



HAL
open science

Méthodes robustes pour l'estimation d'illuminants et la prise en compte de la couleur en comparaison d'images

Baptiste Mazin

► **To cite this version:**

Baptiste Mazin. Méthodes robustes pour l'estimation d'illuminants et la prise en compte de la couleur en comparaison d'images. Vision par ordinateur et reconnaissance de formes [cs.CV]. Télécom Paris-Tech, 2014. Français. NNT : 2014ENST0016 . tel-01547852

HAL Id: tel-01547852

<https://pastel.hal.science/tel-01547852>

Submitted on 27 Jun 2017

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

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

Doctorat ParisTech THÈSE

pour obtenir le grade de docteur délivré par

Télécom ParisTech
Spécialité : **Signal et Images**

Présentée et soutenue publiquement par

Baptiste Mazin

le 28 Mars 2014

Méthodes robustes pour l'estimation d'illuminants et la prise en compte de la couleur en comparaison d'images

Jury :

Mme. Agnès Desolneux, Professeur, MAP5, Université Paris Descartes

M. Patrick Pérez, Chercheur, Technicolor

M. Pascal Monasse, Chercheur, IMAGINE, Ecole des Ponts ParisTech

M. Theo Gevers, Professeur, Université d'Amsterdam

M. Julien Rabin, Maître de conférence, GREYC, Université de Caen

M. Edoardo Provenzi, Chercheur postdoctoral, LTCl, Télécom ParisTech

M. Yann Gousseau, Professeur, LTCl, Télécom ParisTech

Mme. Julie Delon, Professeur, MAP5, Université Paris-Descartes

Présidente

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Directeur

Directeur

THÈSE



Table des matières

Table des matières	I
1 Introduction	1
1.1 Problèmes considérés et contributions	2
1.1.1 Estimation de la couleur des illuminants présents dans une scène	2
1.1.2 Descripteurs locaux couleur pour la comparaison d'images	3
1.1.3 Statistiques couleur des images numériques	3
1.1.4 Autres contributions	4
1.2 Organisation du manuscrit	4
1.2.1 Découpage	4
1.2.2 Liste de publications	5
2 Couleur et illumination	7
2.1 Mesures et représentations de la couleur	7
2.1.1 Formation de l'image couleur	7
2.1.2 Systèmes colorimétriques de la CIE	9
2.1.3 Systèmes perceptuels	12
2.1.4 Caractérisation d'un espace RGB : primaires, point blanc	13
2.1.5 Espaces dérivés des espaces RGB	15
2.2 Illuminants : observations et modélisations	16
2.2.1 Radiateur de Planck	16
2.2.2 Température de couleur proximale (Correlated Color Temperature)	17
2.2.3 Lumière naturelle	18
2.2.4 Modélisation de Von Kries pour le changement d'illuminant	20
3 Estimer l'illuminant	21
3.1 État de l'art	21
3.1.1 Illuminant unique	21
3.1.2 Estimation d'illuminants multiples	23
3.1.3 Contributions	24
3.2 Choisir l'illuminant le plus probable	24
3.3 Raffinement de l'algorithme par une sélection de modes	27
3.4 Expériences	29

3.4.1	Espaces de couleur	29
3.4.2	Estimation mono-illuminant	30
3.4.3	Estimation multi-illuminant	35
3.5	Conclusion	40
4	Couleur et comparaison d'images : état de l'art	47
4.1	Attributs couleur invariants	47
4.1.1	Espaces de couleur avec propriétés d'invariance	49
4.1.2	Invariants pixéliques et leurs liens avec la représentation de la couleur	50
4.2	Descriptions globales d'images basées sur des statistiques couleur	50
4.2.1	Représentations normalisées des couleurs de l'image	51
4.2.2	Statistiques simples et distributions de couleurs	51
4.3	Description locale d'images	53
4.3.1	La description SIFT et ses variantes	53
4.3.2	Variante SIFT utilisée	55
4.3.3	Descripteurs SIFT combinés à des statistiques couleur	56
4.3.4	Descripteurs SIFT appliqués sur des canaux couleur	57
4.3.5	Autres approches	58
5	Descripteurs locaux couleur : choix et appariement	61
5.1	Un nouveau descripteur de teinte	61
5.2	Descripteurs reposant sur l'espace couleur d'Ohta	64
5.3	Validation d'appariements <i>a contrario</i>	67
5.3.1	Critères existants	67
5.3.2	Critère <i>a contrario</i>	69
5.3.3	Regroupement des distances pour le respect de \mathcal{H}_0	71
6	Expérimentations sur les descripteurs locaux couleur	75
6.1	Création de la vérité terrain	75
6.1.1	Expériences synthétiques	75
6.1.2	Expériences en conditions réelles	76
6.1.3	Validation d'une correspondance	79
6.2	Mesures des performances	79
6.3	Terminologie	80
6.4	Validation des descripteurs couleur	81
6.4.1	Robustesse aux changements géométriques	81
6.4.2	Robustesse au changement d'illuminant	85
6.4.3	Combinaison de descripteurs	88
6.4.4	Autres résultats	90
6.4.5	Conclusions	91
6.5	Validation du critère <i>a contrario</i>	92
6.5.1	Résultats de classification	93
7	Statistiques des images naturelles couleur	97
7.1	Travaux connexes	97
7.1.1	Redondance spatiale	97
7.1.2	Redondance chromatique	98
7.1.3	Redondance spatio-chromatique	98

7.2	Présentation de la base d'images	99
7.3	Distributions couleur dans les photographies	100
7.3.1	Distributions de valeurs RGB	100
7.3.2	Liens entre espaces antagonistes et fonctions de sensibilité	103
7.4	Analyse en composantes principales spatio-chromatiques	104
7.5	Conclusion	104
8	Conclusion	117
8.1	Conclusion	117
8.2	Perspectives	118
A	Détection de modes dans les histogrammes pondérés	119
A.1	Modèle de fond pour les histogrammes non pondérés	119
A.2	Modèle de fond pour les histogrammes pondérés	120
A.3	Significativité d'un intervalle et intervalles maximum significatifs	122
A.4	Illustrations de la détection de modes dans les histogrammes pondérés	122
A.4.1	Segmentation d'histogrammes de valeurs de teinte pondérées	123
A.4.2	Application à la détection d'orientations principales	123
B	Modèle à deux paramètres pour le changement d'illuminant	127
B.1	Simplification du modèle de Von Kries pour l'adaptation chromatique	127
B.2	Observation des changements d'illumination de la base ALOI	129
	Bibliographie	133

Introduction

La couleur est une expérience quotidienne. Bien qu'elle nous soit extrêmement familière, son appréhension nécessite pourtant la compréhension des interactions entre la lumière, les matériaux et l'oeil ainsi que la compréhension du processus de perception en œuvre dans le cerveau humain. Résultant de ces interactions, l'étude de la couleur recoupe plusieurs disciplines, telles que la psychologie, les neurosciences, l'informatique, ou encore les mathématiques. La difficulté à maîtriser toutes ces disciplines dans leur ensemble peut expliquer pourquoi la couleur est trop souvent ignorée en traitement de l'image. Elle fait, en effet, souvent figure de parent pauvre de la vision par ordinateur alors même qu'elle est un élément essentiel de la vision humaine.

Cette constatation amène à se demander pourquoi la couleur est cantonnée à un rôle secondaire. Peut-être n'apporte-t-elle pas de plus value suffisante aux applications actuelles. Ou peut-être que ce sont les difficultés posées par son utilisation qui font qu'elle est négligée ? Si la perception de la couleur est inutile à la vision, pourquoi a-t-elle été conservée et perfectionnée par des centaines de milliers d'années d'évolution ? Cette thèse ne prétend pas, bien sûr, répondre à toutes ces questions. Néanmoins, elles en constituent les fondements. Trois axes vont être explorés en particulier.

La représentation robuste des couleurs La couleur de la source de lumière éclairant une scène (l'illuminant) induit de fortes variations indésirables dans les couleurs capturées des objets. Pour diminuer la sensibilité à ces variations des algorithmes de vision par ordinateur, il est possible de mesurer des quantités invariantes ou bien de normaliser les données. Une estimation des conditions d'illumination, et en particulier de la couleur de l'illuminant, permet d'obtenir une représentation canonique des couleurs de l'image en les normalisant.

La description des images couleur par descripteurs locaux Les descripteurs locaux sont des outils omniprésents en vision par ordinateur et ils constituent un bon exemple d'application dans laquelle la couleur est sous-exploitée. Ils sont une brique élémentaire de nombreuses applications, notamment pour la calibration des paramètres internes d'un appareil photographique, ou pour la reconnaissance d'objet.

Les statistiques des images couleur L'observation de statistiques sur les images naturelles constitue la base de la compréhension de la structure des images ainsi qu'un prérequis pour



FIG. 1.1 Différentes photographies du tableau « les noces de Cana » de Véronèse. Ces images proviennent du site de partage en ligne Flickr et ont été capturées par différents appareils photographiques. Elles illustrent la variabilité des couleurs mesurées pour une même scène. Les causes de ces variations sont les conditions d’illumination, les propriétés des surfaces, le capteur, ainsi que des traitements éventuels effectués après l’acquisition.

leur modélisation. Les images en niveaux de gris ont été largement étudiées mais relativement peu de travaux se sont intéressés aux statistiques couleur. Or ces statistiques sont très importantes pour la formation d’*a priori* utilisés dans de nombreuses applications, comme le débruitage ou encore la segmentation d’images.

1.1 Problèmes considérés et contributions

1.1.1 Estimation de la couleur des illuminants présents dans une scène

Les conditions d’acquisition affectent fortement les couleurs capturées pour une scène donnée (cf. Figure 1.1). En particulier, la couleur de la source lumineuse produit des variations très importantes dans les couleurs mesurées. Bien que la vision humaine possède des mécanismes très performants d’adaptation à ces changements, les processus mis en œuvre pour permettre cette aptitude ne sont pas pleinement compris. En effet, les trois hypothèses couramment utilisées, à savoir l’adaptation locale, l’adaptation à la moyenne spatiale de l’image ou l’adaptation à la région la plus lumineuse, ne sont pas suffisantes pour expliquer complètement ces mécanismes dans leur globalité [87]. La recherche d’une solution algorithmique performante pour la reproduction de cette capacité par un appareil photographique est donc encore un problème ouvert.

Contribution 1 : Une méthode automatique d’estimation de l’illuminant Nous présentons au début de cette thèse une méthode automatique pour l’estimation de la couleur de l’illuminant dans les photographies numériques. Cette méthode repose sur la recherche de surfaces grises. Par définition, ces surfaces renvoient un spectre lumineux ayant la même distribution fréquentielle que le spectre de l’illuminant. En supposant que l’ensemble des couleurs que peut prendre un illuminant est limité, il est possible de sélectionner les pixels appartenant à une surface potentiellement grise. Une méthode de vote est alors appliquée, permettant de sélectionner un ou plusieurs illuminants. Cette méthode est fondée sur l’*a priori* physique que la couleur de l’illuminant est proche de celle prédite par l’équation de Planck. L’algorithme

final possède de nombreux avantages : il est efficace, intuitif, ne nécessite pas de phase d'apprentissage, requiert peu de paramètres qui s'avèrent stables. De plus, la méthode de vote permet de s'adapter aux cas où plusieurs sources lumineuses de couleurs différentes éclairent la scène photographiée.

1.1.2 Descripteurs locaux couleur pour la comparaison d'images

Les descripteurs locaux sont des outils très puissants pour comparer finement des images. Cependant, ils font rarement appel à la couleur et de fait, le rôle de la couleur dans l'étape d'appariement a fait l'objet de trop peu d'études. Plusieurs études récentes ont montré l'intérêt de l'information couleur pour une tâche de reconnaissance ou de classification d'objet. Néanmoins, à notre connaissance, aucun travail ne porte sur l'apport de la couleur pour l'étape relativement bas-niveau d'appariement des descripteurs. Le problème principal que nous considérons ici est celui de l'apport de la couleur pour l'appariement de descripteurs locaux. Un usage local de cette dernière permet-il de désambiguïser certaines situations où la luminance seule est insuffisante? Et si oui, dans quelles proportions?

