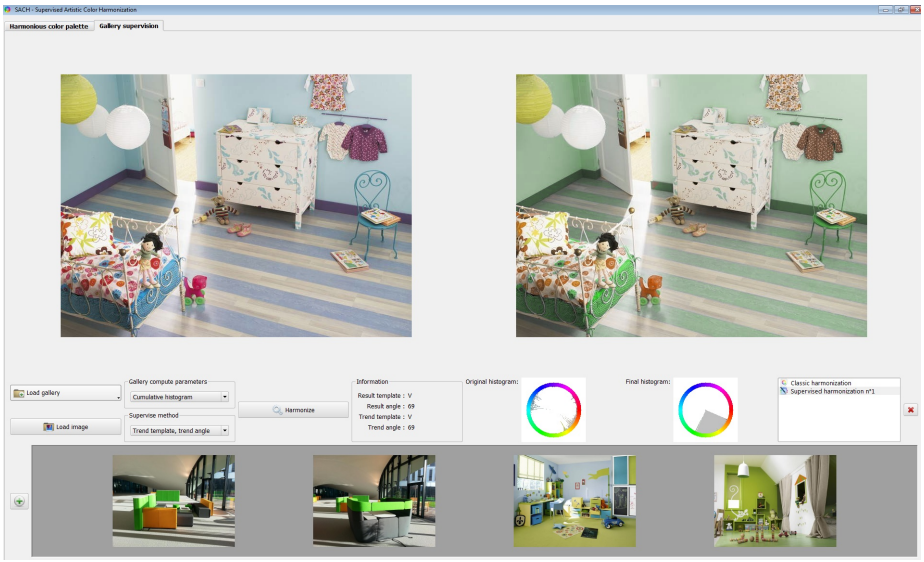
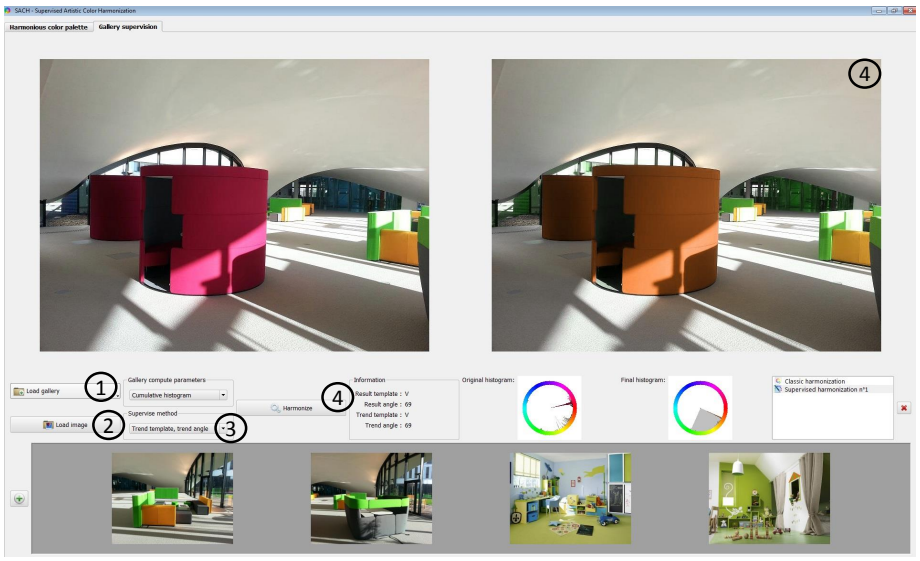


Figure 14.5: Biased Harmonization: screenshot of the User Interface.

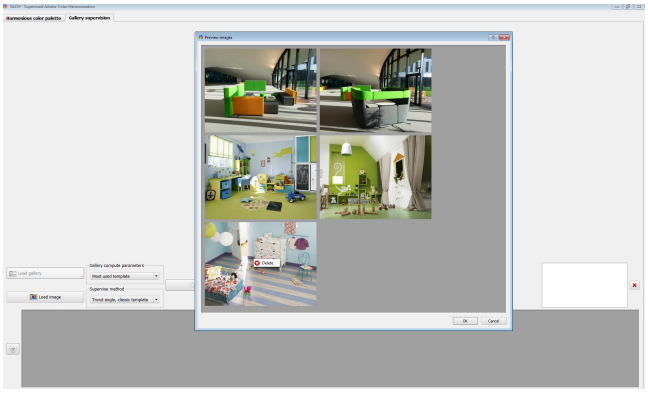
Use

This section describes the different steps and interactional aspects of the UI. The following numbering follows the steps depicted in Figure 14.6.

1. The user uploads a gallery of pictures, either from a local folder or from an internet url (Figure 14.6b).
2. The user uploads the picture to be processed.
3. The user chooses a setting for the bias strategy (details in Section 14.3.3).
4. The user clicks on *Harmonize* button, then the result is displayed. Also, at the same time regular Harmonization process is run and may be display by choosing it on the right-hand side list of processing.



(a)



(b)

Figure 14.6: Biased Harmonization: one additional result. (a) The numbers refer to the one described above. (b) Upload of reference pictures, display of thumbnails version and potential removing of pictures with right click.

14.3.3 Method

Core technology

This tool is based on the method of color harmonization described in Chapter 12. We did not perform any specific changes in the method to get adapted to this application. Only a software architecture effort was necessary to well split the C functions in order to address the way we perform the bias strategy.

Introducing the bias

The bias related to a set of reference pictures may be ingested from different ways. There are actually two settings that may highly influence the final results:

1. which information from the reference pictures we are taken into account,
2. the way those information will be *pushed* to the final harmonization processing

These two paths are depicted in the accessible options for the users. Basically, there are three options for the information extracted from the reference pictures and five options for the way they are taken into account on the harmonization process. Finding the information from reference pictures may follow the idea of computing the trend of a group of pictures, while the ingestion of bias is a shifting or mapping of colors based on previously estimated trend. Both group of options consists simply in finding a candidate template and its associated angle. These options are described below:

Trend from reference pictures

- The most representative template: from a set of consistent pictures, it is likely that the same template, providing the best match with the original hue distribution, occurs for several reference pictures. Thus, the template that appears the most is selected. The selected angle is the average (modulo 360) of all candidate angles.
- Cumulative hue histogram: all hues present in all pictures are aggregated within an unique histogram of hue. Thus, a candidate template is extrapolated from this global histogram.
- Template histogram: each candidate template are described through a 1D hue distribution (for details, see Section 11.3 and Equation (11.3)) and aggregated in one final histogram. The selected template is the one that best match this global template histogram.

A representative template and its angle $(\hat{T}_m^{ref}, \hat{\alpha}_m^{ref})$ are found from this step that is supposed to catch the color spirit of the reference pictures.

Ingestion of bias

The color bias is ingested at this step. It consists in four different strategies that may be chosen by the user. It relies in the trend template and associated angle extrapolated also from different potential methods. Since we may want not to penalize too much the original color distribution, most of strategies keep the original color harmonization parameters $(\hat{T}_m, \hat{\alpha}_m)$ somehow (except the third strategy). It allows keeping the original color *spirit* of the picture in case this latter is very far from the reference pictures. On the uploaded picture, shift the hue histogram based on following parameters:

- use the trend angle combined with the best template from regular harmonization: $(\hat{T}_m, \hat{\alpha}_m^{ref})$
- use the trend template combined with the best angle from regular harmonization: $(\hat{T}_m^{ref}, \hat{\alpha}_m)$
- use the trend template and angle: $(\hat{T}_m^{ref}, \hat{\alpha}_m^{ref})$
- use the trend template and the optimal angle for this template on the original picture: $(\hat{T}_m^{ref}, \alpha_{m\ |T_m=\hat{T}_m^{ref}})$

14.3.4 Use Case

In this section, we provide some results related to the strategies detailed in Section 14.3.3 and illustrated in Figure 14.7. In this picture is chosen the trend from the *Cumulative hue histogram*. Thus, the four ways to ingest the bias are computed. As can be observed, some strategies provide the same visual results, the V template providing a green-yellow-orange rendering. This is conform to the original set of reference pictures. The first strategy of bias ingestion lead to the selection of X template, preserving the blue color, such as present in the original picture, but also in the second picture of the reference set.

14.4 Summary

In this Chapter, we introduced two user-friendly and artistically-guided user interfaces: one for color retouching (Section 14.2) and one for biasing from an existing dataset the harmonization processing (Section 14.3).

Preliminary to the description of these tools, we introduced the state-of-the-art in terms of editing tools including color harmony theory (Section 14.1). Most of the approaches target a specific area of application, such as fashion, magazine covers, website and so on. As a minority, the most generic tools rely on complex formulation of color harmony and some associated distances, making the interface difficult to be manipulated by naive users. The two proposed tools' ambition is to make accessible the color harmony theory through a hidden and intuitive formulation of it and a user-friendly interface.

For the picture retouching tool, we demonstrated an efficient implementation of the quality metric introduced in Chapter 13, where we reinterpreted the use of color template introduced by Matsuda [168]. By means of this editing tool, any user may retouch the color of picture by picking up a harmonious color from the color palette presented by the interface. He is guided to both choose the disharmonious color and replace it with a suitable one.

For the biased harmony tool, the main idea is to harmonize a picture globally by means of a picture dataset. Typically, it could be used in a context of homogenization of colors for a website, a catalog, a book etc.

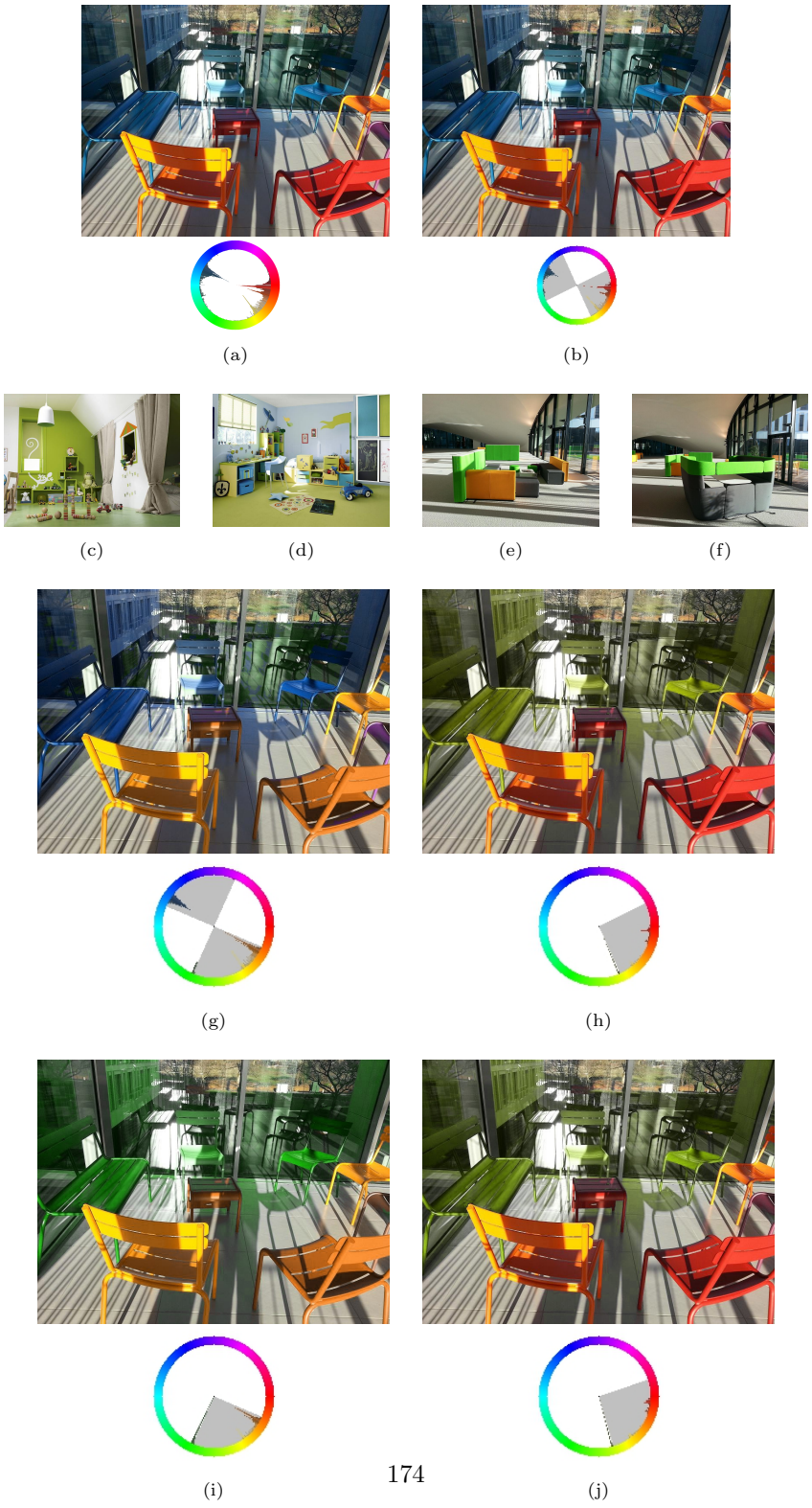


Figure 14.7: Biased Harmonization Usecase. (a) Original picture and hue distribution, (b) Harmonized picture from Chapter 12, (c)-(f) Reference pictures for bias processing, (g)-(j) the four different strategies are applied for biasing the harmonization process