Contribution 2 : De nouveaux descripteurs locaux couleur Nous proposons quatre descripteurs permettant de décrire de manière détaillée le contexte local de points clés dans l'optique de l'appariement de ces derniers. L'idée maîtresse développée ici est qu'une information précise sur le contexte couleur des points clés ne peut être obtenue qu'en décrivant à la fois les distributions de couleurs (teinte) *et* les transitions entre couleurs (gradients de couleur). Remarquons qu'à l'opposé, l'information de distribution locale des niveaux de gris est à la fois peu discriminante et peu robuste. Elle n'est par conséquent pas utilisée pour l'appariement local de points clés entre images monochromes (seul le gradient est exploité en pratique). Les descripteurs proposés montrent clairement l'apport positif que peut produire la couleur pour la fiabilité de la mise en correspondance locale.

Contribution 3 : Un algorithme automatique de validation d'appariements Nous décrivons une méthode pour la comparaison et l'appariement automatique des descripteurs locaux décrits précédemment. L'idée de cette méthode est de tirer profit de l'indépendance supposée entre les informations de luminance et de couleur afin de maîtriser le nombre de fausses alarmes. Cette idée repose sur la méthodologie *a contrario* et en particulier sur les travaux de [33, 125].

1.1.3 Statistiques couleur des images numériques

Quelle est la distribution des couleurs dans le monde? Le vert des arbres et le bleu du ciel sont-elles les couleurs prédominantes? Le monde est-il vraiment gris en moyenne? Il est évidemment difficile de répondre à ces questions, mais il existe peut-être une distribution couleur représentative des photographies numériques.

Contribution 4 : Statistiques sur une grande quantité d'image Partant de ces simples questions, nous montrons des résultats préliminaires sur des statistiques mesurées à partir d'un grand nombre d'images couleur. La question posée est celle de la stabilité à laquelle il est possible de prétendre en effectuant des statistiques sur un nombre donné d'images. Nous analysons en particulier la distribution des couleurs et ses marginales, ainsi que les composantes principales obtenues à partir d'une ACP sur un ensemble d'images basse résolution.

Nous mettons en particulier l'accent sur l'influence du processus d'acquisition de la couleur sur ces statistiques.

1.1.4 Autres contributions

Cette thèse a également conduit à d'autres contributions. Nous décrivons notamment ici un algorithme de sélection de modes, utile pour l'analyse d'histogrammes, ainsi qu'un modèle paramétrique simple de changement de température de couleur.

Contribution 5 : Un algorithme de sélection de mode dans les histogrammes pondérés La problématique de la segmentation d'histogramme est un problème important en traitement des images. Un mode est une région d'un histogramme contenant une masse importante par rapport à la masse moyenne de l'histogramme. Un algorithme pour détecter ce type de structure a été proposé par Desolneux *et coll.* [33]. Cette approche fait l'hypothèse que tous les éléments constitutifs de l'histogramme y contribuent de manière identique. Par exemple, chaque pixel de l'image peut contribuer avec le même poids à la constitution d'un histogramme couleur. Pour de nombreux exemples, cette hypothèse est fautive et l'algorithme est alors inadapté. Nous nous intéressons ici au cas d'histogrammes construits à partir d'éléments contribuant avec des poids variables. En prenant en compte des hypothèses sur la distribution des poids des éléments, nous proposons une méthode d'extraction des modes pour les histogrammes pondérés.

Contribution 6 : Un modèle approché de changement de température de couleur Nous établissons ici un modèle simple, simulant pour une image un changement de la température de couleur de l'illuminant. Pour décrire ce type de transformation, le modèle de Von Kries est généralement utilisé [151]. Le modèle proposé à la fin de cette thèse est une simplification du modèle de Von Kries basée sur l'utilisation de l'équation de Planck. Cette simplification aboutit à la description d'un changement d'illuminant à l'aide de deux paramètres : un paramètre d'intensité et un paramètre de couleur. Bien qu'un modèle proche ait été utilisé par Gevers *et coll.* [60], ce modèle très simple n'a jamais été étudié précisément dans la littérature. Une validation expérimentale de ce modèle est présentée.

1.2 Organisation du manuscrit

1.2.1 Découpage

Le Chapitre 2 propose une brève introduction à la colorimétrie ainsi qu'à la modélisation des sources lumineuses. Il introduit les concepts clés nécessaires à la lecture des chapitres suivants. Dans le Chapitre 3, la méthode d'estimation automatique de l'illuminant que nous avons développée (Contribution 1) est décrite. Un état de l'art sur l'estimation mono et multi-illuminant est présenté. Il est suivi par la description puis par l'évaluation de l'algorithme.

Le Chapitre 4, dresse l'état de l'art des descriptions couleur globales et locales pour la comparaison d'images. Le Chapitre 5 présente les descripteurs locaux couleur développés durant cette thèse ainsi que la procédure associée de validation des correspondances (Contributions 2 et 3). Leurs évaluations respectives font l'objet du chapitre 6.

Le chapitre 7 présente des résultats préliminaires sur les statistiques couleur réalisées sur un grand nombre d'images.

L'annexe A décrit l'algorithme de détection de mode dans les histogrammes pondérés. Et l'annexe B présente le modèle simplifié de changement de température de couleur.

1.2.2 Liste de publications

Cette thèse a conduit à la publication des articles suivants.

Publication de journal :

- **Estimation of illuminants from projections on the Planckian locus**, B. Mazin, J. Delon, Y. Gousseau, (Prépublication, soumise à IEEE I.P.)¹

Publications dans des conférences internationales :

- **Illuminant estimation from projection on the Planckian locus**, B. Mazin, J. Delon, Y. Gousseau, Color and Photometry in Computer Vision, Workshop ECCV 2012,
- **Combining color and geometry for local image matching**, B. Mazin, J. Delon, Y. Gousseau, ICPR, 2012.

Publications dans des conférences francophones :

- **Estimation Automatique d'Illuminants Multiples**, B. Mazin, J. Delon, Y. Gousseau, GRETSI, 2013,
- **Utilisation jointe de la couleur et de la géométrie pour la comparaison locale d'images**, B. Mazin, J. Delon, Y. Gousseau, ORASIS, 2011,

1. Une version peut-être trouvée à l'adresse <http://hal.archives-ouvertes.fr/hal-00915853>.

Couleur et illumination

Ce chapitre présente des notions sur la mesure de la couleur qui seront utiles tout au long de la thèse. En particulier, nous rappelons différentes modélisations du processus d'acquisition d'une image couleur par un ensemble de capteurs. Nous décrivons ensuite les principes élémentaires de la colorimétrie, dont l'objectif est l'obtention d'une représentation de la couleur permettant une reproduction précise de celle-ci. Cela nous amènera à présenter les espaces de couleurs les plus utilisés. Dans une seconde partie, nous décrivons la modélisation des sources lumineuses avec une attention particulière pour l'équation de Planck ainsi que pour la lumière dite naturelle.

2.1 Mesures et représentations de la couleur

Au début du XIX^e siècle, Young a été le premier à émettre l'hypothèse que l'oeil humain ne possédait que trois types de photorécepteurs. Cette hypothèse fut confirmée en 1855 par Maxwell qui démontra la possibilité de synthétiser une couleur quelconque par une combinaison de trois sources de lumière bien choisies. Ces trois sources de lumière sont appelées primaires et ce sont elles qui définissent les caractéristiques d'un système de mesure de la couleur.

2.1.1 Formation de l'image couleur

Différents modèles d'acquisition de la lumière existent dans la littérature. Ces modèles ont pour objectif de décrire les interactions entre une source lumineuse, les objets qui nous entourent et un système d'acquisition. Les modèles lambertien et dichromatique présentés ici sont les plus utilisés dans la littérature. La différence entre ces deux modèles est que le modèle dichromatique prend en compte les réflexions spéculaires alors que le modèle lambertien les ignore.

Modèle lambertien

Une source de lumière L est caractérisée par son spectre d'émission noté $L(\lambda)$, λ représentant la longueur d'onde. Cette lumière interagit avec une surface S dont le spectre de réflectance est noté $S(\lambda)$. Dans le modèle lambertien, le spectre $E(\lambda, \mathbf{m})$ atteignant le capteur au pixel \mathbf{m} résulte de l'interaction entre la lumière émise par la source et la surface, ce qui donne :

$$E(\lambda, \mathbf{m}) = L(\lambda, \mathbf{m})S(\lambda, \mathbf{m}). \quad (2.1)$$

Précisons que le spectre $L(\cdot, \mathbf{m})$ dépend du point \mathbf{m} car l'intensité de la source lumineuse peut varier localement. De plus, plusieurs sources lumineuses peuvent se mélanger ou éclairer différentes parties de la scène, provoquant une variation de la distribution spectrale atteignant chaque pixel. Lorsque la lumière E , réfléchiée par la surface S , atteint le capteur, elle est filtrée par les fonctions de sensibilité du capteur. Bien que certains appareils permettent d'échantillonner le spectre $E(\lambda, \mathbf{m})$ sur plus de trois bandes de fréquences (on parle alors d'imagerie multispectral), nous considérons ici le cas d'un appareil photographique classique. Les fonctions de sensibilité sont alors au nombre de trois, chacune correspondant à une des composantes RGB. Notons $\rho(\lambda) = (\rho_R(\lambda), \rho_G(\lambda), \rho_B(\lambda))$ ces fonctions et Λ un domaine de définition. Ce domaine de définition est en général l'ensemble visible du spectre électromagnétique qui s'étend de 360nm à 830nm. Les composantes du vecteur $\mathbf{p}(\mathbf{m}) = (p_R(\mathbf{m}), p_G(\mathbf{m}), p_B(\mathbf{m}))$ constitué des valeurs mesurées en un pixel \mathbf{m} de l'appareil sont alors

$$p_c(\mathbf{m}) = \int_{\Lambda} E(\lambda, \mathbf{m}) \rho_c(\lambda) d\lambda, \quad c \in \{R, G, B\} \quad (2.2)$$

Ce modèle est couramment utilisé comme première approximation pour la formation des images couleur. Néanmoins, il ne prend pas en compte une partie importante des réflexions : les réflexions spéculaires. Le modèle dichromatique, lui, permet de prendre en compte ces dernières.

Modèle dichromatique [134]

Les réflexions spéculaires correspondent physiquement à la réflexion directe de la source lumineuse, sans autres interactions avec la surface. Dans ce cas, le spectre $E(\lambda, \mathbf{m})$ réfléchi par la surface S est donné par

$$E(\lambda, \mathbf{m}) = L(\lambda, \mathbf{m})[\alpha_l(\mathbf{m})S(\lambda, \mathbf{m}) + \alpha_s(\mathbf{m})]. \quad (2.3)$$

Les coefficients $\alpha_l(\mathbf{m})$ et $\alpha_s(\mathbf{m})$ permettent de pondérer l'influence entre les composantes spéculaires et lambertiennes. Le paramètre $\alpha_s(\mathbf{m})$ est ici considéré indépendant de la longueur d'onde, ce qui implique qu'un reflet spéculaire possède le même spectre que la source lumineuse. La plupart des matériaux vérifient en pratique cette propriété [83]. Ces coefficients dépendent de la géométrie de la scène et, en particulier, des angles formés par la lumière incidente et la normale à la surface. La valeur $\mathbf{p}(\mathbf{m})$ peut alors s'écrire comme la somme de deux composantes

$$p_c(\mathbf{m}) = \underbrace{\alpha_l(\mathbf{m}) \int_{\Lambda} L(\lambda) S(\lambda, \mathbf{m}) \rho_c(\lambda) d\lambda}_{\text{Composante lambertienne}} + \underbrace{\alpha_s(\mathbf{m}) \int_{\Lambda} L(\lambda) \rho_c(\lambda) d\lambda}_{\text{Composante spéculaire}}, \quad c \in \{R, G, B\}. \quad (2.4)$$

Le modèle de Kubelka-Monk [57] est un autre modèle utilisé dans la littérature. De manière semblable au modèle dichromatique, il permet de prendre en compte les réflexions spéculaires. Dans le modèle, les coefficients $\alpha_l(\mathbf{m})$ et $\alpha_s(\mathbf{m})$ sont liés par la relation $\alpha_l(\mathbf{m}) = (1 - \alpha_s(\mathbf{m}))^2$. La réflectance de la surface S peut être modélisée comme une fonction d'un coefficient d'absorption $k(\lambda)$ et d'un coefficient de diffusion $s(\lambda)$ tous deux propres à chaque surface. Ce modèle permet de tenir compte du cas des surfaces non opaques, c'est-à-dire transmettant de la lumière.