Chapter 15

General Conclusion

This chapter concludes the development elaborated in this thesis. The topic concerns the Color Harmony, that has been addressed under two mains axis: an experimental and a computational perspectives. First, we summarize the achievements of the proposed work by reminding the limitations about the literature. Second, we propose some perspectives for future directions of research.

15.1 Achievements

Literature status

The first part of the thesis presents the literature in relation with the different fields tackled in this thesis. A non-exhaustive review of visual attention principles (Chapter 2) as well as eye-tracking processes (Chapter 3) are presented in order to lay the foundation of the employed protocols and methods. Chapter 4 depicts a large view of the theory and models of Color Harmony as well as their field of application, notably in the image processing community. Additionally, we also provide a review of editing tools in Section 14.1.

This literature review points out a number of limitations and non-investigated paths (Chapter 5):

- Despite the fact that color preferences and color mood (or emotion) have been recently studied experimentally, there are still pending questions about their universality. They seem highly related to culture, age, background and reveal tricky to be evaluated. In such context, we position the topic of color harmony as being less subject to be interfered by such side factors. The existence of empirical work, with universal rules (complementarity, similarity of colors...), leading to geometric representation of harmony on color wheel testifies of such feeling.

- The validity of collected data remains questionable. The measured data and more precisely the inter-rater agreement have not been intensively studied. Thus, the generalization of numerical models computed from

these data has not been convincingly demonstrated. In addition, they are based on two or three color combination samples. Regarding the geometric models (wheel-based), they do not propose an objective framework for their application; they are open to interpretation. Consequently, their practical nature is challenged.

- The algorithms employing color harmony theory as a core concept to perform image processing suffer from being quantitatively tested and validated. This is typically due to the lack of available ground truth, that can be easily explained as follows: how can we reasonably design a perfectly harmonious picture for any kind of content that would serve as benchmark? Can we ensure a high agreement without bias related to age, gender etc.?

Experimental view

We assumed a strong relationship between visual attention and color harmony. Therefore, we investigated the measurement of such factor through an eye-tracking protocol under a search task. We evidenced a high inter-observer agreement that demonstrated the universal aspect of the color harmony notion. Such way of experiencing Color Harmony was pioneer and could suffer from deeper analysis; however, it paved the way to further investigations.

From the experiment 1 on color factor (Chapter 7), we found that the color factor does not influence the visual attention deployment. The human fixation maps compared between color stimuli and their grey counterparts were not significantly different. This result is delivered partly by the visual attention community. Anyway, it made us confident to apprehend the effects related to color harmony in experiment 2 (Chapter 8) without expecting (low-level) color-related attentional interferences.

From the experiment 2 on color harmony, we ended up with several conclusions. During a task protocol with eye-tracking for measuring color harmony, the inter-observer consistency was high (confirmed by several employed metrics). There was an influence of color distribution and color diversity in the assessment of color harmony. The diversity of color distribution did not seem to be the main factor, rather the spatial arrangement of color variety.

From the data collected in experiment 2, the eye movements and individuals' assessment of color harmony were relevant enough to create a ground truth (Chapter 9). Two aspects prevented from using directly the eye fixation maps recorded with a color harmony task. First, they were too noisy due to parasite fixations related to search task and low level attentional mechanisms. Second, the complexity of stimuli measured by their color distribution varied and lead us to clustered them according to observers' consistency in their assessment.

Globally, we concluded that the concept of color harmony is well understood by observers, consistent and close to universality. Thus, we are convinced that the concept is homogeneous enough to design any computational models that

would predict in some extend a common human behavior and assessment.

Computational view

In Chapter 12, a color harmonization method, outperforming the state-of-the-art, is introduced and validated. Building on the work of Cohen-Or [54], the method proposed three main contributions to reduce possible artifacts and to obtain a visually consistent result. First, a cost function measuring the statistical distance between a weighted hue histogram and the possible templates was used, jointly with the information of saliency. Consequently, the selection of harmony templates better fits the original image color distribution. Second, a new color mapping function contracted less the hues within the desired template and leads to a more consistent mapping. Third, a dedicated color segmentation was able to segment small histogram modes, while leading to satisfactory result.

Also in this chapter, we attempted to convince that the method is achieving good performances by tackling different tracks for validation. We provided traditional visual results, but we also carefully answered the questions raised during the exhaustive exposition of limitations. A major contribution dealt with the introduction of a pairwise protocol in the context of color harmonization. We demonstrated that harmonized pictures are objectively rated more harmonious than their original counterpart (19 over 23 stimuli).

In Chapter 13, a perceptual and no-reference harmony-guided quality metric has been designed as a new way to assess picture quality. It relied on perceptual masking effects that are important properties of Human Visual System, and on experimental harmony templates. The integration of these two concepts led to the design of perceptual harmony-guided map and an associated score.

Qualitative results showed that the harmony-guided map reflects the perception of color harmony by the human eye. Perceptual maps closely mimicked the human visual system to take into account masking effects and to discard disharmonious regions that are not perceived. The computation of score was consistent with potential changes of colors that can be done to improve picture harmony. The quantitative analysis revealed a correlation between the computed outputs and the results of two experiments serving respectively for the validation of the harmony score and the perceptual harmony maps.

In Chapter 14, we introduced two user-friendly and artistically-guided user interfaces. The two proposed tools made accessible the color harmony theory through a hidden and intuitive formulation of it and a user-friendly interface.

For the first tool, we demonstrated an efficient implementation of the quality metric introduced in Chapter 13. Also, we reinterpreted the use of color template introduced by Matsuda [168], by assuming that each template has a role in the assessment of color harmony. By means of this editing tool, any user may retouch the color of picture by picking up a harmonious color from the color palette presented by the interface. He is guided to both choose the disharmonious color and replace it with a suitable one.

For the second tool, the main idea is to harmonize a picture globally by means of a picture set. Typically, it could be used in a context of homogenization of colors for a website, a catalog, a book...

15.2 Perspectives

The task of characterizing the mechanisms related to color harmony is indisputably challenging. Many factors (age, culture, personal background, stimuli features and so on) may potentially play their role, while they may be difficult to quantify or measure. The path is still long before being able to get good prediction and to prove it by means of a relevant ground truth. However, we hope to have brought first answers to this problem in this thesis.

When thinking about the perspectives or the future directions of this work, a remaining challenge lies in **the establishment of a solid link between the experimental findings and the modeling of Color Harmony**. If an unique regret should be mentioned about this thesis, it would be to have failed achieving a strong relationship and transfer of knowledge between the experimental investigations and the computational modeling.

Less conceptually, but pretty relevant, a nice way of investigations would be to establish a relationship experimentally between Color Harmony and other concepts, such as e.g. Color Mood, Color Preferences. Indubitably, it would be a first step to the specialization of models per categories (age, nationality...) of subjects.

In addition to these global perspectives, several ideas can be mentioned to extend the contributions of the thesis.

15.2.1 Experimental perspectives

Ground truth

- The dataset used for creating the ground truth could be extended with natural pictures and designed on image categories (street, countryside and so on).
- A more reliable computational method to derive a ground truth from the collected eye fixation data would be relevant. More intensive (or cross-) validation could be conducted.

Stimuli impact

- The experimental work could be extended to a different dataset. As for the creation of the ground truth, a dataset of natural pictures could be relevant, on condition that the stimuli are controlled in some way regarding

their low-level features (spatial frequencies, color etc.) and their semantic meaning (faces, objects etc.). Also, an interesting investigation would be to measure the effects of color harmony on image categories (street, countryside and so on).

- The impact of color distribution on color harmony ranking could be more investigated. Particularly, the spatial arrangement and color complexity could be controlled and specific behavior could then be evidenced.

Brain mechanisms

More investigations could be conducted to understand at which stage/area the color harmony functions is activated in the brain. By proposing to measure such concept, it could help to better understand the mechanisms and strategy inferred in the human brain. This could be performed by means of specific medical apparatus, such as EEG, MRI and so on. Maybe it could help precising the color inference located in different areas of the brain.

Dedicated protocols

It would be of great benefits to define dedicated protocols to specifically evaluate the performances between different algorithms. As an example, we still do not know **objectively** if our Harmonization algorithm performs better than the one of Cohen-Or, even though we demonstrated the gain of harmonizing pictures (with our algorithm) versus not harmonizing them.

Generally, this is a typical problem met when comparing algorithms (such as Color Transfer, Stylization and so on) where the results are subjective and that a ground truth is neither available nor able to be generated.

15.2.2 Computational perspectives

Algorithm improvements

- The harmony quality map could be certainly improved by ingesting a Contrast Sensitivity Function dedicated to color. An interesting work could be also to extend this metric to video content and then to employ the principles of temporal masking.

- Another interesting idea would be to use additional harmony templates such as those inferred from large scale data.

- A certain number of thresholds or fixed values are present in the different algorithms: power β in HQA, number of salient pixels in the harmonization and so on. The ground truth could help tuning the values to find the suitable set of parameters.

Color Harmony models

In the thesis, we did use color harmony model or template as being a core technology. We did question their reliability, but we employed them as they were available and commonly adopted. Finally, the representation of color harmony through the different families illustrated in the Part 1 has not been addressed in the thesis. Interestingly, it would be relevant to establish a link between the families of color harmony models. Their different representation maybe made them complementary for the prediction of color harmony.

Related to the ground truth as well, it could make sense to define specific models or subfamily of models in relation with image categories. Finally, higher level of information related to the semantic of the picture could be added to the design of color harmony models.

Benchmark

The harmony quality score could be benchmarked with different competitive methods:

- the extension of Ou and Luo [195] to natural pictures by Solli and Lenz [239] (it was not possible at this time),
- the Itten scores produced by the implementation of Sauvaget *et al.* [227]

The editing tools would benefit from being evaluated regarding their time saving and attractiveness in terms of usage.

15.2.3 Industrial perspectives

This work has been developed in an industrial context within Technicolor company. There are two points which motivated this work under the industrial perspcetive.

First, Technicolor is a key actor in Hollywood for the post-production of movies and is well recognized for its Color expertize (since it colorized the first black and white movies). Thus, the choice of Color Harmony has been motivated by these reasons, but also in order to facilitate the discussion between scientists and colorists and fill the gap between these two communities evolving within Technicolor.