2.1.2 Systèmes colorimétriques de la CIE

Colorimétrie Wyszecki et Stiles [155] donnent la définition suivante de la colorimétrie :

- « La colorimétrie est la branche des sciences de la couleur concernée par la spécification numérique de la couleur d'un stimulus visuel de manière à ce que :
- Deux stimulus, possédant les mêmes spécifications, vus sous des conditions identiques par un observateur ayant une vision normale, soient semblables.
 - Deux stimulus semblables aient les mêmes spécifications.
 - Les nombres caractérisant la spécification soient des fonctions continues de paramètres physiques définissant la distribution du spectre du stimulus. ».

Dans le but de définir un système de référence, la CIE (Commission Internationale de l'Éclairage) a défini en 1931 l'observateur de référence qui est communément appelé observateur standard. Cet observateur établit un lien entre l'espace des couleurs perceptibles par l'œil humain et l'ensemble des couleurs qu'il est possible de créer à l'aide de trois sources lumineuses monochromatiques. Ces trois sources $[R]$, $[G]$ et $[B]$ sont d'intensité variable et sont caractérisées par leurs longueurs d'onde respectives de 435nm pour le bleu, 546nm pour le vert et 700nm pour le rouge.

Ce lien est défini par trois fonctions colorimétriques issues d'expériences d'égalisation. Dans ces expériences, un utilisateur est soumis à une lumière monochromatique requête d'une longueur d'onde λ_r et il doit chercher à égaliser ce stimulus $[S_{\lambda_r}]$ par une combinaison des trois sources monochromatiques $[R]$, $[G]$ et $[B]$ (aussi appelées primaires). On cherche donc \bar{r}_{λ_r} , \bar{g}_{λ_r} et \bar{b}_{λ_r} telles que :

$$[S_{\lambda_r}] = \bar{r}_{\lambda_r}[R] + \bar{g}_{\lambda_r}[G] + \bar{b}_{\lambda_r}[B] \quad (2.5)$$

Comme une source polychromatique peut-être vue comme la superposition de plusieurs sources monochromatiques, le stimulus RGB associé à une lumière polychromatique $s(\lambda)$ vue par un être humain, peut être reproduit par la combinaison linéaire des coordonnées associées à la somme de chaque stimulus monochromatique, soit

$$R_s = \int_{\Lambda} \bar{r}(\lambda)s(\lambda)d\lambda, \quad G_s = \int_{\Lambda} \bar{g}(\lambda)s(\lambda)d\lambda, \quad B_s = \int_{\Lambda} \bar{b}(\lambda)s(\lambda)d\lambda. \quad (2.6)$$

Ce système est défini par les fonctions colorimétriques $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ et $\bar{b}(\lambda)$, qui sont obtenues en faisant parcourir au stimulus requête $[S_{\lambda_r}]$ l'ensemble des longueurs d'onde du visible Λ . Ces fonctions colorimétriques sont illustrées par la Figure 2.1(a). Les valeurs correspondant aux sources $[R]$, $[G]$ et $[B]$ sont visibles sur les courbes, car elles correspondent à l'annulation de deux des trois fonctions colorimétriques. En définissant une équivalence entre un spectre et un stimulus visuel, un tel système fournit un moyen de spécifier précisément une couleur et de la reproduire fidèlement. Cependant, des valeurs négatives sont présentes dans chacune des trois courbes et cela est particulièrement visible pour la courbe $\bar{r}(\lambda)$. Ces valeurs négatives s'expliquent par le fait que les fonctions de sensibilité de l'œil se recouvrent. À cause de ce recouvrement ainsi que du choix des primaires servant à l'égalisation, chacune des primaires excite au moins deux types de photorécepteur. Dans le cas où le stimulus n'en excite qu'un seul, l'égalisation est donc impossible.

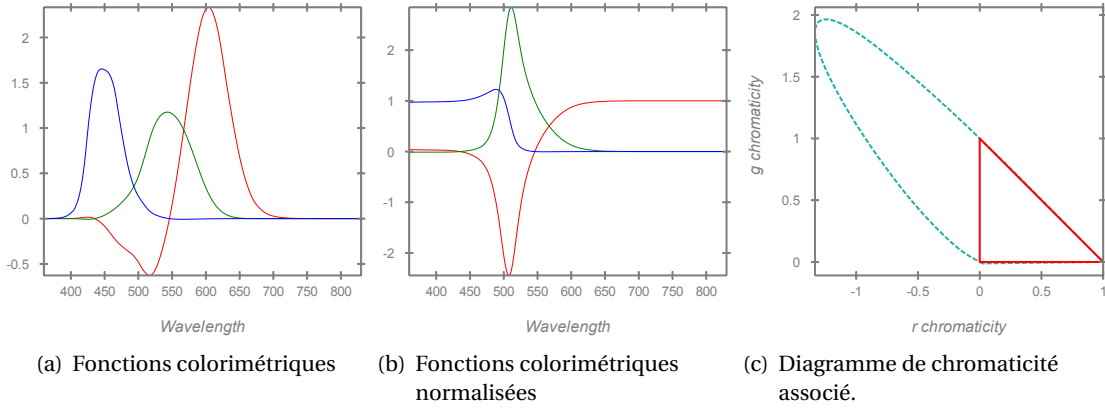


FIG. 2.1 Fonctions colorimétriques définies par la CIE en 1931 : $\bar{r}(\lambda)$ (en rouge), $\bar{g}(\lambda)$ (en vert) et $\bar{b}(\lambda)$ (en bleu), ainsi que leurs versions normalisées $r(\lambda)$, $g(\lambda)$ et $b(\lambda)$ telles que $r(\lambda) + g(\lambda) + b(\lambda) = 1$. La figure de droite représente le diagramme de chromaticité issu des primaires mesurées. La courbe pointillée représente le lieu du spectre tandis que le triangle rouge donne le gamut de l'espace CIE RGB 1931.

L'astuce consiste à « désaturer » le stimulus que l'on souhaite reproduire, en lui « soustrayant » une des primaires. En pratique, cela revient à ajouter une des primaires servant à l'égalisation, du côté du stimulus à égaliser, pour parvenir au résultat. Par exemple

$$[S_{\lambda_r}] + \bar{r}_{\lambda_r}[R] = \bar{g}_{\lambda_r}[G] + \bar{b}_{\lambda_r}[B]. \quad (2.7)$$

Espaces de chromaticité Les expériences d'égalisation visuelle confirment que la sensation colorée produite par une lumière quelconque, peut être reproduite par une combinaison de trois couleurs primaires bien choisies. Une fois ces primaires définies, l'Équation (2.6) permet d'associer à chaque spectre un triplet (R, G, B) . L'information chromatique, qui est relative à la distribution fréquentielle du spectre, réside dans les valeurs relatives de ce triplet tandis que l'information d'intensité est donnée par leur somme. Il est possible de définir des coordonnées chromatiques (r, g, b) (également appelées chromaticités) associées au triplet de valeurs (R, G, B) en normalisant ces composantes par l'intensité

$$r = \frac{R}{R+G+B} \quad g = \frac{G}{R+G+B} \quad b = \frac{B}{R+G+B}. \quad (2.8)$$

La somme de ces chromaticités vérifie donc $r + g + b = 1$. Cette équation définit un plan dans l'espace RGB. L'intersection de ce plan avec le cube RGB forme un triangle appelé le triangle de Maxwell. Pour se repérer dans ce plan, deux coordonnées sont suffisantes. Par convention, on utilise les coordonnées (r, g) . En appliquant cette normalisation aux fonctions colorimétriques $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ et $\bar{b}(\lambda)$ on obtient

$$\begin{cases} r(\lambda) = \frac{\bar{r}(\lambda)}{\bar{r}(\lambda) + \bar{b}(\lambda) + \bar{g}(\lambda)} \\ g(\lambda) = \frac{\bar{g}(\lambda)}{\bar{r}(\lambda) + \bar{b}(\lambda) + \bar{g}(\lambda)} \\ b(\lambda) = \frac{\bar{b}(\lambda)}{\bar{r}(\lambda) + \bar{b}(\lambda) + \bar{g}(\lambda)} \end{cases}, \quad \lambda \in \Lambda. \quad (2.9)$$

Lieu du spectre et gamut La courbe paramétrée par les fonctions $\bar{r}(\lambda)$ et $\bar{g}(\lambda)$, tracée en pointillés sur la Figure 2.1(c), donne les coordonnées chromatiques de chaque spectre monochromatique dans le diagramme de chromaticité et définit le domaine de l'ensemble des couleurs visibles par un œil humain. Sur cette même figure, le triangle rouge délimite l'ensemble des couleurs qu'il est possible de reproduire à l'aide d'une combinaison positive des trois primaires $[R]$, $[G]$ et $[B]$. Cet ensemble est appelé gamut. Les extrémités de ce triangle correspondent aux coordonnées chromatiques de chacune des primaires.

Système de primaires XYZ Le système de primaires CIE RGB, défini en 1931, a l'inconvénient de générer des valeurs négatives. Pour remédier à cela, un changement de primaire a simultanément été proposé. Ce changement correspond à une combinaison linéaire des fonctions colorimétriques $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ et $\bar{b}(\lambda)$. Ces nouvelles primaires, nommées $[X]$, $[Y]$ et $[Z]$, sont obtenues à partir des primaires $[R]$, $[G]$ et $[B]$ présentées précédemment par le système suivant :

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (2.10)$$

Les fonctions colorimétriques correspondantes sont illustrées sur la Figure 2.2(a). Plusieurs raisons ont guidé le choix de cette transformation. Les deux principales sont l'absence de valeurs négatives et la représentation par la primaire $[Y]$ de l'information d'intensité visuelle [155].

Ce système est toujours très largement utilisé pour la spécification de couleur. Un diagramme de chromaticité est défini de manière similaire à celui de l'espace CIE RGB 1931. Ce diagramme, appelé xy , est obtenu en normalisant les coordonnées X , Y et Z par leur somme $X + Y + Z$. Les fonctions colorimétriques de cet espace ainsi que le diagramme de chromaticité associé sont présentés sur la Figure 2.2.

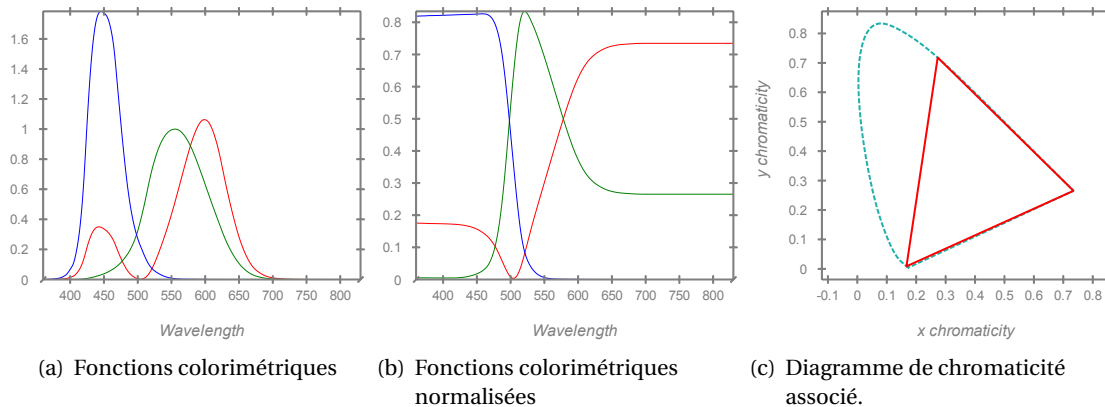


FIG. 2.2 Le système de primaires XYZ : la fonction colorimétrique $\bar{x}(\lambda)$ associée à la primaire X est tracée en rouge, $\bar{y}(\lambda)$ correspondant à la primaire Y est tracée en vert et $\bar{z}(\lambda)$ la primaire Z en bleu. Les versions normalisées $x(\lambda)$, $y(\lambda)$ et $z(\lambda)$ de ces fonctions sont tracées sur le diagramme du milieu. Le diagramme de chromaticité associé à cet espace est tracé sur la figure de droite. En pointillé est représenté le lieu du spectre tandis que le triangle rouge délimite l'ensemble des couleurs qu'il est possible de reproduire par une combinaison linéaire positive des sources monochromatiques utilisées pour définir l'espace CIE RGB 1931.