Second, there are a number of concrete applications in the framework of Technicolor business. Any editing or post-production tool integrating color harmony rules could potentially benefit to save time of creatives. The insertion of Computer Graphics or Virtual objects within an existing background/scene is a typical daily task realized by a high number of creatives. Also, these tools could be potentially licensed to external company.

There are many potential directions to continue after this thesis, from short- to long-term investigations, from experimental to computational sensitivities.

Appendix A

Benchmark of the Technicolor's Visual Attention Model

The visual attention model has been compared to most recent models via the MIT Saliency Benchmark[1] [38, 123]. Figure A.1 depicts the results numerically for the seven available metrics.

Since the Technicolor's model has been purposely simplified to address real-time application, this is not surprising that it points not at the top level, as very algorithmically sophisticated model. Nonetheless, we can see the good performance of the *center bias* baseline. Basically, all models that do not implement such prior knowledge perform worse than this baseline. This is our case.

Figures A.2, A.3, A.4, A.5, A.6,A.7 and A.8 localize the Technicolor's model visually for each metric.

[1] Available at: http://saliency.mit.edu

	AUC_Judd	SIM	EMD	AUC_Borji	sAUC	CC	NSS
Baseline: Chance [?]	0.5	0.31	5.73	0.5	0.5	0	0
Achanta	0.52	0.29	5.77	0.52	0.52	0.04	0.13
IttiKoch	0.6	0.2	5.17	0.54	0.53	0.14	0.43
SUN saliency	0.67	0.38	5.1	0.66	0.61	0.25	0.68
Baseline: Permutation Control [?]	0.68	0.33	4.73	0.59	0.5	0.2	0.49
Torralba saliency	0.68	0.39	4.99	0.68	0.62	0.25	0.69
Murray model (Chromatic Induction Wavelet Model)	0.7	0.38	5.18	0.69	0.65	0.27	0.73
Rosin Saliency 1	0.71	0.4	4.86	0.7	0.62	0.29	0.76
Stochastic fixation prediction (SFP)	0.71	0.41	4.56	0.7	0.62	0.3	0.8
Self-resemblance by LARK	0.71	0.41	4.55	0.69	0.64	0.31	0.83
Technicolor	0.71	0.4	5.03	0.7	0.64	0.29	0.78
NARFI saliency	0.73	0.38	4.75	0.61	0.55	0.33	0.83
Aboudib Magnification Saliency	0.74	0.44	4.24	0.72	0.58	0.39	0.99
Weighted Maximum Phase Alignment Model (WMAP)	0.74	0.42	4.49	0.67	0.63	0.34	0.97
Generalized Nonlocal Mean Saliency (GNM)	0.74	0.42	4.49	0.67	0.63	0.34	0.97
Context-Aware saliency	0.74	0.43	4.46	0.73	0.65	0.36	0.95
Co-Occurrence Histogram based Saliency	0.74	0.44	4.49	0.71	0.66	0.36	1.01
Adaptive Whitening Saliency Model (AWS)	0.74	0.43	4.62	0.73	0.68	0.37	1.01
Random Center Surround Saliency	0.75	0.44	3.81	0.74	0.55	0.38	0.95
Quantum-Cuts (QCUT)	0.75	0.39	4.57	0.67	0.57	0.4	1.07
Visual Conspicuity (VICO)	0.75	0.44	4.38	0.71	0.6	0.37	0.97
IttiKoch2	0.75	0.44	4.26	0.74	0.63	0.37	0.97
Image Signature	0.75	0.43	4.49	0.74	0.66	0.38	1.01
Local+Global Saliency Model (LGS)	0.76	0.42	4.63	0.76	0.66	0.39	1.02
Multi-Resolution AIM (MR-AIM)	0.77	0.43	4.04	0.76	0.55	0.39	0.96
LMF	0.77	0.45	4.22	0.76	0.64	0.41	1.07
Saliency for Image Manipulation	0.77	0.46	4.17	0.76	0.64	0.43	1.14
AIM	0.77	0.4	4.73	0.75	0.66	0.31	0.79
RARE2012	0.77	0.46	4.11	0.75	0.67	0.42	1.15
Baseline: Center [?]	0.78	0.39	4.81	0.77	0.51	0.38	0.92
Rosin Saliency 2	0.78	0.48	3.43	0.73	0.53	0.45	1.13
Aboudib Magnification Saliency (Bottom-up v2)	0.78	0.48	3.56	0.75	0.56	0.45	1.12
MKL-based model	0.78	0.42	4.4	0.78	0.61	0.42	1.08
Local Saliency Model (LS)	0.78	0.43	4.4	0.77	0.64	0.39	1.02
Sampled Template Collation	0.79	0.39	4.79	0.78	0.54	0.4	0.97
CWS model	0.79	0.46	3.81	0.78	0.55	0.45	1.11
Region Contrast (RC)	0.79	0.48	3.48	0.78	0.55	0.47	1.18
Multiresolution CNN (Mr-CNN)	0.79	0.48	3.71	0.75	0.69	0.48	1.37
Fast and Efficient Saliency (FES)	0.8	0.49	3.36	0.73	0.59	0.48	1.27
Baseline: one human [?]	0.8	0.38	3.48	0.66	0.63	0.52	1.65
CovSal	0.81	0.47	3.39	0.67	0.57	0.45	1.22
Spatially Weighted Dissimilarity Saliency (SWD)	0.81	0.46	3.89	0.8	0.59	0.49	1.27
Judd Model	0.81	0.42	4.45	0.8	0.6	0.47	1.18
Graph-Based Visual Saliency (GBVS)	0.81	0.48	3.51	0.8	0.63	0.48	1.24
Ensembles of Deep Networks (eDN)	0.82	0.41	4.56	0.81	0.62	0.45	1.14
Mixture of Saliency Models	0.82	0.44	4.22	0.81	0.62	0.52	1.34
Outlier Saliency (OS)	0.82	0.5	3.33	0.81	0.62	0.54	1.38
Boolean Map based Saliency (BMS)	0.83	0.51	3.35	0.82	0.65	0.55	1.41
SalNet	0.83	0.52	3.31	0.82	0.69	0.58	1.51
Deep Gaze 1	0.84	0.39	4.97	0.83	0.66	0.48	1.22
DeepFix	0.87	0.67	2.04	0.8	0.71	0.78	2.26
SALICON	0.87	0.6	2.62	0.85	0.74	0.74	2.12

Figure A.1: Benchmark of the Technicolor's visual attention model on the MIT dataset. Arrangement is performed regarding the *AUC Judd* metric.

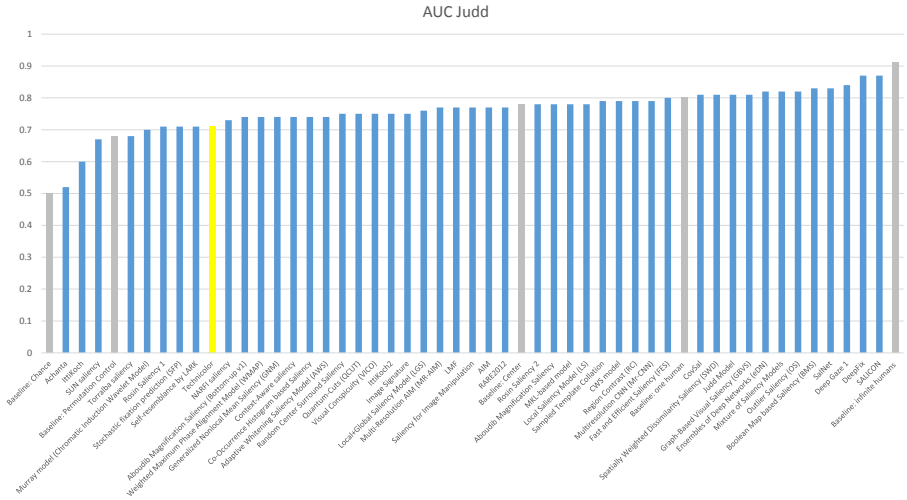


Figure A.2: Benchmark of the Technicolor's visual attention model on the MIT dataset: AUC Judd metric.

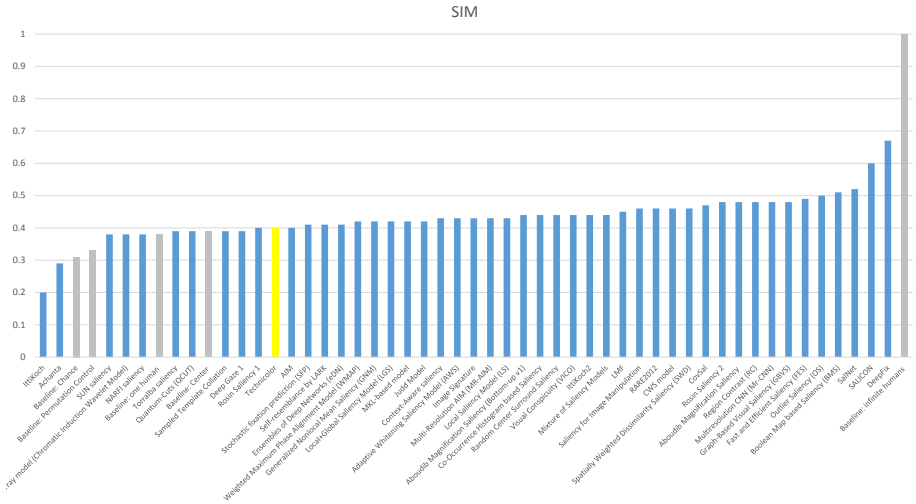


Figure A.3: Benchmark of the Technicolor's visual attention model on the MIT dataset: Similarity metric.

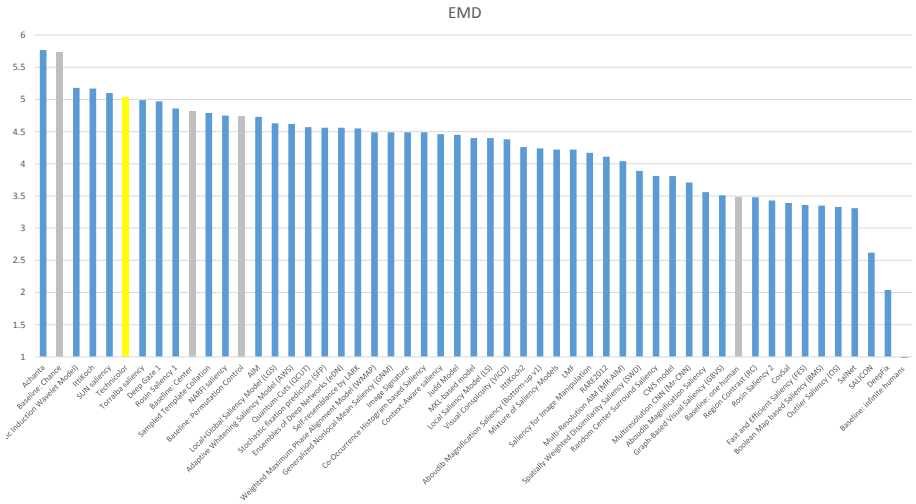


Figure A.4: Benchmark of the Technicolor's visual attention model on the MIT dataset: EMD metric.

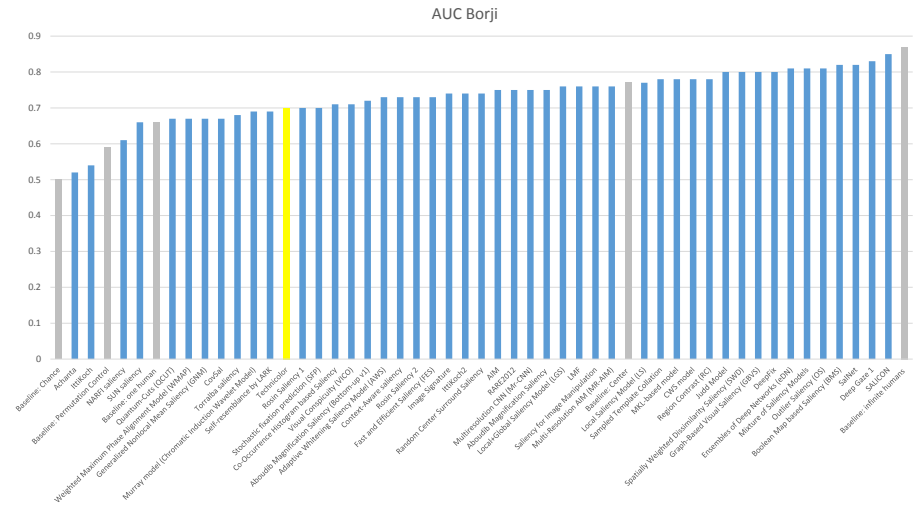


Figure A.5: Benchmark of the Technicolor's visual attention model on the MIT dataset: AUC Borji metric.

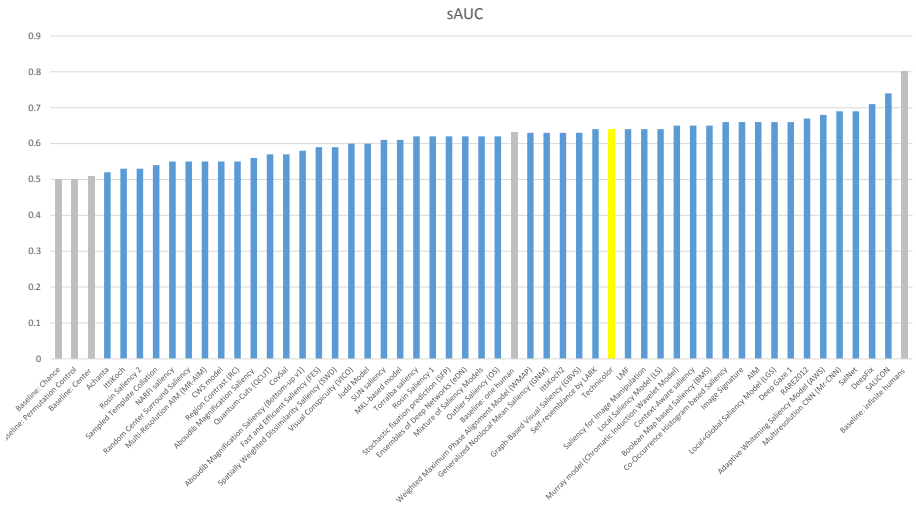


Figure A.6: Benchmark of the Technicolor's visual attention model on the MIT dataset: sAUC metric.

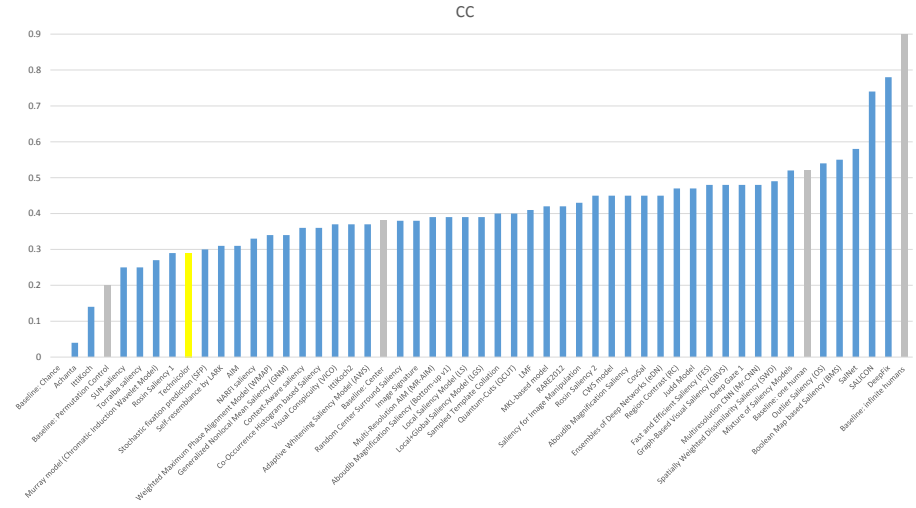


Figure A.7: Benchmark of the Technicolor's visual attention model on the MIT dataset: Coefficent Correlation metric.

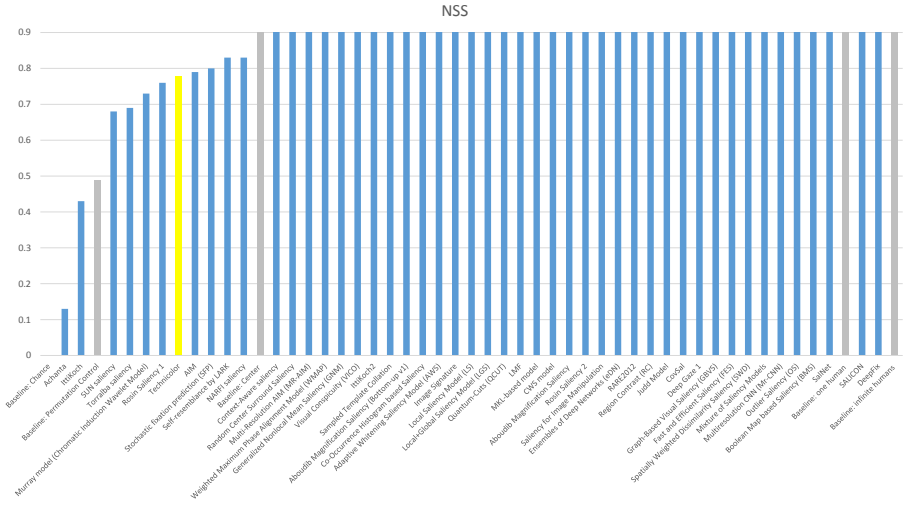


Figure A.8: Benchmark of the Technicolor's visual attention model on the MIT dataset: Normalized Saliency Scanpath metric.

Bibliography

[1] Adobe kuler. 47, 65

[2] Colourlovers. 47, 65

[3] Dataset of annotated images. 77

[4] Web page of ral design system. 50

[5] CIE 2004. Colorimetry. *3rd edn, Commission Internationale de l'eclairage*, 2004. 62

[6] R. Achanta and S. Susstrunk. Saliency detection for content-aware image resizing. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 1005–1008, Nov 2009. 16

[7] Alper Acik, Adjmal Sarwary, Rafael Schultze-Kraft, Selim Onat, and Peter Konig. Developmental changes in natural viewing behavior: Bottom-up and top-down differences between children, young adults and older adults. *Front Psychol*, 1(0):207, 2010. 30

[8] C. Ackerman and L. Itti. Robot steering with spectral image information. *IEEE Transactions on Robotics*, 21(2):247–251, Apr 2005. 16

[9] Kinjiro Amano and David H Foster. Influence of local scene color on fixation position in visual search. *JOSA A*, 31(4):A254–A262, 2014. 71, 72, 75, 114, 115

[10] Kinjiro Amano, David H Foster, Matthew S Mould, and John P Oakley. Visual search in natural scenes explained by local color properties. *JOSA A*, 29(2):A194–A199, 2012. 71, 75, 114, 115

[11] James R. Antes. The time course of picture viewing. *Journal of Experimental Psychology*, 103(1):62–70, 1974. 96

[12] Marc Antonini, Michel Barlaud, Pierre Mathieu, and Ingrid Daubechies. Image coding using wavelet transform. *Image Processing, IEEE Transactions on*, 1(2):205–220, 1992. 19

[13] Rudolf Arnheim. *A Review of Proportion. In Module, Proportion, Symmetry, Rhythm.* Vision and Value. Gyorgy Kepes, ed. New York: George Braziller, 1966. 40

[14] A Terry Bahill, Deborah Adler, and Lawrence Stark. Most naturally occurring human saccades have magnitudes of 15 degrees or less. *Investigative Ophthalmology & Visual Science*, 14(6):468–469, 1975. 95

[15] Mokryun Baik, Hyeon-Jeong Suk, Jeongmin Lee, and Kyungah Choi. Investigation of eye-catching colors using eye tracking. In *Proc. SPIE*, volume 8651, pages 86510W–86510W–6, 2013. 71

[16] Dana H. Ballard and Mary M. Hayhoe. Modelling the role of task in the control of gaze. *Vis cogn.*, 17(6-7):1185–1204, August 2009. 69

[17] Antoine Barbot, Michael S. Landy, and Marisa Carrasco. Differential effects of exogenous and endogenous attention on second-order texture contrast sensitivity. *Journal of Vision*, 12(8), 2012. 52

[18] Yoann Baveye, Jean-Noel Bettinelli, Emmanuel Dellandrea, Liming Chen, and Christel Chamaret. A large video data base for computational models of induced emotion. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 0:13–18, 2013. 6, 58

[19] Yoann Baveye, Christel Chamaret, Emmanuel Dellandréa, and Liming Chen. A protocol for cross-validating large crowdsourced data: The case of the liris-accede affective video dataset. In *Proceedings of the 2014 International ACM Workshop on Crowdsourcing for Multimedia*, pages 3–8. ACM, 2014. 6

[20] Yoann Baveye, Emmanuel Dellandrea, Christel Chamaret, and Liming Chen. From crowdsourced rankings to affective ratings. In *Multimedia and Expo Workshops (ICMEW), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014. 6

[21] Yoann Baveye, Emmanuel Dellandréa, Christel Chamaret, and Liming Chen. Deep learning vs. kernel methods: Performance for emotion prediction in videos. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on*, pages 77–83. IEEE, 2015. 6

[22] Yoann Baveye, Emmanuel Dellandrea, Christel Chamaret, and Liming Chen. Liris-accede: A video database for affective content analysis. *Affective Computing, IEEE Transactions on*, 6(1):43–55, 2015. 6

[23] Yoann Baveye, Fabrice Urban, and Christel Chamaret. Image and video saliency models improvement by blur identification. In *Computer Vision and Graphics*, pages 280–287. Springer Berlin Heidelberg, 2012. 6

[24] Yoann Baveye, Fabrice Urban, Christel Chamaret, Vincent Demoulin, and Pierre Hellier. Saliency-guided consistent color harmonization. In *Computational Color Imaging*, volume 7786 of *Lecture Notes in Computer Science*, pages 105–118. Springer Berlin Heidelberg, 2013. 57, 88, 89, 92, 117, 138, 151

[25] V.A. Billock, D.W. Cunningham, P.R. Haviq, and B.H. Tsou. Perception of spatiotemporal random fractals: an extension of colorimetric methods to the study of dynamic texture. *J Opt Soc Am A Opt Image Sci Vis.*, 18(10):2404–13, 2001. 91

[26] F. Birren. *Principles of Color: A Review of Past Traditions and Modern Theories of Color Harmony*. Van Nostrand Reinhold, 1969. ii, 1

[27] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. State-of-the-Art of Visualization for Eye Tracking Data. In R. Borgo, R. Maciejewski, and I. Viola, editors, *EuroVis - STARs*. The Eurographics Association, 2014. 29

[28] P. Bodrogi and T.Q. Khan. *Illumination, Color and Imaging: Evaluation and Optimization of Visual Displays*. Wiley Series in Display Technology. Wiley, 2012. 62

[29] Nadia Bolognini, Francesca Frassinetti, Andrea Serino, and Elisabetta Làdavas. "acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, 160(3):273–282, 2005. 8

[30] Ali Borji and Laurent Itti. State-of-the-Art in Visual Attention Modeling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 35(1):185–207, January 2013. 14, 15, 20