2.1.3 Systèmes perceptuels

Le but originel de la standardisation des systèmes de colorimétrie est la reproduction fidèle des couleurs. Cette reproduction précise requiert la mesure des erreurs. Or, si les systèmes décrits précédemment permettent un repérage précis des couleurs, ils ne sont pas adaptés à la mesure des différences entre les couleurs. En effet, la mesure de distance dans ces espaces ne reflète pas du tout la sensibilité aux variations de couleur du système visuel humain. Ainsi, le seuil à partir duquel il est possible de différencier deux couleurs, dépend très fortement de ces couleurs. Différents systèmes permettant des mesures de distances dites perceptuelles, ont été proposés. Actuellement, les systèmes CIELAB et CIELUV sont recommandés par la CIE et sont en conséquence les principaux utilisés.

CIELUV Les travaux les plus connus sur la mesure des seuils différentiels sont ceux de MacAdam [95]. L'expérience consiste à déterminer, pour 25 centres de couleur, la distance minimale pour laquelle une différence est décelable par un observateur humain. Ces travaux ont mené aux célèbres ellipses de MacAdam. Ces ellipses, une fois placées dans le diagramme de chromaticité xy , varient fortement suivant leur localisation. Dans le but de transformer ces ellipses en cercles de mêmes diamètres, plusieurs transformations projectives ont été proposées. Seule cette classe de transformations a été étudiée car elle permet de conserver les propriétés d'un diagramme de chromaticité (par exemple, le mélange de deux couleurs décrit une droite). Cependant, aucune transformation de ce type ne permet d'atteindre parfaitement le but recherché, et donc, plusieurs possibilités ont été retenues. Un premier diagramme fut proposé par MacAdam en 1960. Ce diagramme, considéré aujourd'hui comme obsolète, est toujours utilisé pour déterminer une caractéristique liée aux sources lumineuses, la température proximale de couleur (cf. 2.2.2). Les équations suivantes permettent de passer des coordonnées chromatiques xy aux coordonnées CIE uv 1960

$$u = \frac{4x}{-2x + 12y + 3} \quad v = \frac{6y}{-2x + 12y + 3}. \quad (2.11)$$

Il peut paraître étrange qu'il existe une relation paramétrique, permettant de passer des coordonnées xy à des coordonnées « perceptuelles ». Cette relation est bien entendu une approximation.

Ce diagramme a été remplacé en 1976 lorsque la CIE a défini l'espace chromatique uniforme $L^*u^*v^*$. Les Équations (2.12) définissent l'espace CIELUV

$$\begin{cases} L^* = 116(Y/Y_n)^{1/3} - 16 \\ u^* = 13L^*(u' - u'_n) \\ v^* = 13L^*(v' - v'_n) \end{cases} \quad (2.12)$$

avec,

$$\begin{cases} u' = \frac{4X}{X+15Y+3Z} = \frac{4x}{-2x+12y+3} = u \\ v' = \frac{9Y}{X+15Y+3Z} = \frac{9y}{-2x+12y+3} = \frac{3}{2}v \end{cases} .$$

Les variables Y_n , u'_n et v'_n représentent respectivement l'intensité et les coordonnées chromatiques de l'illuminant. Les diagrammes associés aux coordonnées chromatiques uv et $u'v'$ sont tracés sur la Figure 2.3.

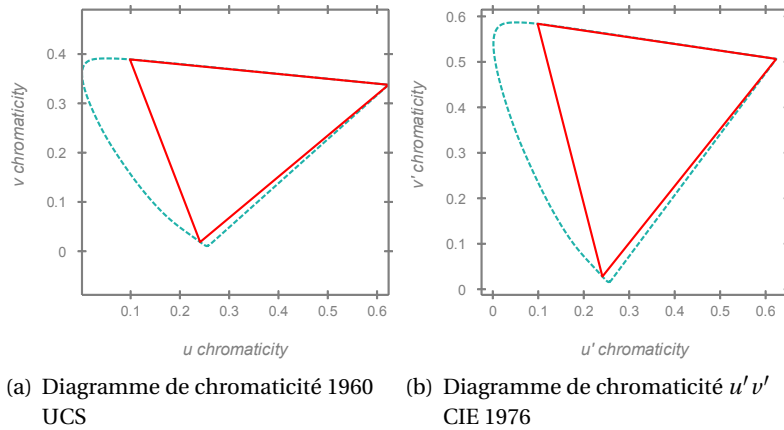


FIG. 2.3 Différents espaces de chromaticités uniformes définis par la CIE.

CIELAB En même temps que l'espace CIELUV, la CIE a proposé l'espace CIELAB ou espace $L^*a^*b^*$ CIE 1976.

$$\begin{cases} L^* = 116(Y/Y_n)^{1/3} - 16 \\ a^* = 500 \left[\left(\frac{X}{X_n} \right)^{1/3} - \left(\frac{Y}{Y_n} \right)^{1/3} \right] \\ b^* = 200 \left[\left(\frac{Y}{Y_n} \right)^{1/3} - \left(\frac{Z}{Z_n} \right)^{1/3} \right] \end{cases} \quad (2.13)$$

où X_n, Y_n et Z_n représentent les coordonnées de l'illuminant dans l'espace XYZ. Cet espace possède des propriétés similaires à celles de l'espace CIELUV mais a l'inconvénient de ne pas posséder un diagramme de chromaticité (cela est dû aux transformations non linéaires utilisées pour calculer les coordonnées a^* et b^*). Bien que l'espace CIELAB soit aujourd'hui de loin le plus utilisé des deux, aucun de ces espaces ne permet une meilleure uniformité que l'autre. La norme utilisée pour comparer des couleurs dans ces espaces est la norme euclidienne.

Il est intéressant de remarquer que ces espaces sont faits pour être utilisés sur des mesures effectuées en environnement contrôlé. En particulier, ils nécessitent la connaissance de l'illuminant pour être précis. S'ils s'avèrent très utiles pour des applications spécifiques dans des conditions connues, l'avantage lié à l'utilisation de ces espaces dans des applications génériques en vision par ordinateur est plus incertain.

2.1.4 Caractérisation d'un espace RGB : primaires, point blanc

Comme nous l'avons vu précédemment, un système d'acquisition RGB est caractérisé par ses fonctions de sensibilité. Ainsi, pour deux capteurs possédant des fonctions de sensibilité différentes, des valeurs RGB identiques ne refléteront pas nécessairement une distribution spectrale identique. Inversement, les valeurs RGB mesurées par ces deux capteurs seront différentes pour une surface donnée. Pour pallier ce problème, les espaces de la CIE associent à chaque distribution spectrale des coordonnées liées à celles que le système visuel humain attribue à cette distribution. Ces espaces constituent ainsi des espaces de référence pour décrire les couleurs. Parmi eux, le plus utilisé est l'espace CIE XYZ 1931.

Définition des primaires En considérant les fonctions colorimétriques $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ et $\bar{z}(\lambda)$, les coordonnées associées à une distribution spectrale $s(\lambda)$ sont données par

$$X_s = \int_{\Lambda} \bar{x}(\lambda)s(\lambda)d\lambda, \quad Y_s = \int_{\Lambda} \bar{y}(\lambda)s(\lambda)d\lambda, \quad Z_s = \int_{\Lambda} \bar{z}(\lambda)s(\lambda)d\lambda. \quad (2.14)$$

À chacune des fonctions de sensibilité d'un appareil photo numérique correspond un triplet de valeurs XYZ. Ces trois triplets définissent les trois primaires de l'appareil considéré. Le gamut de l'appareil, c'est-à-dire le sous-espace des couleurs qu'il peut reproduire, est défini par l'ensemble des combinaisons linéaires de ces primaires. Les primaires définissent les « couleurs » de chacun des récepteurs de l'appareil photo. Le point blanc de l'appareil est une couleur de référence obtenue lorsque les trois primaires sont combinées dans une certaine proportion, par exemple $[1, 1, 1]^T$. La transformation projetant les coordonnées de l'espace RGB de l'appareil vers un espace de référence (ici XYZ) s'obtient en déterminant le vecteur $[k_R, k_G, k_B]^T$ dans le système suivant.

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \end{bmatrix} = \begin{bmatrix} X_R & X_G & X_B \\ Y_R & Y_G & Y_B \\ Z_R & Z_G & Z_B \end{bmatrix} \begin{bmatrix} k_R \\ k_G \\ k_B \end{bmatrix} \quad (2.15)$$

Ces variables contrôlent le mélange des trois primaires correspondant au point blanc. Cela permet de compenser les différences de sensibilité entre les différents capteurs $[R]$, $[G]$, et $[B]$.

Cas de l'espace sRGB L'espace sRGB est souvent utilisé comme espace RGB de référence. En tant qu'espace RGB, il est caractérisé par un gamut et un point blanc. Son gamut est défini par les coordonnées chromatiques de 3 primaires et son point blanc par les coordonnées chromatiques de la source lumineuse utilisée comme référence. Ces coordonnées, le plus souvent fournies dans l'espace de chromaticité CIE xy 1931, sont présentées dans la Table 2.1. La composante d'intensité de l'illuminant est arbitrairement fixée à un. Pour les primaires, les composantes d'intensité permettent déterminer les paramètres du vecteur $[k_R, k_G, k_B]^T$. et de fixer les proportions de primaires permettant d'égaliser le point blanc. Cet ensemble de valeurs

Composante	R	G	B	Point blanc
x	0.6400	0.3000	0.1500	0.3127
y	0.3300	0.6000	0.0600	0.3290
Y	0.2126	0.7153	0.0721	1.0000

TAB. 2.1 Primaires et point blanc de l'espace sRGB.

permet de définir la matrice de conversion de l'espace sRGB vers l'espace XYZ donnée par l'Équation (2.16).

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4123 & 0.3577 & 0.1802 \\ 0.2126 & 0.7153 & 0.0721 \\ 0.0193 & 0.1192 & 0.9493 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (2.16)$$

L'espace sRGB n'est pas défini uniquement par ses primaires et son point blanc, il sous entend également qu'une correction gamma identique est appliquée à chacun des canaux RGB.

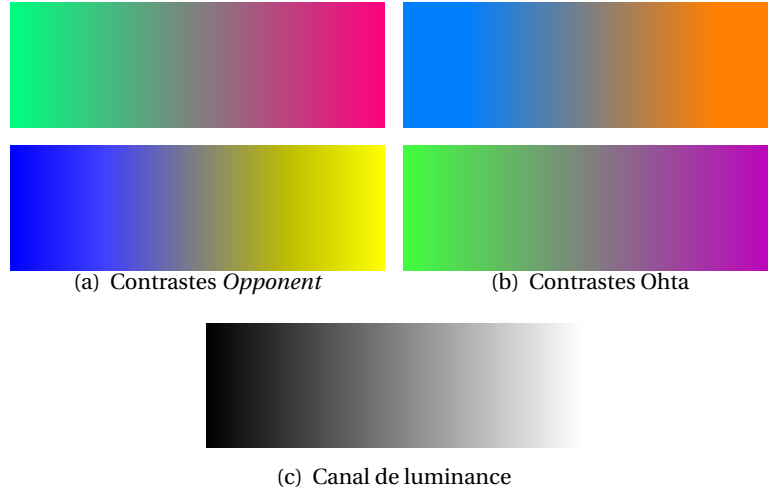


FIG. 2.4 Comparaison entre les contrastes couleur des espaces antagonistes *Opponent* et *Ohta*.