[31] Ali Borji and Laurent Itti. Defending yarbus: Eye movements reveal observers' task. *Journal of Vision*, 14(3):29, 2014. 69, 117

[32] Y. Y Boykov and M. P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In *Proceedings. Eighth IEEE International Conference on Computer Vision, 2001. ICCV 2001.*, volume 1, page 105–112, 2001. 133

[33] Yuri Y Boykov and M-P Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in nd images. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 105–112. IEEE, 2001. 55

[34] R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs, i. the method of paired comparisons. *Biometrika*, 39:324–345, 1952. 139

[35] R.P. Brent. *Algorithms for Minimization Without Derivatives*. Dover Books on Mathematics. Dover Publications, 1973. 54

[36] Kenneth E. Burchett. Color harmony attributes. *Color Research & Application*, 16(4):275–278, 1991. 48

[37] Kenneth E. Burchett. Color harmony. *Color Research & Application*, 27(1):28–31, 2002. 39

[38] Zoya Bylinskii, Tilke Judd, Ali Borji, Laurent Itti, Fredo Durand, Aude Oliva, and Antonio Torralba. Mit saliency benchmark. 182

[39] Manuel G Calvo and Peter J Lang. Parafoveal semantic processing of emotional visual scenes. *Journal of Experimental Psychology: Human Perception and Performance*, 31(3):502, 2005. 75

[40] ManuelG. Calvo and PeterJ. Lang. Gaze patterns when looking at emotional pictures: Motivationally biased attention. *Motivation and Emotion*, 28(3):221–243, 2004. 75, 82

[41] Roger HS Carpenter. *Movements of the eyes (2nd rev.* Pion Limited, 1988. 11

[42] Marisa Carrasco. Visual attention: The past 25 years. *Vision research*, 51(13):1484–1525, 2011. 9

[43] Monica S. Castelhano, Michael L. Mack, and John M. Henderson. Viewing task influences eye movement control during active scene perception. *Journal of Vision*, 9(3):6, March 2009. 30, 95, 96, 117

[44] C. Chamaret. Color impact in visual attention deployment considering emotional images. volume 8291, pages 82911T–82911T–10, 2012. 6

[45] C. Chamaret, J.C. Chevet, and O. Le Meur. Spatio-temporal combination of saliency maps and eye-tracking assessment of different strategies. In *Image Processing (ICIP), 2010 17th IEEE International Conference on*, pages 1077–1080, Sept 2010. 6, 18

[46] Christel Chamaret, Claire-Helene Demarty, Vincent Demoulin, and Gwenaelle Marquant. Experiencing the interestingness concept within and between pictures. In *SPIE Conference Human Vision and Electronic Imaging*, 2016. 6

[47] Christel Chamaret, Sylvain Godeffroy, Patrick Lopez, and Olivier Le Meur. Adaptive 3d rendering based on region-of-interest. In *IS&T/SPIE Electronic Imaging*, pages 75240V–75240V. International Society for Optics and Photonics, 2010. 6

[48] Christel Chamaret and Olivier Le Meur. Attention-based video reframing: Validation using eye-tracking. In *ICPR*, pages 1–4. IEEE, 2008. 6

[49] Christel Chamaret, Fabrice Urban, and Lionel Oisel. Harmony-guided image editing. In *IEEE International Conference on Image Processing 2014 (ICIP 2014)*, pages 2176–2178, Paris, France, October 2014. 164

[50] Kuen-Meau Chen and Ming-Jen Wang. Computer aided three-dimensional colour planning and visualisation system for product design. In *Journal of the International Colour Association*, 2012. 121, 163

[51] C.I. Cheng and D.S.M. Liu. Mobile fashion advisor - a novel application in ubiquitous society. *special issue on "Future Generation Smart Space", International Journal of Smart Home*, 2(2):59–76, 2008. 164

[52] Michel E. Chevreul. *The Principles of Harmony and Contrast of Colours, and Their Applications to the Arts*. London: Longman, Brown, Green, and Longmans., 2nd edition edition, 1855. ii, 1, 40

[53] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46, 1960. 65

[54] D. Cohen-Or, O. Sorkine, R. Gal, T. Leyvand, and Y. Q Xu. Color harmonization. In *ACM Transactions on Graphics (TOG)*, volume 25, page 624–630, 2006. iv, vi, 44, 52, 53, 54, 60, 65, 121, 123, 126, 127, 128, 129, 133, 137, 146, 147, 150, 178

[55] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002. 50, 165, 166

[56] T. N. Cornsweet and H. D. Crane. Accurate two-dimensional eye tracker using first and fourth purkinje images. *J. Opt. Soc. Am.*, 63(8):921–928, Aug 1973. 23

[57] Vincent Courboulay and Matthieu Perreira Da Silva. Real-time computational attention model for dynamic scenes analysis: from implementation to evaluation. In *SPIE Photonics Europe*, pages 84360O–84360O. International Society for Optics and Photonics, 2012. 18

[58] A. Coutrot and N. Guyader. Toward the introduction of auditory information in dynamic visual attention models. In *2013 14th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS)*, pages 1–4, July 2013. 15

[59] A. Coutrot and N. Guyader. How saliency, faces, and sound influence gaze in dynamic social scenes. *Journal of Vision*, 14(8):5, 2014. 15

[60] Filipe Cristino, Sebastiaan Mathôt, Jan Theeuwes, and Iain D Gilchrist. Scanmatch: A novel method for comparing fixation sequences. *Behavior research methods*, 42(3):692–700, 2010. 32

[61] S. Daly. A visual model for optimizing the design of image processing algorithms. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, volume 2, pages 16–20 vol.2, Nov 1994. 17

[62] Scott Daly. In *Digital images and human vision*, chapter The visible differences predictor: an algorithm for the assessment of image fidelity, pages 179–206. MIT Press, Cambridge, MA, USA, 1993. 154, 155

[63] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Ze Wang. Studying aesthetics in photographic images using a computational approach. In *ECCV (3)'06*, page 288–301, 2006. 149

[64] Marianne DeAngelus and Jeff B. Pelz. Top-down control of eye movements: Yarbus revisited. *Visual Cognition*, 17(6-7):790–811, 2009. iii, 3, 11, 30, 69

[65] J. Delon, A. Desolneux, J-L Lisani, and A-B Petro. A nonparametric approach for histogram segmentation. *IEEE Transactions on Image Processing*, 16(1):253–261, 2007. 134, 166

[66] Xiaoyan Deng, Sam K Hui, and J Wesley Hutchinson. Consumer preferences for color combinations: An empirical analysis of similarity-based color relationships. *Journal of Consumer Psychology*, 20(4):476–484, 2010. 63

[67] Gunilla Derefeldt, Tiina Swartling, Ulf Berggrund, and Peter Bodrogi. Cognitive color. *Color Research & Application*, 29(1):7–19, 2004. 62

[68] Richard Dewhurst, Marcus Nyström, Halszka Jarodzka, Tom Foulsham, Roger Johansson, and Kenneth Holmqvist. It depends on how you look at it: Scanpath comparison in multiple dimensions with multimatch, a vector-based approach. *Behavior Research Methods*, 44(4):1079–1100, 2012. 32

[69] N. Dhavale and L. Itti. Saliency-based multi-foveated mpeg compression. *Proc. IEEE Seventh International Symposium on Signal Processing and its Applications, Paris, France*, pages 229–232, Jul 2003. 16

[70] Michael Dorr, Thomas Martinetz, Karl R. Gegenfurtner, and Erhardt Barth. Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision*, 10(10):1–17, August 2010. 95, 97

[71] J Duncan. Cooperating brain systems in selective perception and action. In T Inui and J L McClelland, editors, *Attention and Performance*, pages 85–105. MIT Press, Cambridge, MA, 1996. 106

[72] MA Dzulkifli and MF Mustafar. The influence of colour on memory performance: A review. *he Malaysian Journal of Medical Sciences*, 20(2):3–9, 2013. 62

[73] C. W. Eriksen and Y. Y. Yeh. Allocation of attention in the visual field. *Journal of experimental psychology. Human perception and performance*, 11(5):583–597, October 1985. 10

[74] Charles W Eriksen and Robert L Colegate. Selective attention and serial processing in briefly presented visual displays. *Perception & Psychophysics*, 10(5):321–326, 1971. 11

[75] Hans J Eysenck. A critical and experimental study of colour preferences. *The American Journal of Psychology*, pages 385–394, 1941. 63

[76] H Sheikh Faridul, T Pouli, C Chamaret, J Stauder, E Reinhard, D Kuzovkin, and A Tremeau. Colour mapping: A review of recent methods, extensions and applications. *Computer Graphics Forum*, 2015. 6, 53

[77] Frank H. Farley and Alfred P. Grant. Arousal and cognition: Memory for color versus black and white multimedia presentation. *The Journal of Psychology*, 94(1):147–150, 1976. 62

[78] Tom Fawcett. An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874, June 2006. 34

[79] Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971. 65, 139

[80] Francesca Frassinetti, Nadia Bolognini, and Elisabetta Làdavas. Enhancement of visual perception by crossmodal visuo-auditory interaction. *Experimental Brain Research*, 147(3):332–343, 2002. 8

[81] Hans-Peter Frey, Christian Honey, and Peter König. What's color got to do with it? the influence of color on visual attention in different categories. *Journal of Vision*, 8(14):6, 2008. 71, 74, 115

[82] Hans-Peter Frey, Kerstin Tanja Wirz, Verena Willenbockel, Torsten Betz, Cornell Schreiber, Tomasz Troscianko, and Peter Konig. Beyond correlation: do color features influence attention in rainforest? *Frontiers in human neuroscience*, 5:36, 2011. 71, 115

[83] Erin Goddard, Damien J. Mannion, J. Scott McDonald, Samuel G. Solomon, and Colin W. G. Clifford. Color responsiveness argues against a dorsal component of human v4. *Journal of Vision*, 11(4), 2011. 13

[84] Johann W. Goethe. *Theory of colours*. Cambridge, Mass: M.I.T. Press, 1970. 40, 46

[85] Melvyn A. Goodale and Milner. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25, January 1992. 13

[86] Peter De Graef, Dominie Christiaens, and Géry d'Ydewalle. Perceptual effects of scene context on object identification. *Psychological Research*, 52(4):317–329, 1990. 96

[87] Walter C. Granville. Color harmony: What is it? *Color Research & Application*, 12(4):196–201, 1987. 39, 40

[88] Michelle R. Greene, Tommy Liu, and Jeremy M. Wolfe. Reconsidering yarbus: a failure to predict observers' task from eye movement patterns. *Vision Research*, 62:1–8, 2012. iii, 3, 11, 69, 98, 117

[89] ThomasC. Greene, PaulA. Bell, and WilliamN. Boyer. Coloring the environment: Hue, arousal, and boredom. *Bulletin of the Psychonomic Society*, 21(4):253–254, 1983. 62

[90] Lukas Gruber, Denis Kalkofen, and Dieter Schmalstieg. Color harmonization for augmented reality. In *ISMAR*, pages 227–228. IEEE, 2010. 56, 57

[91] N. K. Hadjikhani, A.K Liu, P. Cavanagh, A.M Dale, and R. B. H. Tootell. Retinotopy and color sensitivity in human visual cortical area v8. *Nature Neuroscience*, 1:235–241, 1998. 62

[92] Shahrbanoo Hamel, Nathalie Guyader, Denis Pellerin, and Dominique Houzet. Contribution of Color Information in Visual Saliency Model for Videos. In *6th International Conference on Image and Signal Processing 2014 (ICISP 2014)*, number 8509 in Lecture Notes in Computer Science, pages 213–220, Cherbourg, France, June 2014. Springer International Publishing. 72, 75

[93] Shahrbanoo Hamel, Dominique Houzet, Denis Pellerin, and Nathalie Guyader. Does color influence eye movements while exploring videos? *Journal of Eye Movement Research*, 8(1):1–10, 2015. 72

[94] N. Harkness. The colour wheels of art, perception, science and physiology. *Optics Laser Technology*, 38:219–229, June 2006. ii, 1

[95] M. Hascoet. Visual color design. In *Information Visualisation (IV), 2012 16th International Conference on*, pages 62–67, July 2012. 163, 164

[96] Sabine E. Hasler, David; Suesstrunk. Measuring colorfulness in natural images. *Proceedings of the SPIE*, 5007:87–95, 2003. 77

[97] H. Von Helmholtz. *Handbuch der Physiologischen Optik*. Leipzig : Leopold Voss, 1867. 2, 9, 12

[98] Andrea Helo, Sebastian Pannasch, Louah Sirri, and Pia Rämä. The maturation of eye movement behavior: Scene viewing characteristics in children and adults. *Vision Research*, 103(0):83 – 91, 2014. 30

[99] John M. Henderson, James R. Brockmole, Monica S. Castelhano, and Michael Mack. Visual saliency does not account for eye movements during visual search in real-world scenes. In *In Eye movements: A window on mind and brain*, pages 537–562. Oxford: Elsevier., 2007. 32

[100] JohnM. Henderson and GrahamL. Pierce. Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review*, 15(3):566–573, 2008. 28

[101] E. Hering. *Outlines of a theory of the light sense*. Harvard University Press, 1964. 13

[102] CA Heywood, A Gadotti, and A Cowey. Cortical area v4 and its role in the perception of color. *The Journal of Neuroscience*, 12(10):4056–4065, 1992. 62

[103] Tien Ho-Phuoc, Nathalie Guyader, and Anne Guérin-Dugué. A functional and statistical bottom-up saliency model to reveal the relative contributions of low-level visual guiding factors. *Cognitive Computation*, 2(4):344–359, 2010. 71, 74, 75

[104] Tien Ho-Phuoc, Nathalie Guyader, Frederic Landragin, and Anne Guerin-Dugue. When viewing natural scenes, do abnormal colors impact on spatial or temporal parameters of eye movements? *Journal of Vision*, 12(2):1–13, February 2012. 71, 72, 95, 106, 114

[105] Schupp HT, Stockburger J, Codispoti M, Junghöfer M, Weike AI, and Hamm AO. Selective visual attention to emotion. *J Neurosci.*, 27(5):1082 – 9, 2007. 75, 79, 80, 82

[106] Guosheng Hu, Zhigeng Pan, Mingmin Zhang, De Chen, Wenzhen Yang, and Jian Chen. An interactive method for generating harmonious color schemes. *Color Research & Application*, 39(1):70–78, 2014. 121, 163

[107] Katherine Humphrey, Geoffrey Underwood, and Tony Lambert. Salience of the lambs: A test of the saliency map hypothesis with pictures of emotive objects. *Journal of vision*, 12(1):22, 2012. 75, 76

[108] Xing Huo and Jieqing Tan. An improved method for color harmonization. In *Image and Signal Processing, 2009. CISP '09. 2nd International Congress on*, pages 1–4, oct 2009. 56, 133

[109] David E. Irwin. Visual memory within and across fixations. In Keith Rayner, editor, *Eye Movements and Visual Cognition*, Springer Series in Neuropsychology, pages 146–165. Springer New York, 1992. 25

[110] Johannes Itten. *The art of color : the subjective experience and objective rationale of color*. Van Nostrand Reinhold, New York, 1973. ii, 1

[111] L. Itti. Quantifying the Contribution of Low-Level Saliency to Human Eye Movements in Dynamic Scenes. *Visual Cognition*, 12(6):1093–1123, August 2005. 18

[112] L. Itti, N. Dhavale, and F. Pighin. Realistic avatar eye and head animation using a neurobiological model of visual attention. *Proc. SPIE 48th Annual International Symposium on Optical Science and Technology*, 5200:64–78, Aug 2003. 16

[113] L. Itti, C. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, Nov 1998. 15, 76, 81

[114] ITU-T. Subjective video quality assessment methods for multimedia applications., 1999. 139

[115] Margaret C Jackson, Chia-Yun Wu, David EJ Linden, and Jane E Raymond. Enhanced visual short-term memory for angry faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2):363, 2009. 63

[116] Ali Jahanian, Jerry Liu, Qian Lin, Daniel Tretter, Eamonn O'Brien-Strain, Seungyon Claire Lee, Nic Lyons, and Jan Allebach. Recommendation system for automatic design of magazine covers. In *Proceedings of the 2013 International Conference on Intelligent User Interfaces*, IUI '13, pages 95–106, New York, NY, USA, 2013. ACM. 164

[117] Ali Jahanian, Jerry Liu, Qian Lin, Daniel R. Tretter, Eamonn O'Brien-Strain, Seungyon Lee, Nic Lyons, and Jan P. Allebach. Automatic design of colors for magazine covers. In *Proc. SPIE*, volume 8664, 2013. 164

[118] William James. *The principles of psychology*. H. Holt and company, New York, 1890. 9

[119] Halszka Jarodzka, Kenneth Holmqvist, and Marcus Nyström. A vector-based, multidimensional scanpath similarity measure. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 211–218, New York, NY, USA, 2010. ACM. 32

[120] D. Joshi, R. Datta, E. Fedorovskaya, Quang-Tuan Luong, J. Z Wang, Jia Li, and Jiebo Luo. Aesthetics and emotions in images. *Signal Processing Magazine, IEEE*, 28(5):94–115, 2011. 58

[121] Timothée Jost, Nabil Ouerhani, Roman von Wartburg, René Müri, and Heinz Hügli. Assessing the contribution of color in visual attention. *Computer Vision and Image Understanding*, 100(1–2):107 – 123, 2005. Special Issue on Attention and Performance in Computer Vision. 72

[122] D.B. Judd and G. Wyszecki. *Color in business, science, and industry*. Pure and Applied Optics Series. Wiley, 1975. 40

[123] Tilke Judd, Fredo Durand, and Antonio Torralba. A benchmark of computational models of saliency to predict human fixations. In *MIT Technical Report*, 2012. 182

[124] Christopher Kanan, Dina NF Bseiso, Nicholas A Ray, Janet H Hsiao, and Garrison W Cottrell. Humans have idiosyncratic and task-specific scanpaths for judging faces. *Vision research*, 108:67–76, 2015. 69

[125] Christopher Kanan, Nicholas A. Ray, Dina N. F. Bseiso, Janet H. Hsiao, and Garrison W. Cottrell. Predicting an observer's task using multi-fixation pattern analysis. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '14, pages 287–290, New York, NY, USA, 2014. ACM. 69

[126] Kai Kaspar, Teresa-Maria Hloucal, Jürgen Kriz, Sonja Canzler, Ricardo Ramos Gameiro, Vanessa Krapp, and Peter König. Emotions' impact on viewing behavior under natural conditions. *PloS one*, 8(1):e52737, 2013. 75, 76, 115

[127] Kai Kaspar and Peter König. Overt attention and context factors: the impact of repeated presentations, image type, and individual motivation. *PloS one*, 6(7):e21719, 2011. 76

[128] Naz Kaya and Helen H. Epps. Color-emotion associations: Past experience and personal preference. 62

[129] D. Y. Kim. The interactive effects of colors on visual attention and working memory: In case of images of tourist attractions. *Proceedings of 2010 International CHRIE Conference,*, 2010. 74

[130] Shigenobu Kobayashi. The aim and method of the color image scale. *Color research & application*, 6(2):93–107, 1981. 62

[131] C. Koch and S. Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4:219–227, 1985. 14, 15, 91

[132] John Krauskopf, David R. Williams, and David W. Heeley. Cardinal directions of color space. *Vision Research*, 22(9):1123 – 1131, 1982. 17, 19

[133] K. Krippendorff. Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement*, 30(1):61–70, apr 1970. 65, 139

[134] R.G. Kuehni. *Color: An Introduction to Practice and Principles*. Wiley, 2004. 39

[135] Dmitry Kuzovkin, Christel Chamaret, and Tania Pouli. Descriptor-based image colorization and regularization. In *Computational Color Imaging*, pages 59–68. Springer International Publishing, 2015. 6

[136] Nummenmaa L., Hyönä J., and Calvo M.G. Eye movement assessment of selective attentional capture by emotional pictures. *Emotion*, 6(2):257 – 68, 2006. 72, 75

[137] M Land, N Mennie, and J Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999. 117

[138] M. F. Land and D. N. Lee. Where we look when we steer. *Nature*, 369(6483):742–4, 1994. 69

[139] J. Richard Landis and Gary G Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174, March 1977. 140

[140] Peter J Lang, Margaret M Bradley, and Bruce N Cuthbert. International affective picture system (iaps): Affective ratings of pictures and instruction manual. *Technical report A-8*, 2008. 115

[141] Patrick Le Callet, Abdelhakim Saadane, and Dominique Barba. Orientation selectivity of opponent-colour channels. In *Perception ECVP Abstract Supplement*, volume 28, 1999. 17, 160

[142] Patrick Le Callet, Abdelhakim Saadane, and Dominique Barba. Frequency and spatial pooling of visual differences for still image quality assessment. In *Proc. SPIE*, volume 3959, pages 595–603, 2000. 17

[143] Olivier Le Meur and Thierry Baccino. Methods for comparing scanpaths and saliency maps: strengths and weaknesses. *Behavior Research Methods*, 45(1):251–266. 10.3758/s13428-012-0226-9. 32, 35

[144] Olivier Le Meur, Patrick Le Callet, and Dominique Barba. Predicting visual fixations on video based on low-level visual features. *Vision research*, 47(19):2483–2498, 2007. 15, 16, 17, 18, 76

[145] Anat Lechner. Mapping critical color-harmony contingencies. In *International Conference on Colour Harmony, Budapest, Hungary*, April 24-26 2007. 51

[146] T. R. Lee, D. L. Tang, and C. M. Tsai. Exploring color preference through eye tracking. *AIC Colour*, 3(5):333–336, 2005. 72

[147] VI Levenshtein. Binary Codes Capable of Correcting Deletions, Insertions and Reversals. *Soviet Physics Doklady*, 10:707, 1966. 32

[148] Anat Levin, Dani Lischinski, and Yair Weiss. Colorization using optimization. In *ACM SIGGRAPH 2004 Papers*, SIGGRAPH '04, pages 689–694, New York, NY, USA, 2004. ACM. 57

[149] Congcong Li and Tsuhan Chen. Aesthetic visual quality assessment of paintings. *Selected Topics in Signal Processing, IEEE Journal of*, 3(2):236–252, April 2009. 58, 149

[150] Jing Li, Marcus Barkowsky, and Patrick Le Callet. Boosting paired comparison methodology in measuring visual discomfort of 3DTV: performances of three different designs. In *SPIE 8648*, volume 86481V, March 2013. 139

[151] Ting Li, Yoann Baveye, Christel Chamaret, Emmanuel Dellandréa, and Liming Chen. Continuous arousal self-assessments validation using real-time physiological responses. In *International Workshop on Affect and Sentiment in Multimedia (ASM)*, 2015. 6

[152] Hantao Liu and Ingrid Heynderickx. Visual attention modeled with luminance only: from eye-tracking data to computational models. In *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics - VPQM 2010*, jan 2010. 18, 71, 75, 82, 106

[153] Ningning Liu, Emmanuel DellandréA, Liming Chen, Chao Zhu, Yu Zhang, Charles-Edmond Bichot, StéPhane Bres, and Bruno Tellez. Multimodal recognition of visual concepts using histograms of textual concepts and selective weighted late fusion scheme. *Comput. Vis. Image Underst.*, 117(5):493–512, May 2013. 58