2.1.5 Espaces dérivés des espaces RGB

Espaces antagonistes [155] Ce type d'espace représente l'information de l'espace RGB sous la forme d'une composante d'intensité O_3 et de deux composantes chromatiques O_1 et O_2 exprimant des contrastes de couleur

$$O_1 = \frac{1}{\sqrt{2}}(R - G), \quad O_2 = \frac{1}{\sqrt{6}}(R + G - 2B), \quad O_3 = \frac{1}{\sqrt{3}}(R + G + B). \quad (2.17)$$

La composante O_1 correspond à un contraste rouge/vert et la composante O_2 à un contraste jaune/bleu. La vision humaine, par exemple, utilise une représentation de ce type [16]. Dans la suite, nous ferons référence à cette représentation précise par la dénomination « espace antagoniste *Opponent* ». Une autre représentation antagoniste est fréquemment utilisée, il s'agit de l'espace d'Ohta [117]. Il a été obtenu à partir des valeurs RGB moyennes de 100 régions, résultant de la segmentation de 8 images. Ces valeurs RGB sont prises comme base pour une analyse en composantes principales. La transformation qui décorrèle ces valeurs est la suivante :

$$Oh_1 = \frac{1}{\sqrt{2}}(R - B), \quad Oh_2 = \frac{1}{\sqrt{6}}(R + B - 2G), \quad Oh_3 = \frac{1}{\sqrt{3}}(R + G + B). \quad (2.18)$$

Dans la suite, nous nous référerons à cet espace par « espace antagoniste Ohta ». La Figure 2.4 illustre les différences entre les deux espaces antagonistes.

Espaces de coordonnées cylindriques [69, 138] Les espaces à coordonnées cylindriques représentent la couleur sous forme polaire. Parmi les deux composantes chromatiques, l'une décrit la teinte et l'autre la saturation. Du point de vue spectral, la teinte peut être interprétée comme une longueur d'onde moyenne, et la saturation comme une information sur la variance autour de cette longueur d'onde, une forte saturation signifiant une faible variance. La teinte H peut être définie par l'angle formé par les canaux antagonistes tandis que la saturation est définie par le module de ces canaux, soit

$$H = \arctan\left(\frac{O_1}{O_2}\right), \quad S = \sqrt{O_1^2 + O_2^2}, \quad L = O_3. \quad (2.19)$$

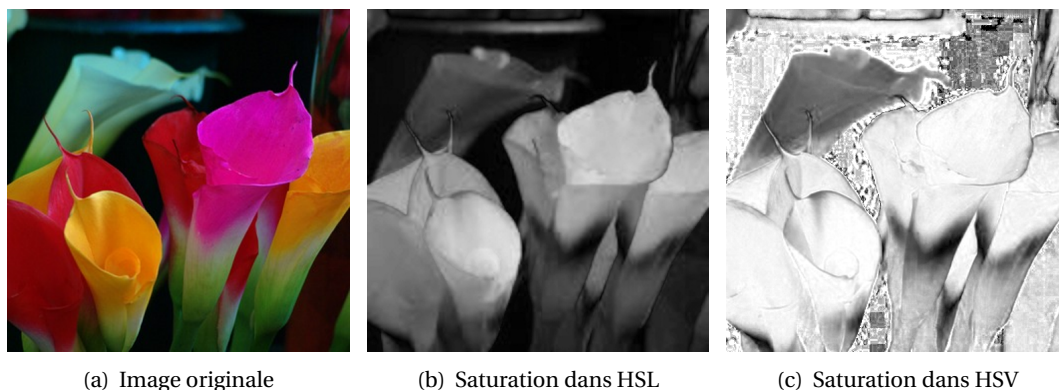


FIG. 2.5 Deux définitions différentes de la saturation.

Définitions de la saturation Différentes définitions de la saturation existent. Par exemple, Handbury *et coll.* [69] proposent comme définition $S = \max(R, G, B) - \min(R, G, B)$. Cette définition, largement utilisée, permet d'obtenir une distribution plus régulière et symétrique dans le cas d'une image de bruit blanc uniforme. Les deux définitions précédentes possèdent une dynamique variable proportionnelle à l'intensité. Cependant, d'autres définitions éliminent cette dépendance. Par exemple, dans HSV, la saturation est normalisée par l'intensité soit $S_{HSV} = 1 - \frac{\min(R, G, B)}{\max(R, G, B)}$. Cette définition introduit une singularité en zéro, ce qui produit des artefacts qui sont illustrés par la Figure 2.5.

2.2 Illuminants : observations et modélisations

Nous appelons illuminant une source de rayonnement électromagnétique. En pratique, on parle de source primaire lorsqu'elle transforme de l'énergie en rayonnement électromagnétique, et on parle de source à incandescence lorsque l'énergie est transformée en chaleur. Un illuminant peut avoir un spectre discontinu quand seulement certaines longueurs d'onde sont représentées, ou continu lorsque toutes les longueurs d'onde sont présentes, ou encore à spectre mixte lorsqu'une superposition des deux cas précédents est observée. Le rayonnement visible constitue la base de l'étude de la couleur. On définit généralement la partie visible du spectre électromagnétique par des longueurs d'onde comprises entre 360nm et 830nm. Cependant, les limites de 380nm et 780nm sont plus fréquemment utilisées en colorimétrie. L'étude théorique du rayonnement électromagnétique des corps chauffés ainsi que la mesure de divers spectres de la lumière du jour, ont permis de définir des illuminants standards et de caractériser de manière simple un grand nombre de sources lumineuses.

2.2.1 Radiateur de Planck

Au début du xx^e siècle, Planck formula certains principes et certaines lois liés aux rayonnements électromagnétiques diffusés par un corps chauffé. Ces lois stipulent qu'un corps dont la température est supérieure au zéro absolu émet un rayonnement dont le spectre est continu. Elles indiquent également que pour une température et une longueur d'onde fixée, le spectre émis est borné par celui émis par un corps physique idéal : le radiateur de Planck.

Le radiateur de Planck (également appelé corps noir) est donc un corps physique qui émet un rayonnement d'énergie maximale pour une température donnée. La loi d'émission que suit ce corps est donnée par l'Équation (2.20)

$$L(T, \lambda) = c_1 \lambda^{-5} \left[\exp\left(\frac{c_2}{\lambda T}\right) - 1 \right]^{-1}, \quad (2.20)$$

où T est la température en kelvins, λ est la longueur d'onde et $c_1 = 3.74183 \times 10^{-16} \text{ Wm}^2$ et $c_2 = 1.4388 \times 10^{-2} \text{ mK}$ sont deux constantes.

Approximation En remarquant que lorsque $T \leq 10000\text{K}$ et $\lambda \leq 700 \times 10^{-9}\text{m}$, on a $\exp\left(\frac{c_2}{\lambda T}\right) \gg 1$, l'équation de Planck peut alors être simplifiée en

$$L(T, \lambda) = c_1 \lambda^{-5} \exp\left(-\frac{c_2}{\lambda T}\right). \quad (2.21)$$

Lieu de Planck Chacun de ces spectres, une fois échantillonné par les fonctions colorimétriques de l'observateur standard XYZ CIE 1931 [155], fournit un triplet de valeur XYZ. Les chromaticités xy associées à l'ensemble de ces triplets décrivent une courbe dans le diagramme de chromaticité. Cette courbe est appelée le lieu de Planck et elle est tracée sur la Figure 2.6(b). Une approximation de cette courbe peut être obtenue à l'aide d'une spline [84].

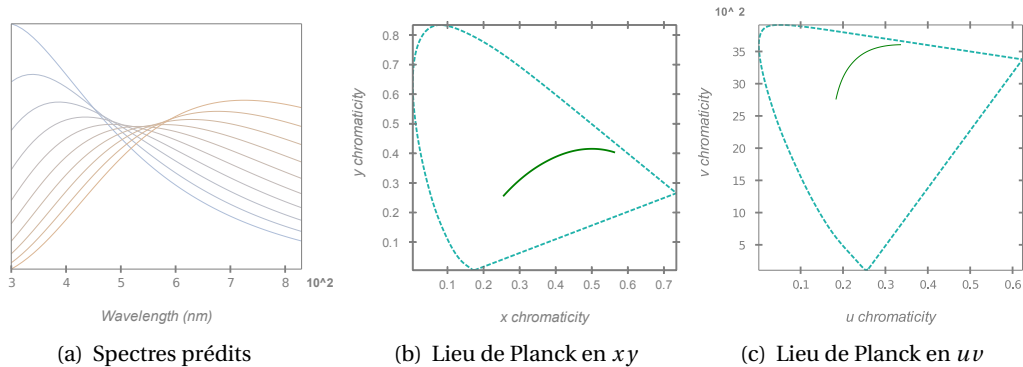


FIG. 2.6 Sur la figure de gauche, les spectres prédits par l'équation de Planck pour des températures allant de 4000K (orange) à 10000K (bleu). Au centre, le lieu de Planck est représenté en vert dans le diagramme de chromaticité xy . À droite, le lieu de Planck est représenté dans le diagramme de chromaticité uv .

2.2.2 Température de couleur proximale (Correlated Color Temperature)

L'équation de Planck prédit le spectre d'une source théorique. Dans le cas d'une source réelle, par exemple pour un matériau chauffé à une température donnée, le comportement diffère de celui prédit par cette équation. En effet, il a été remarqué [133] que l'intensité émise est généralement beaucoup plus faible que celle prédite. Quant à la distribution spectrale, elle correspond habituellement à celle d'un radiateur de Planck, mais chauffé cependant à une température différente. Par exemple, dans le cas des métaux, la température du radiateur de Planck possédant un spectre proche est nettement supérieure.

Cette remarque montre qu'il est possible de spécifier la distribution spectrale d'une source de lumière réelle par la température d'un radiateur de Planck dont la distribution spectrale est proche. Plusieurs définitions sont possibles pour préciser cette notion de proximité.

La définition de température de luminance énergétique repose sur la luminance. Elle est définie pour une longueur d'onde donnée par la température à laquelle un radiateur de Planck possède la même luminance. Cette définition donne le plus souvent un résultat différent selon la longueur d'onde considérée, rendant son utilisation compliquée. Une seconde définition possible repose sur la distribution spectrale. La température d'une source réelle est alors celle du radiateur de Planck dont le spectre minimise la distance quadratique au spectre de la source étudiée.

La dernière définition est la plus couramment utilisée. Elle repose sur la couleur de la source lumineuse. Elle est appelée température de couleur proximale (*Correlated Color Temperature* en anglais ou CCT en abrégé). La température correspond alors à celle d'un radiateur de Planck dont la chromaticité du rayonnement est la plus proche de celle de la source observée. Cette proximité chromatique est définie par la distance euclidienne entre les coordonnées chromatiques de la source et celles d'un point du lieu de Planck. La CIE préconise de calculer ces distances dans le diagramme de chromaticité uniforme CIE uv 1960 [155]. L'ensemble des points dont la projection orthogonale correspond au même radiateur de Planck forme une droite dans le diagramme de chromaticité. Ces droites sont appelées isothermes et sont visibles en bleu sur la Figure 2.7.