[154] Gerald L Lohse and DJ Wu. Eye movement patterns on chinese yellow pages advertising. *Electronic Markets*, 11(2):87–96, 2001. 23

[155] Peng Lu, Zhijie Kuang, Xujun Peng, and Ruifan Li. Discovering harmony: A hierarchical colour harmony model for aesthetics assessment. In *Computer Vision–ACCV 2014*, pages 452–467. Springer, 2015. 47

[156] Peng Lu, Xujun Peng, Ruifan Li, and Xiaojie Wang. Towards aesthetics of image: A bayesian framework for color harmony modeling. *Signal Processing: Image Communication*, 2015. 47

[157] Peng Lu, Xujun Peng, Xinshan Zhu, and Ruifan Li. An el-lda based general color harmony model for photo aesthetics assessment. *Signal Processing*, 120:731 – 745, 2016. 47

[158] Yiwen Luo and Xiaoou Tang. Photo and video quality evaluation: Focusing on the subject. In *Proceedings of the 10th European Conference on Computer Vision: Part III*, ECCV '08, pages 386–399, Berlin, Heidelberg, 2008. Springer-Verlag. 58

[159] Paul Lyons and Giovanni Moretti. Nine tools for generating harmonious colour schemes. In *in Proc APCHI 2004 elsewhere in this volume. 2004.* SpringerVerlag, 2004. 162, 163

[160] Paul Lyons and Giovanni Moretti. Incorporating groups into a mathematical model of color harmony for generating color schemes for computer interfaces. In *IEEE International Conference on Virtual Environments, Huamn-Computer Interfaces, and Measurement Systems*, 2005. 163

[161] Paul Lyons, Giovanni Moretti, and Mark Wilson. Color group selection for computer interfaces. In *Proc. SPIE*, volume 3959, pages 302–313, 2000. 121, 163

[162] Jana Machajdik and Allan Hanbury. Affective image classification using features inspired by psychology and art theory. In *Proceedings of the international conference on Multimedia*, MM '10, page 83–92, New York, NY, USA, 2010. ACM. 58, 77

[163] Thomas J Madden, Kelly Hewett, and Martin S Roth. Managing images in different cultures: A cross-national study of color meanings and preferences. *Journal of international marketing*, 8(4):90–107, 2000. 63

[164] S. Mannan, K.H. Ruddock, and D.S. Wooding. Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-d images. *Spatial Vision*, 9(3):363–386, 1995. 32

[165] S.K. Mannan, K.H. Ruddock, and D.S. Wooding. The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision*, 10(3):165–188, 1996. 32

[166] Mark A. Masry and Sheila S. Hemami. A metric for continuous quality evaluation of compressed video with severe distortions. *Signal Processing: Image Comm*, 19:133–146, 2004. 149

[167] Sebastiaan Mathôt, Filipe Cristino, Iain D. Gilchrist, and Jan Theeuwes. A simple way to estimate similarity between pairs of eye movement sequences. *Journal of Eye Movement Research*, 5(1):1–15, 2012. 32, 36, 37, 94, 97, 98

[168] Y. Matsuda. In *Color Design*. Asakura Shoten (in Japanese), 1995. ii, 2, 44, 53, 65, 90, 117, 121, 123, 126, 141, 173, 178

[169] Robert D. McIntosh and Thomas Schenk. Two visual streams for perception and action: Current trends. *Neuropsychologia*, 47(6):1391 – 1396, 2009. Perception and Action. 13

[170] Olivier Le Meur, Patrick Le Callet, Dominique Barba, and Dominique Thoreau. A coherent computational approach to model Bottom-Up visual attention. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(5):802–817, May 2006. 15, 16, 17, 130

[171] Mark Mills, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman, and Michael D Dodd. Examining the influence of task set on eye movements and fixations. *Journal of vision*, 11(8):17, 2011. 30

[172] Parry Moon and Domina E. Spencer. Geometric formulation of classical color harmony. *Journal of the Optical Society of America*, (34):46–50, 1944. 41, 42, 43, 44, 65, 123

[173] PARRY MOON and DOMINA EBERLE SPENCER. Aesthetic measure applied to color harmony. *J. Opt. Soc. Am.*, 34(4):234–242, Apr 1944. 43, 160

[174] Anush Moorthy, Pere Obrador, and Nuria Oliver. Towards computational models of the visual aesthetic appeal of consumer videos. In *Proceedings of the 11th European conference on Computer vision: Part V*, pages 1–14, Heraklion, Crete, Greece, 2010. Springer-Verlag. 58, 149

[175] J. Moran and R. Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science (New York, N.Y.)*, 229(4715):782–784, August 1985. 13

[176] Giovanni Moretti, Paul Lyons, and Stephen Marsland. Computational production of colour harmony. part 1: A prototype colour harmonization tool. *Color Research & Application*, 38(3):203–217, 2013. 163

[177] Albert Munsell. A grammar of color, a basic treatise on the color system. from the original version of 1921. *New York: Van Nostrand Reinhold*, 1969. ii, 1, 41, 43, 50

[178] Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2408–2415. IEEE, 2012. 160

[179] M. Nadenau. *Integration of the Human Color Vision Models into High Quality Image Compression*. Ph.D thesis, Ecole Polytechnique Federale de Lausanne, 2000. 155

[180] A. Nemcsics. Colour dynamics: Environmental colour design. *Color Research & Application*, 19(4):310–311, 1994. 50

[181] I. Newton and W. Innys. *Opticks:: Or, A Treatise of the Reflections, Refractions, Inflections and Colours of Light*. William Innys at the West-End of St. Paul's., 1730. ii, 1

[182] PhiBang Nguyen, J. Fleureau, C. Chamaret, and P. Guillotel. Calibration-free gaze tracking using particle filter. In *Multimedia and Expo (ICME), 2013 IEEE International Conference on*, pages 1–6, July 2013. 23, 37

[183] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. On the performance of human visual system based image quality assessment metric using wavelet domain. In *SPIE Conference Human Vision and Electronic Imaging XIII*, 2008. 149, 154

[184] A. Ninassi, O. Le Meur, P. Le Callet, and D. Barba. Which semi-local visual masking model for wavelet based image quality metric? In *Image Processing, 15th IEEE International Conference on*, pages 1180–1183, 2008. 153, 154

[185] Alexandre Ninassi, Olivier Le Meur, Patrick Le Callet, and Dominique Barba. On the performance of human visual system based image quality assessment metric using wavelet domain. In *SPIE Conference Human*

Vision and Electronic Imaging XIII, volume 6806, pages 680610–1, 2008. 19

[186] Masashi Nishiyama, Takahiro Okabe, Imari Sato, and Yoichi Sato. Aesthetic quality classification of photographs based on color harmony. In *CVPR'11*, page 33–40, 2011. 58, 149, 160

[187] Yaqing Niu, Rebecca M Todd, and Adam K Anderson. Affective salience can reverse the effects of stimulus-driven salience on eye movements in complex scenes. *Frontiers in psychology*, 3, 2012. 75, 76

[188] David Noton and Lawrence Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968):308–311, 1971. 11

[189] David Noton and Lawrence Stark. Scanpaths in eye movements during pattern perception. *Science*, 171(3968):308 –311, 1971. 26

[190] David Noton and Lawrence Stark. Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11(9):929 – IN8, 1971. 26

[191] Lauri Nummenmaa, Jukka Hyönä, and Manuel G Calvo. Emotional scene content drives the saccade generation system reflexively. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2):305, 2009. 72

[192] P. Obrador. Region based image appeal metric for consumer photos. In *Multimedia Signal Processing, 2008 IEEE 10th Workshop on*, pages 696–701, Oct 2008. 77

[193] Zena O'Connor. Colour harmony revisited. *Color Research & Application*, 35(4):267–273, 2010. iii, 2, 52

[194] Peter O'Donovan, Aseem Agarwala, and Aaron Hertzmann. Color compatibility from large datasets. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, pages 63:1–63:12, New York, NY, USA, 2011. ACM. 47, 65, 122

[195] L. C. Ou and M. Ronnier Luo. A colour harmony model for two-colour combinations. *Color Research & Application*, 31(3):191–204, 2006. 48, 49, 50, 68, 88, 102, 105, 121, 123, 181

[196] Li-Chen Ou, M Ronnier Luo, Andrée Woodcock, and Angela Wright. A study of colour emotion and colour preference. part i: Colour emotions for single colours. *Color Research & Application*, 29(3):232–240, 2004. 63

[197] Li-Chen Ou, M Ronnier Luo, Andrée Woodcock, and Angela Wright. A study of colour emotion and colour preference. part iii: Colour preference modeling. *Color Research & Application*, 29(5):381–389, 2004. 62

[198] Li-Chen OU, Ming Ronnier LUO, and Guihua CUI. A colour design tool based on empirical studies. In *Undisciplined! Design Research Society Conference 2008*, July 2008. 163, 164

[199] Sonia Ouni, Ezzeddine Zagrouba, and Majed Chambah. A new no-reference method for color image quality assessment. *International Journal of Computer Applications*, 40(17):24–31, February 2012. Published by Foundation of Computer Science, New York, USA. 149

[200] Vuilleumier P. How brains beware: neural mechanisms of emotional attention. *Trends Cogn Sci.*, 12(9):585 – 94, 2005. 76, 82

[201] Yi Pan. Attentional capture by working memory contents. *Canadian journal of experimental psychology*, 64(2):124–8, 2010. 62

[202] S. Pannasch, J.R. Helmert, K. Roth, A.K. Herbold, and H. Walter. Visual fixation durations and saccade amplitudes: Shifting relationship in a variety of conditions. *Journal of Eye Movement Research*, 4(2):1–19, 2008. 79

[203] Niebur E. Parkhurst D, Law K. Modeling the role of salience in the allocation of overt visual attention. *"Vision Research"*, 42(1):107–23, 2002. 74, 81

[204] Jeff B. Pelz and Roxanne Canosa. Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, 41(25-26):3587–96, 2001. 69

[205] Robert J. Peters, Asha Iyer, Christof Koch, and Laurent Itti. Components of bottom-up gaze allocation in natural scenes. *Journal of Vision*, 5(8):692, 2005. 35

[206] Robert J. Peters, Asha Iyer, Christof Koch, and Laurent Itti. Components of bottom-up gaze allocation in natural scenes. *Journal of Vision*, 5(8):692, 2005. 94

[207] MatthewS. Peterson, ArthurF. Kramer, and DavidE. Irwin. Covert shifts of attention precede involuntary eye movements. *Perception & Psychophysics*, 66(3):398–405, 2004. 10

[208] François Pitié, Anil C. Kokaram, and Rozenn Dahyot. Automated colour grading using colour distribution transfer. *Comput. Vis. Image Underst.*, 107(1-2):123–137, July 2007. 168

[209] M Pomplun, H Ritter, and B Velichkovsky. Disambiguating complex visual information: Towards communication of personal views of a scene. *Perception*, 25(8):931–948, 1996. 28

[210] M. I. Posner, C. R. Snyder, and B. J. Davidson. Attention and the detection of signals. *Journal of experimental psychology*, 109(2):160–174, jun 1980. vi, 10, 11, 70

[211] Michael I Posner. Orienting of attention. *Quarterly journal of experimental psychology*, 32(1):3–25, 1980. 11

[212] Tania Pouli, Ronan Boitard, Christel Chamaret, Mekides Abebe, Catherine Serré, David Touzé, Edouard François, and Erik Reinhard. Hdr in the living room. In *ACM SIGGRAPH 2014 Studio*, page 5. ACM, 2014. 6

[213] Tania Pouli and Erik Reinhard. Progressive color transfer for images of arbitrary dynamic range. *Computers & Graphics*, 35(1):67 – 80, 2011. Extended Papers from Non-Photorealistic Animation and Rendering (NPAR) 2010. 168

[214] Tania Pouli, Erik Reinhard, and Douglas W Cunningham. *Image Statistics in Visual Computing*. CRC Press, 2013. 91

[215] D. Radha, J. Amudha, P. Ramyasree, R. Ravindran, and S. Shalini. Detection of unauthorized human entity in surveillance video. *International Journal of Engineering and Technology*, 5(3), 2013. 16

[216] Justus J Randolph. *Free-Marginal Multirater Kappa (multirater κ_free): An Alternative to Fleiss ' Fixed-Marginal Multirater Kappa*. Paper presented at the \textitJoensuu University Learning and Instruction Symposium 2005, October 2005. 65, 140

[217] Keith Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, pages 372–422, 1998. 25, 27

[218] Marilyn A Read and Deborah Upington. Young children's color preferences in the interior environment. *Early Childhood Education Journal*, 36(6):491–496, 2009. 63

[219] Malcolm James Ree. Applied power analysis for the behavioral sciences by christopher aberson. *Personnel Psychology*, 65(3):716–717, 2012. 31

[220] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *Computer Graphics and Applications, IEEE*, 21(5):34–41, Sep 2001. 168

[221] Erik Reinhard, Edouard François, Ronan Boitard, Christel Chamaret, Catherine Serré, and Tania Pouli. High dynamic range video chains. *International Broadcasting Convention (IBC), Amsterdam*, 2014. 6

[222] Erik Reinhard, Edouard François, Ronan Boitard, Christel Chamaret, Catherine Serré, and Tania Pouli. High dynamic range video production, delivery and rendering. *Motion Imaging Journal, SMPTE*, 124(4):1–8, 2015. 6

[223] Erik Reinhard, Peter Shirley, Michael Ashikhmin, and Tom Troscianko. Second order image statistics in computer graphics. In *Proceedings of the 1st Symposium on Applied Perception in Graphics and Visualization*, APGV '04, page 99–106, New York, NY, USA, 2004. ACM. 91

[224] Austin Roorda and David R. Williams. The arrangement of the three cone classes in the living human eye. *Nature*, 397(6719):520–522, 1999. 62

[225] Frank E. Saal, Ronald G. Downey, and Mary A. Lahey. Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, 88(2):413–428, sept 1980. 65

[226] Dario D Salvucci and Joseph H Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 symposium on Eye tracking research & applications*, pages 71–78. ACM, 2000. 27

[227] Catherine Sauvaget and Vincent Boyer. Harmonic colorization using proportion contrast. In *Proceedings of the 7th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, AFRIGRAPH '10, pages 63–69, New York, NY, USA, 2010. ACM. vi, 42, 46, 47, 57, 181

[228] N. Sawant and N. J Mitra. Color harmonization for videos. In *Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, page 576–582, 2008. 56

[229] Nikhil Sawant and Niloy J. Mitra. Color harmonization for videos. In *Proceedings of the 2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*, ICVGIP '08, pages 576–582, Washington, DC, USA, 2008. IEEE Computer Society. 57, 133

[230] KB Schloss and SE Palmer. Aesthetic response to color combinations: preference, harmony, and similarity. *Atten Percept Psychophys.*, 73(2):551–71, 2011. 63, 64, 68, 88, 102, 117

[231] Gerald E. Schneider. Two visual systems. *Science*, 163(3870):895–902, 1969. 13

[232] H. Senane, A. Saadane, and D. Barba. Image coding in the context of a psychovisual image representation with vector quantization. In *Image Processing, 1995. Proceedings., International Conference on*, volume 1, pages 97–100 vol.1, Oct 1995. 17

[233] Yunyu Shi, Youdong Ding, Ranran Zhang, and Jun Li. Structure and hue similarity for color image quality assessment. In *Electronic Computer Technology, 2009 International Conference on*, pages 329–333, 2009. 149

[234] Lars Sivik and Anders Hård. Some reflections on studying colour combinations. *Color Research & Application*, 19(4):286–295, 1994. 39, 43, 48, 51

[235] Przemyslaw Skurowski and Michal Kozielski. Investigating human color harmony preferences using unsupervised machine learning. In *CGIV*, pages 59–64, 2012. 47, 48, 65

[236] N. Smirnov. Table for estimating the goodness of fit of empirical distributions. *Ann. Math. Statist.*, 19(2):279–281, 06 1948. 93

[237] Tim J. Smith and Parag K. Mital. Attentional synchrony and the influence of viewing task on gaze behavior in static and dynamic scenes. *Journal of Vision*, 13(8), 2013. 30

[238] Robert J Snowden. Visual attention to color: Parvocellular guidance of attentional resources? *Psychological Science*, 13(2):180–184, 2002. 1, 70, 115

[239] M. Solli and R. Lenz. Color harmony for image indexing. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1885 –1892, October 2009. 50, 59, 66, 105, 123, 160, 181

[240] Kalyan Sunkavalli, Micah K. Johnson, Wojciech Matusik, and Hanspeter Pfister. Multi-scale image harmonization. *ACM Trans. Graph.*, 29(4):125:1–125:10, July 2010. 55

[241] Ferenc Szabo, Peter Bodrogi, and János Schanda. Experimental modeling of colour harmony. *Color Research & Application*, 35(1):34–49, 2010. 50, 65, 68, 88, 105, 121

[242] P.D. Tafti and Xiaolin Wu. Automatic tonal harmonization for multispectral mosaics. In *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, pages 1876–1879, Oct 2008. 55

[243] Z. Tang, Z. Miao, and Y. Wan. Image composition with color harmonization. In *Image and Vision Computing New Zealand (IVCNZ), 2010 25th International Conference of*, page 1 –8, November 2010. 56, 57, 127, 128, 147

[244] Z. Tang, Z. Miao, Y. Wan, and F. F Jesse. Colour harmonisation for images and videos via two-level graph cut. *Image Processing, IET*, 5(7):630 –643, October 2011. 133

[245] Z. Tang, Z. Miao, Y. Wan, and F.F. Jesse. Color harmonization for images and videos via two-level graph cut. *IET Image Processing*, 5:630–643(13), October 2011. 56, 57

[246] Benjamin W. Tatler, Roland J. Baddeley, and Benjamin T. Vincent. The long and the short of it: spatial statistics at fixation vary with saccade amplitude and task. *Vision Research*, 46(12):1857–62, June 2006. 69, 95, 117, 118

[247] Benjamin W Tatler and Benjamin T Vincent. Systematic tendencies in scene viewing. *Journal of Eye Movement Research*, 2(2):1–18, 2008. 81

[248] Benjamin W. Tatler, Nicholas J Wade, Hoi Kwan, John M Findlay, and Boris M Velichkovsky. Yarbus, eye movements, and vision. *i-Perception*, 1:7–27, 2010. iii, 3, 11, 69

[249] BW Tatler, RJ Baddeley, and ID Gilchrist. Visual correlates of fixation selection effects of scale and time. *Vision Research*, 45:643 – 659, 2005. 35, 70, 71, 74, 81, 114

[250] Niveditta Thakur and Swapna Devi. A new method for color image quality assessment. *International Journal of Computer Applications*, 15(2):10–17, February 2011. Published by Foundation of Computer Science. 149

[251] Jan Theeuwes. Abrupt luminance change pops out; abrupt color change does not. *Perception & psychophysics*, 57(5):637–644, 1995. 70

[252] M. Tokumaru, N. Muranaka, and S. Imanishi. Color design support system considering color harmony. In *Fuzzy Systems, 2002. FUZZ-IEEE'02. Proceedings of the 2002 IEEE International Conference on*, volume 1, page 378 –383, 2002. 45, 46, 53, 56, 65, 121, 150

[253] Antonio Torralba, Aude Oliva, Monica S Castelhano, and John M Henderson. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological review*, 113(4):766, 2006. 36

[254] Anne M. Treisman and Garry Gelade. A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97 – 136, 1980. 10, 15, 70, 71

[255] P. Unema, S. Pannasch, M. Joos, and B. M. Velichkovsky. Time-course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition*, 12(3):473–494, 2005. 79, 80, 96

[256] Fabrice Urban, Brice Follet, Christel Chamaret, Olivier Le Meur, and Thierry Baccino. Medium spatial frequencies, a strong predictor of salience. *Cognitive Computation*, 3(1):37–47, 2011. vii, 6, 18, 19, 27, 81, 83

[257] A.H.C. van der Heijden. Visual information processing and selection. In Odmar Neumann and Wolfgang Prinz, editors, *Relationships Between Perception and Action*, pages 203–226. Springer Berlin Heidelberg, 1990. 10

[258] Artem Violentyev, Shinsuke Shimojo, and Ladan Shams. Touch-induced visual illusion. *Neuroreport*, 16(10):1107–1110, 2005. 8

[259] Thomas Wachtler, Terrence J Sejnowski, and Thomas D Albright. Representation of color stimuli in awake macaque primary visual cortex. *Neuron*, 37(4):681–691, 2003. 62

[260] Yanli Wan, Zhen Tang, Zhenjiang Miao, and Bo Li. Image composition with color harmonization. *IJPRAI*, 26(3), 2012. 56

[261] Baoyuan Wang, Yizhou Yu, and Ying-Qing Xu. Example-based image color and tone style enhancement. In *ACM SIGGRAPH 2011 Papers*, SIGGRAPH '11, pages 64:1–64:12, New York, NY, USA, 2011. ACM. 55

[262] Lujin Wang and Klaus Mueller. Harmonic colormaps for volume visualization. In *Proceedings of the Fifth Eurographics / IEEE VGTC Conference on Point-Based Graphics*, SPBG'08, pages 33–39, Aire-la-Ville, Switzerland, Switzerland, 2008. Eurographics Association. 56, 57

[263] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *Image Processing, IEEE Transactions on*, 13(4):600–612, 2004. 149

[264] Zhou Wang, Eero P. Simoncelli, and Alan C. Bovik. Multi-scale structural similarity for image quality assessment. In *IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 1398–1402, 2003. 154

[265] Andrew B. Watson, Robert Borthwick, and Mathias Taylor. *Proc. SPIE 3016, Human Vision and Electronic Imaging*, pages 2–12, 1997. 17, 153

[266] Michel Wedel and Rik Pieters. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing science*, 19(4):297–312, 2000. 23

[267] Stephen Westland, Kevin Laycock, Vien Cheung, Phil Henry, and Forough Mahyar. Colour harmony. *Colour: Design & Creativity*, 1(1):1–15, 2007. ii, 2, 40

[268] Felix A. Wichmann, Lindsay T. Sharpe, and Karl R. Gegenfurtner. The contributions of color to recognition memory for natural scenes. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 28:509–520, 2002. 62

[269] Florian Wickelmaier and Christian Schmid. A matlab function to estimate choice model parameters from paired-comparison. *Behav Res Methods Instrum Comput.*, 36(1):29–40, 2004. 139

[270] Stefan Winkler. A perceptual distortion metric for digital color video. In *in Proc. SPIE*, pages 175–184, 1999. 149

[271] Harry K Wolfe. On the color-vocabulary of children. 1890. 1

[272] DavidS. Wooding. Eye movements of large populations: II. deriving regions of interest, coverage, and similarity using fixation maps. *Behavior Research Methods, Instruments, & Computers*, 34(4):518–528, 2002. 28, 29

[273] Xuezhong Xiao and Lizhuang Ma. Gradient-preserving color transfer. *Computer Graphics Forum*, 28(7):1879–1886, 2009. 168

[274] Chun-Yu Yang, Hsin-Ho Yeh, and Chu-Song Chen. Video aesthetic quality assessment by combining semantically independent and dependent features. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1165–1168, 2011. 149

[275] A. L. Yarbus. *Eye movements and vision*. Springer, 1967. iii, 3, 11, 26, 30, 69