Approximation du calcul de la température de couleur proximale Le calcul des températures de couleur proximales est le seul usage qui est fait du diagramme uv . Plusieurs approximations de ce calcul ont été décrites dans la littérature. Bien que les projections soient définies dans le diagramme de chromaticité uv , les approximations sont généralement effectuées dans le diagramme xy . Ces approximations reposent sur l'idée que dans le diagramme de chromaticité CIE xy 1931, les droites isothermes sont quasiment concourantes en un point appelé foyer. Une approximation reposant sur ce principe a ainsi été proposé par [74]. La valeur de la température de couleur proximale pour un couple de chromaticité (x, y) est alors donnée par l'équation suivante

$$T(n) = A_0 + A_1 \exp(-n/t_1) + A_2 \exp(-n/t_2) + A_3 \exp(-n/t_3) \quad (2.22a)$$

$$n = \frac{x - x_e}{y - y_e} \quad (2.22b)$$

avec $x_e = 0.3366$, $y_e = 0.1735$ les coordonnées du foyer, $A_0 = -949.86315$, $A_1 = 6253.80338$, $A_2 = 28.70599$, $A_3 = 0.00004$ $t_1 = 0.92159$, $t_2 = 0.20039$ et $t_3 = 0.07125$ des constantes.

Échantillonnage du lieu de Planck Une différence de température de couleur n'est pas représentative de la différence de couleur perçue par un observateur. Pour établir un lien entre une différence de température de couleur et une différence perçue, Priest [122] a proposé l'échelle MIRE (Micro REciprocal Degree). Cette échelle est définie proportionnellement à l'inverse de la température de couleur soit $M = 10^6 / T$.

2.2.3 Lumière naturelle

L'étude de la lumière naturelle est essentielle à la caractérisation des sources de lumière puisqu'elle a été la seule source présente durant une grande partie de l'évolution humaine.

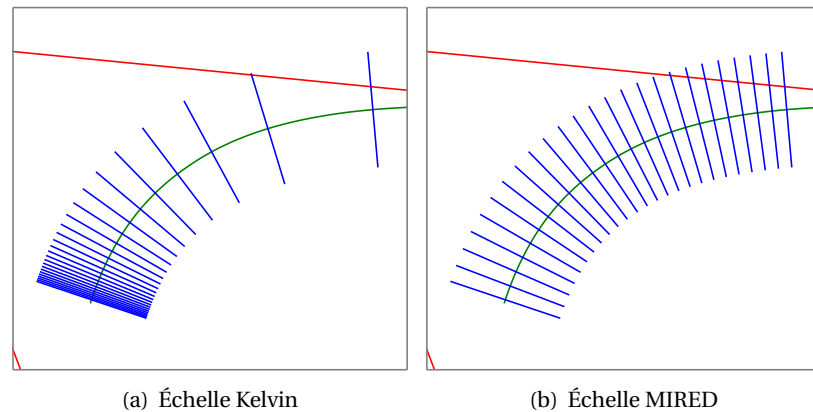


FIG. 2.7 Différence entre l'échelle proposée MIRED et l'échelle Kelvin dans le diagramme CIE uv 1960.

Le rayonnement solaire est différent de celui du radiateur de Planck en particulier en raison de l'absence de certaines fréquences dans son spectre. Il est de plus fortement modifié par l'atmosphère. Le rayonnement mesuré au sol est ainsi le résultat de l'addition de deux composantes : une composante directe et une composante diffuse. La composante directe varie en fonction des conditions météorologiques ainsi qu'en fonction de la hauteur du soleil dans le ciel. Lorsque le soleil descend dans le ciel, la lumière traverse une plus grande distance dans l'atmosphère qui filtre ainsi davantage la partie bleue du spectre. Lorsque le soleil est au plus haut, la température de couleur proximale associée est d'environ $5600K$ et peut descendre sous les $2000K$ lorsque le soleil se couche. La composante diffuse est, quant à elle, toujours présente et correspond à la partie diffractée par l'atmosphère. La température proximale de couleur de cette composante s'étale entre $15000K$ et $100000K$.

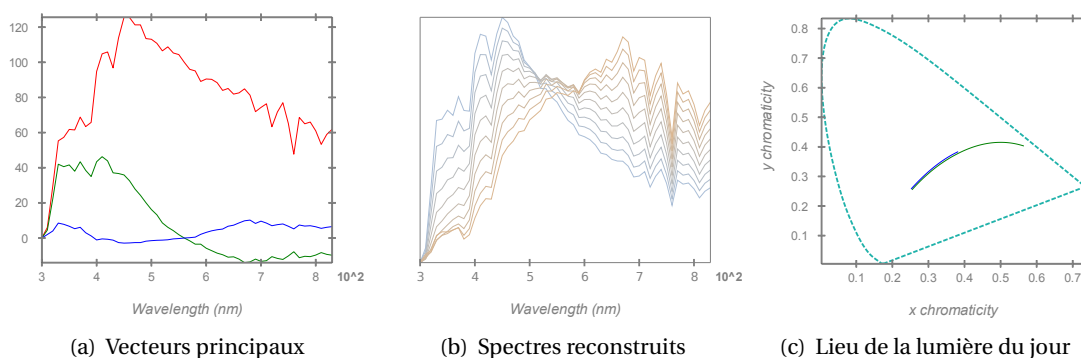


FIG. 2.8 Sur la figure de gauche, le vecteur moyen est représenté en rouge \bar{E} , le vecteur V_1 est en vert et le vecteur V_2 est en bleu. Sur la figure centrale, les spectres correspondent à une température de couleur proximale allant de $4000K$ (orange) à $10000K$ (bleu). Et sur la figure de droite, le lieu de la lumière du jour est représenté en bleu au dessus du lieu de Planck qui est en vert. Malgré des différences, il est clairement raisonnable de considérer le lieu de Planck comme une bonne description de la couleur des illuminants.

Modélisation Judd *et coll.* [82] ont remarqué que les coordonnées chromatiques (x, y) associées à 622 distributions spectrales de la lumière du jour sont regroupées autour de la courbe d'équation $y = 2.870x - 3.000x^2 - 0.275$ longeant le lieu de Planck (voir la courbe bleue sur la Figure 2.8(c)). À partir des trois premiers vecteurs \bar{E} , V_1 et V_2 issus de l'analyse en composantes principales de ces 622 distributions, ils reconstituent le spectre E_λ d'une phase quelconque de la lumière jour comme une combinaison linéaire de ces trois vecteurs, $E_\lambda = \bar{E} + M_1 V_{1,\lambda} + M_2 V_{2,\lambda}$. Les coefficients M_1 et M_2 sont alors fonction des coordonnées chromatiques de la source que l'on souhaite reconstituer. Ces coordonnées dépendent elles-mêmes d'une température de couleur proximale. L'illuminant D_{65} , qui est très souvent utilisé, correspond grossièrement au spectre moyen \bar{E} et son spectre est caractéristique d'un temps couvert.

2.2.4 Modélisation de Von Kries pour le changement d'illuminant

Le modèle de Von Kries suppose que les fonctions de sensibilité de l'appareil photographique numérique (APN) sont des fonctions de Dirac. Le modèle lambertien présenté par l'Équation (2.4) devient alors

$$p_c(\mathbf{m}) = \int_{\Lambda} L(\lambda) S(\lambda, \mathbf{m}) \delta(\lambda - \lambda_c) d\lambda, \quad c \in \{R, G, B\} \quad (2.23)$$

$$= L_{\lambda_c} S_{\lambda_c}(\mathbf{m}), \quad c \in \{R, G, B\} \quad (2.24)$$

Considérons alors deux mesures \mathbf{p}^1 et \mathbf{p}^2 d'une même surface \mathbf{s} éclairée sous deux illuminants différents \mathbf{l}^1 et \mathbf{l}^2 , on a

$$\begin{bmatrix} p_R^1 \\ p_G^1 \\ p_B^1 \end{bmatrix} = \begin{bmatrix} l_R^1 & 0 & 0 \\ 0 & l_G^1 & 0 \\ 0 & 0 & l_B^1 \end{bmatrix} \begin{bmatrix} s_R \\ s_G \\ s_B \end{bmatrix}, \quad \begin{bmatrix} p_R^2 \\ p_G^2 \\ p_B^2 \end{bmatrix} = \begin{bmatrix} l_R^2 & 0 & 0 \\ 0 & l_G^2 & 0 \\ 0 & 0 & l_B^2 \end{bmatrix} \begin{bmatrix} s_R \\ s_G \\ s_B \end{bmatrix}. \quad (2.25)$$

La relation liant \mathbf{p}^1 et \mathbf{p}^2 est alors

$$\begin{bmatrix} p_R^1 \\ p_G^1 \\ p_B^1 \end{bmatrix} = \begin{bmatrix} l_R^1/l_R^2 & 0 & 0 \\ 0 & l_G^1/l_G^2 & 0 \\ 0 & 0 & l_B^1/l_B^2 \end{bmatrix} \begin{bmatrix} p_R^2 \\ p_G^2 \\ p_B^2 \end{bmatrix}. \quad (2.26)$$

Cette relation constitue le modèle de Von Kries.

L'hypothèse de fonctions de Dirac pour les fonctions de sensibilité, et plus généralement l'hypothèse de capteurs à bandes étroites est fautive. En pratique, les fonctions de sensibilité des appareils photographiques se recouvrent fortement [81]. Malgré cela, cette relation est considérée comme étant une modélisation satisfaisante du changement d'illuminant [42]. De plus, il est possible de simuler un nouveau jeu de fonctions de sensibilité à l'aide d'une combinaison linéaire des fonctions de sensibilité de l'appareil photo [43]. Ce nouveau jeu respectant mieux l'hypothèse de Von Kries. Dans l'annexe B, nous proposons une version simplifiée du modèle de Von Kries reposant sur l'équation de Planck.

Pour la correction d'illuminant, il existe des méthodes plus élaborées. Elles reposent généralement sur l'idée que la correction doit être effectuée dans l'espace couleur de l'œil humain. Cette étape porte alors le nom d'adaptation chromatique. On peut en particulier citer le modèle d'apparence couleur CIECAM02 [111].

Estimer l'illuminant

Les couleurs des objets présents dans une scène dépendent fortement de la source de lumière. Malgré cela, les êtres humains perçoivent ces couleurs quasiment indépendamment de cette source. Cette capacité s'appelle la constance de couleur. À l'inverse, cette influence des illuminants constitue un facteur limitant pour les applications de vision par ordinateur telles que la reconnaissance d'objet ou la classification. L'objectif de la constance de couleur automatique est de réduire autant que possible les variations de couleurs dues à l'illumination, et de ce fait, de reproduire les couleurs indépendamment de la source lumineuse, imitant ainsi le système visuel humain. La capacité de produire une représentation de l'image aussi indépendante que possible de la source d'éclairage est également utile pour la photographie numérique, domaine dans lequel ce procédé est appelé *balance des blancs*. La première étape d'un processus de constance de couleur automatique consiste à estimer la couleur de l'illuminant, pour une scène donnée. Dans une seconde étape, et dans le cas où seul un illuminant est présent, l'image peut être corrigée à l'aide de l'estimation obtenue. Dans ce chapitre, nous nous concentrons sur l'étape d'estimation de l'illuminant. Nous nous intéressons aux situations dans lesquelles une ou plusieurs sources lumineuses de couleurs différentes sont impliquées.

3.1 État de l'art

3.1.1 Illuminant unique

Comme le montre une récente étude [63], l'estimation d'illuminants constitue un champ de recherche très actif depuis plusieurs décennies. Dans le cas d'une unique source de lumière éclairant la scène photographiée, les méthodes les plus simples reposent sur des *a priori* concernant des statistiques liées à la distribution des couleurs dans les images. Par exemple, une méthode très populaire consiste à supposer que la couleur moyenne d'une image acquise sous un illuminant neutre est la couleur grise. C'est sur cette hypothèse que repose l'algorithme dit de *monde gris* [16]. Dans cet algorithme, la couleur moyenne d'une image est supposée fixée par la couleur de l'illuminant. La compensation de l'influence de la source lumineuse est alors réalisée en corrigeant les couleurs de l'image de manière à ce que la couleur moyenne respecte l'*a priori* initial. Plusieurs variantes de cette approche ont été proposées, comme la restriction de l'hypothèse à un ensemble de surfaces bien choisies dans l'image, supposées grises [92].

Une autre hypothèse abondamment utilisée, est celle dite de *White-Patch* [89]. Elle suppose qu'une surface de réflectance parfaite est présente dans la scène et que cette surface

correspond aux points les plus lumineux de l'image. L'algorithme Max-RGB [89] résulte de cette hypothèse. L'illuminant est alors déduit à partir du niveau maximal de chaque canal RGB. Des variantes proposent d'effectuer un filtrage préliminaire de l'image [4]. Un modèle unifiant les hypothèses de *Monde-Gris* et de *White-Patch* a été proposé par Finlayson *et coll.* [50]. Plus récemment, ce même modèle a été étendu aux dérivées de premier et second ordre de l'image par les méthodes de type *Grey-Edge* [20, 66, 145].

Il est possible de s'appuyer sur le modèle de réflexion dichromatique [37] pour estimer la chromaticité de la source lumineuse. Ce modèle prend en compte la part spéculaire des réflexions en chaque pixel, en plus de la part lambertienne. Les pixels d'une surface monochromatique appartiennent alors à un plan dans l'espace de couleur RGB de l'appareil photographique ou à une droite dans un espace de chromaticité, voir [48]. Trouver l'intersection entre deux droites dans cet espace donne alors l'illuminant. Dans le cas où l'image ne contient qu'une seule surface, l'illuminant peut toujours être estimé en faisant l'hypothèse que la source de lumière appartient au lieu de Planck. L'illuminant est alors donné par l'intersection entre le lieu de Planck et la droite associée à la surface.

Des contraintes physiques sur l'ensemble des illuminants ainsi que sur le contenu de la scène ont également été proposées par Sapiro dans [132]. Plus précisément, il est supposé que le spectre de la réflectance ainsi que la distribution spectrale de l'illuminant peuvent être approchés par une combinaison linéaire d'un petit nombre de fonctions de base connues. Une procédure de vote est alors mise en oeuvre, dans laquelle chaque pixel vote pour un ensemble d'illuminants possibles. L'illuminant le plus probable est alors retrouvé à l'aide d'une transformée de Hough. Les pixels potentiellement gris peuvent aussi être sélectionnés comme étant proches du lieu de Planck¹ dans un diagramme de chromaticité [88] (cf. 2.1.2). Cette approche peut cependant échouer dès que plusieurs surfaces grises sont présentes dans la scène.

Des techniques plus élaborées s'appuient sur une phase d'apprentissage, dans laquelle des statistiques couleur sont apprises sur un ensemble d'images acquises sous une source de lumière connue. Parmi ces méthodes, les approches basées sur l'étude du gamut présument que l'ensemble des couleurs observables sous un illuminant donné est limité [51]. Ces approches nécessitent donc l'apprentissage de l'ensemble des couleurs observables sous un illuminant afin de trouver la meilleure correspondance possible entre les couleurs observées dans une scène donnée et cet ensemble. La correspondance trouvée permet de déduire la chromaticité de l'illuminant. Le problème de mise en correspondance de gamut reste néanmoins mal posé et des hypothèses additionnelles sont généralement nécessaires [40, 63]. Par exemple, les hypothèses peuvent concerner la diversité des couleurs dans la scène observée [51], ou encore l'ensemble possible des sources lumineuses [40]. Pour aller plus loin, certains auteurs explorent non seulement l'ensemble des couleurs possibles dans la scène observée mais également la distribution de ces couleurs sous un illuminant canonique [47, 127]. Cela permet de reformuler le problème d'estimation de l'illuminant en un problème de maximum de vraisemblance [47]. D'autres méthodes font également appel à un apprentissage simple de statistiques couleur [53].

Bien qu'offrant une estimation correcte dans un grand nombre de situations, toutes les approches mentionnées présentent également des cas d'échec. Une dernière catégorie de techniques d'estimation de l'illuminant propose de combiner la sortie de plusieurs de ces

1. Dans un espace de chromaticité, le lieu de Planck désigne l'ensemble possible des chromaticités d'un corps noir. De plus amples détails sont disponibles au paragraphe 2.2.1.

méthodes [10, 68, 149]. La stratégie pour obtenir une combinaison optimale est apprise sur des bases d'apprentissage, et repose sur des propriétés statistiques des images naturelles, ou sur de l'information sémantique plus évoluée. Récemment, Vasquez-Corral *et coll.* [150] ont proposé de raffiner les approches de constance de couleur en ajoutant des contraintes perceptuelles et en pondérant l'ensemble des illuminants "selon leur capacité à faire correspondre les couleurs de l'image corrigée à certaines couleurs", choisies selon des catégories de couleurs universelles.

3.1.2 Estimation d'illuminants multiples

Toutes les méthodes précédentes supposent la présence d'un illuminant unique et uniforme sur la scène. Une moindre attention a été portée aux cas où de multiples illuminants sont présents ainsi qu'aux cas d'illuminants non uniformes, bien qu'ils puissent être considérés comme plus réalistes en pratique. Les images prises en extérieur sont des exemples typiques. Elles contiennent des objets illuminés directement par la lumière du soleil tandis que d'autres, à l'ombre, sont éclairés par de la lumière réfléchiée par le ciel. Un autre exemple est celui d'images prises en intérieur avec une source de lumière incandescente ainsi qu'une source de lumière naturelle provenant d'une fenêtre.

La théorie Retinex proposée par Land et McCann [89] est l'un des premiers modèles abordant le problème de la constance de couleur locale. Pour distinguer un changement dû à une variation d'illumination d'un changement causé par une fluctuation de la réflectance, la méthode pose comme hypothèse que l'illuminant varie spatialement lentement : les faibles changements sont alors imputés à une variation d'illumination tandis que les forts le sont à des changements de la réflectance. De nombreux papiers reposent sur cette idée à laquelle ils incorporent des contraintes additionnelles, soit sur l'ensemble possible des illuminants [5, 44] (supposés être sur le lieu de Planck) soit sur l'ensemble des réflectances [5]. Ces approches [5, 44] exploitent également les variations d'illumination sur les surfaces mates pour contraindre la solution². Le travail de Finlayson [44] est étendu dans [85] pour les scènes extérieures, en s'appuyant sur les différences entre régions ombragées et non-ombragées.

Ebner [36] propose d'utiliser une version locale et itérative de l'algorithme de *Monde-Grise* pour calculer une estimation de l'illuminant en chaque pixel. Cet algorithme est affiné par [38], par l'utilisation d'une moyenne non uniforme des pixels locaux dans le but de tenir compte d'une variation brutale d'illumination. Plus récemment, les auteurs de [11] proposent d'appliquer différents algorithmes de constance de couleur sur des superpixels et de combiner leurs résultats. D'une manière similaire, Gijzen *et coll.* [67] ont évalué les algorithmes de constance de couleur sur des régions de l'image, sélectionnées selon différentes stratégies.

Quelques auteurs utilisent les reflets spéculaires [79] pour estimer la distribution spectrale de puissance par un dispositif d'imagerie spectrale.

Enfin, différentes méthodes guidées par l'utilisateur ont été proposées dans la littérature [14, 76]. La méthode décrite dans [76] permet une estimation, pour chaque pixel de l'image, du mélange entre deux illuminants spécifiés par l'utilisateur. Cette approche fournit de bons résultats tant que l'illuminant varie lentement mais échoue lorsque des transitions abruptes de l'illuminant sont présentes. Pour surmonter ce problème, Boyadzhiev [14] propose un algorithme permettant de corriger une image en se basant sur les indications d'un utilisateur, spécifiant si les surfaces sont neutres ou uniformes.

2. Ces surfaces doivent être préalablement identifiées, par exemple à l'aide d'une étape de segmentation [5].

3.1.3 Contributions

Ce chapitre introduit une méthode simple et efficace pour l'estimation de la chromaticité des sources de lumière sans recours à une phase d'apprentissage. Notre approche est basée sur une procédure de vote originale combinée à deux hypothèses physiques. L'une concerne le contenu de la scène observée et la seconde l'ensemble des illuminants possibles. Premièrement, la scène est supposée contenir au moins quelques pixels achromatiques (ou parfaitement réfléchissants). Deuxièmement, l'ensemble des chromaticités pouvant être associé à un illuminant est supposé être au voisinage du lieu de Planck. Une version raffinée de la procédure de vote est également proposée. Elle permet de s'adapter aux situations où plusieurs illuminants sont présents, en estimant automatiquement le nombre de ces illuminants ainsi que leurs couleurs respectives. L'ensemble de l'algorithme ne nécessite pas de phase d'apprentissage et fait appel à des paramètres intuitifs. De nombreuses expériences viennent illustrer l'efficacité de la méthode proposée. Une comparaison sur différentes bases de données entre cette approche et d'autres méthodes de l'état de l'art est également présentée et montre l'efficacité de la méthode présentée. Cet algorithme a fait l'objet de deux publications à l'occasion de conférences [101, 102] ainsi que d'une soumission au journal *IEEE Transactions on Image Processing*.

Certaines notions, telles que la modélisation de la formation des images couleur et l'équation de Planck, qui sont nécessaires à la compréhension de ce chapitre, ont été présentées dans le chapitre 2.1. La suite de ce chapitre est dédié au problème de constance de couleur. La Section 3.2 est dédiée à la description des principales étapes de notre algorithme dans le cas où un seul illuminant est présent dans la scène considérée. La version multi-illuminant est pour sa part détaillée dans la Section 3.3. Les Sections 3.4.2 et 3.4.3, sont consacrées à l'évaluation de la capacité de l'algorithme à retrouver la couleur d'une, puis de plusieurs sources lumineuses. Cette évaluation est menée sur plusieurs bases de données, couvrant à la fois le cas mono et multi-illuminant.

3.2 Choisir l'illuminant le plus probable

Nous présentons maintenant une première version de notre algorithme d'estimation d'illuminant. Nous supposons pour l'instant qu'un seul illuminant est présent dans la scène étudiée. Comme décrit précédemment, estimer le triplet (L_R, L_G, L_B) à partir des mesures $p_c(\mathbf{m})$ est un problème insuffisamment contraint et il nécessite des hypothèses supplémentaires. Nous nous appuyons ici sur deux hypothèses.

La première concerne l'ensemble des illuminants possibles que nous supposons bien décrit par le lieu de Planck et son voisinage. Cet *a priori* est commun dans la littérature et constitue une hypothèse raisonnable pour la plupart des illuminants [45, 82].

Notre seconde hypothèse concerne le contenu de la scène. Elle est inspirée de la méthode *White-Patch* : nous supposons que la scène contient au moins une surface achromatique ou parfaitement réfléchissante. Bien qu'il soit bien évidemment possible de trouver ou de créer un contre-exemple, nous avons observé qu'en pratique, cette hypothèse est très souvent satisfaite, particulièrement dans les scènes contenant des objets fabriqués par l'homme. Ces deux hypothèses seront discutées dans la section dédiée à l'évaluation de l'algorithme.

L'idée principale de notre méthode est de sélectionner les pixels proches du lieu de Planck et de faire voter ces pixels pour l'illuminant le plus probable.

Avant de décrire l'algorithme plus précisément, un détail important doit être spécifié sur l'espace couleur utilisé pour les images. L'espace RGB d'un appareil photographique est défini par les fonctions de sensibilité de cet appareil. Bien qu'il soit possible d'estimer l'illuminant directement dans cet espace, il est alors difficile d'évaluer la qualité des résultats, puisque les mesures d'erreurs ne sont pas comparables d'un espace RGB à l'autre. De plus les distances entre couleurs calculées dans différents espaces RGB sont différentes. Dans le but d'obtenir un algorithme générique, il est important que l'image d'entrée soit donnée dans un espace couleur unique tel que sRGB (on peut noter qu'aujourd'hui, la plupart des fabricants d'appareils photographiques utilisent ou permettent d'utiliser cet espace pour sauvegarder les images.). Cet espace bien connu est choisi car il très répandu parmi les appareils modernes et permet des conversions faciles vers d'autres espaces colorimétriques comme ceux de la CIE.

La première partie de l'algorithme fonctionne de la manière suivante. Premièrement, pour tous les pixels \mathbf{m} , le triplet couleur observé $p_c(\mathbf{m})$ ($c \in \{R, G, B\}$) est transformé en un vecteur 2D $\mathbf{c}(\mathbf{m}) = (u(\mathbf{m}), v(\mathbf{m}))$ dans le diagramme uv CIE 1960 (cf. Section 2.1.3). Travailler dans un espace de chromaticité est raisonnable car l'information d'intensité de l'illuminant ne peut pas être récupérée. Dans cet espace, tous les pixels situés à une distance du lieu de Planck \mathcal{L} supérieure à un seuil δ sont alors rejetés. Ceci signifie que nous ne conservons que les pixels \mathbf{m} vérifiant

$$(\min_{\mathbf{r} \in \mathcal{L}} \|\mathbf{c}(\mathbf{m}) - \mathbf{r}\|_2) < \delta, \quad (3.1)$$

En pratique, le seuil δ est choisi de manière à englober les principaux illuminants. Bien que cela n'ait pas été considéré, il peut également être choisi à l'aide une hypothèse sur la distribution des illuminants.

Ces pixels sont considérés comme probablement gris dans l'image. Nous projetons ensuite orthogonalement chacune des chromaticités conservées $\mathbf{c}(\mathbf{m})$ sur le lieu de Planck :

$$\mathbf{c}_{Planck}(\mathbf{m}) := \text{proj}_{\mathcal{L}}(\mathbf{c}(\mathbf{m})) \in \underset{\mathbf{r} \in \mathcal{L}}{\text{argmin}} \|\mathbf{c}(\mathbf{m}) - \mathbf{r}\|_2. \quad (3.2)$$

Pour calculer cette projection, la solution la plus simple consiste à discrétiser le lieu de Planck, puis à effectuer une recherche exhaustive. Bien que cela ne concerne que des chromaticités loin du lieu de Planck, si l'argmin dans l'équation (3.2) n'est pas défini de manière unique, un choix aléatoire peut être opéré. En pratique, l'approximation de la projection présentée au paragraphe 2.2.2 (Équations 2.22) est utilisée. Comme cela est rappelé au paragraphe, cette assignation permet d'associer à chaque pixel \mathbf{m} une température de couleur proximale $T(\mathbf{m})$ correspondant à la température du corps noir dont la chromaticité est $\mathbf{c}_{Planck}(\mathbf{m})$.

Seuls sont conservés les pixels dont la température de couleur appartient à l'intervalle $[T_{min}, T_{max}]$. Cette intervalle associé au seuil δ décrit l'ensemble des couleurs liées aux illuminants que l'algorithme considère possible. En pratique, cet ensemble représente la plupart des couleurs des illuminants [82]. Concernant les sources artificielles, l'approximation est correcte tant que la projection du spectre de cette source dans le diagramme de chromaticité est proche du lieu de Planck. C'est-à-dire tant que la couleur perçue de la source est proche d'une de celles prédites par le corps noir. Ce qui est souvent le cas puisque les sources artificielles sont souvent conçues pour imiter la couleur de la lumière naturelle (à l'exception de besoins ponctuels tels que l'éclairage de scènes *etc.*).

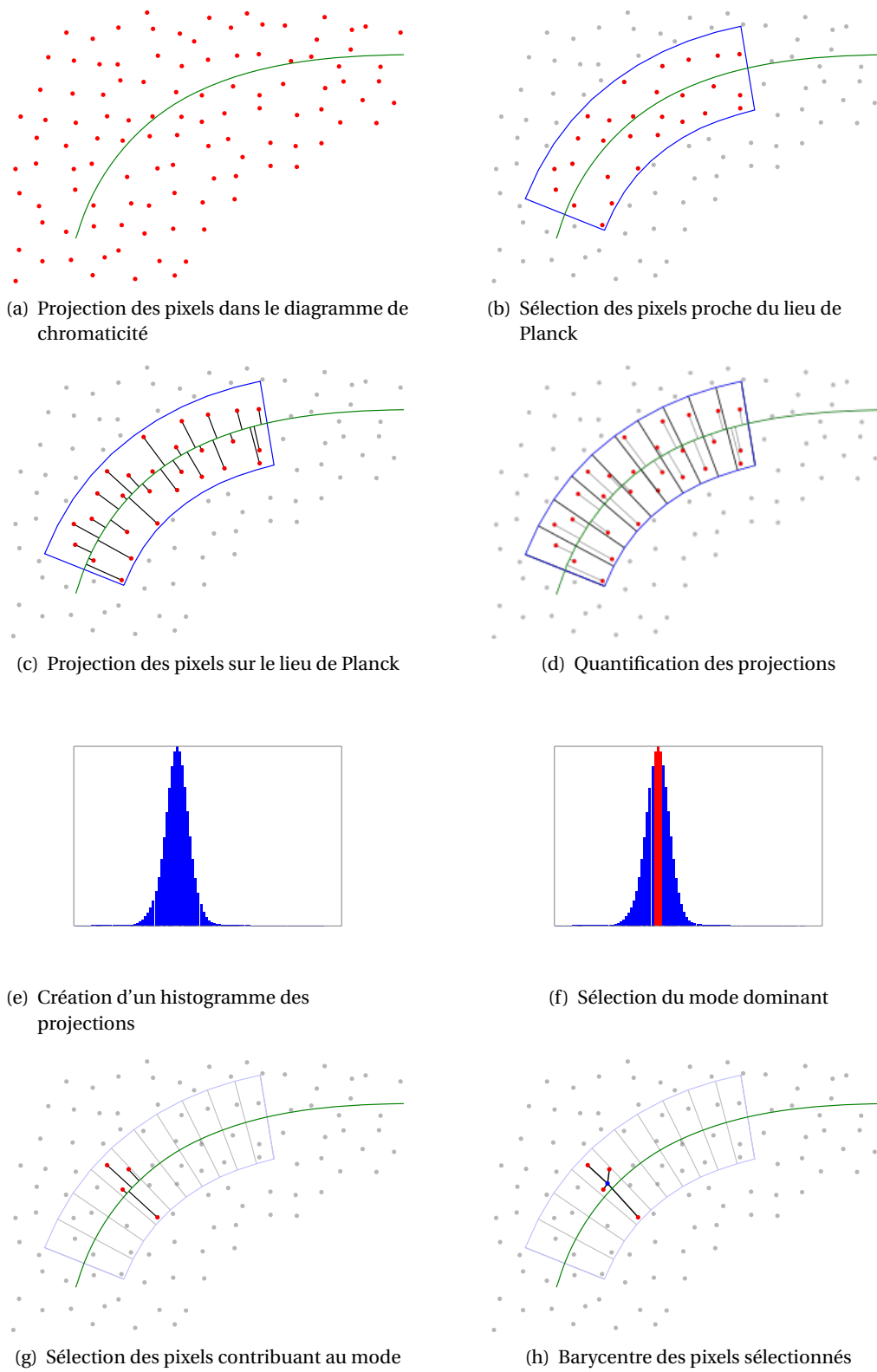


FIG. 3.1 Illustration des différentes étapes pour la sélection des pixels gris par notre algorithme.

La seconde étape de l'algorithme est une procédure de vote permettant de retrouver la température de couleur T^0 la plus probable pour un illuminant. À la fin de la première étape, les points restants dans le diagramme de chromaticité sont utilisés pour construire un histogramme des températures de couleur. Pour calculer cet histogramme discret, nous convertissons la température de couleur proximale T (mesurée en Kelvin) dans l'échelle MIREC (Micro reciprocal degree) :

$$T_{Mired} = \frac{10^6}{T}$$

avant la discrétisation. Comme cela a été expliqué au paragraphe 2.2.2, l'échelle de température en Kelvin ne permet pas un échantillonnage satisfaisant du lieu de Planck. La conversion à l'échelle MIREC a été proposée par Priest [122] comme étant un bon moyen pour échantillonner la température de couleur en accord avec la perception humaine.

Lors du calcul de l'histogramme des températures de couleur, la contribution de chaque pixel est pondérée par une puissance de sa luminance. Cette pondération permet de favoriser les pixels les plus lumineux de l'image, et ainsi de limiter l'influence des pixels les plus sombres sans pour autant les ignorer³.

La température T^0 de l'illuminant est alors estimée comme étant la plus représentée dans l'histogramme pondéré. Par simplicité, l'argmax est utilisé ici. Mais une procédure plus élaborée, basée sur une détection de mode, peut également être appliquée. Nous verrons dans la prochaine section comment une telle procédure peut être utilisée pour la détection de sources lumineuses de couleurs différentes.

Ce choix fait apparaître une hypothèse implicite de notre méthode. Nous supposons que la couleur de l'illuminant appartient à un sous-domaine des couleurs mesurables par l'appareil et que la scène contient une surface grise. Mais, en choisissant la température ayant reçu la plus grande masse de vote, nous supposons également que la surface grise présente dans l'image contribue avec un poids plus important à l'histogramme que les autres surfaces ayant une couleur proche du lieu de Planck. Par exemple sous un illuminant neutre (une lumière blanche), si une scène contient une surface grise et une surface bleue dont la couleur est proche du lieu de Planck, alors nous supposons que la surface grise contribue de manière plus importante à l'histogramme que ne le fait la surface bleue. Dans le cas inverse, une mauvaise estimation sera fournie par l'algorithme.

Comme dernière étape de notre algorithme, nous calculons le barycentre de tous les pixels \mathbf{m} ayant participé au choix de T^0 (c'est à dire les pixels gardés pour construire l'histogramme et dont la température proximale de couleur appartient à la classe T^0). Ce barycentre, pouvant être en dehors du lieu de Planck mais proche de lui, est choisi comme estimation finale de l'illuminant. L'Algorithme 1 donne une description complète de la procédure.

3.3 Raffinement de l'algorithme par une sélection de modes

L'algorithme décrit précédemment fournit une estimation de l'illuminant après avoir sélectionné le mode principal de l'histogramme des températures par un moyen assez frustré (argmax de l'histogramme). En pratique, l'histogramme peut contenir plusieurs modes, en

3. D'une certaine manière, cela peut être vu comme un compromis entre les hypothèses de « Monde-Gris » et de *White-Patch* [50].

Données : Image \mathbf{p} dans l'espace sRGB. seuils δ , T_{\min} , T_{\max} , nombre N de classes dans l'histogramme, puissance n , illuminant canonique $(u^{\text{ref}}, v^{\text{ref}})$.

Sorties : Chromaticités (u^e, v^e) estimées dans l'espace CIE 1960 uv .

Initialiser l'histogramme $H = 0$ sur N classes;

pour chaque pixel \mathbf{m} faire

```

     $\mathbf{c}(\mathbf{m}) = \text{Convert}_{s\text{RGB} \rightarrow uv}(p_c(\mathbf{m}));$ 
     $[\mathbf{c}_{\text{Planck}}(\mathbf{m}), T(\mathbf{m})] = \text{Proj}_{\text{Planck}}(\mathbf{c}(\mathbf{m}));$ 
    //  $\mathbf{c}_{\text{Planck}}(\mathbf{m})$  est la projection des chromaticités  $\mathbf{c}(\mathbf{m})$  sur le lieu de
    Planck dans l'espace de chromaticité  $uv$  et  $T(\mathbf{m})$  est la
    température (en Kelvin) de cette projection;
     $d(\mathbf{m}) = \text{distance}(\mathbf{c}(\mathbf{m}), \mathbf{c}_{\text{Planck}}(\mathbf{m}));$ 
    si  $d(\mathbf{m}) < \delta$  et  $T_{\min} \leq T(\mathbf{m}) \leq T_{\max}$  alors
        |  $T_{\text{MIRE}}(\mathbf{m}) = 10^6 / T(\mathbf{m});$ 
        |  $\text{poids}(\mathbf{m}) = \text{luminance}(\mathbf{m})^n;$ 
        | Ajouter  $\text{poids}(\mathbf{m})$  à la classe  $T_{\text{MIRE}}(\mathbf{m})$  dans  $H$ ;
    fin

```

fin

si $H == \emptyset$ **alors**

```

    | retourner  $(u^{\text{ref}}, v^{\text{ref}})$ ;

```

fin

Trouver $T^0 = \arg \max(H)$;

$P =$ liste de pixels \mathbf{m} tels que $T_{\text{MIRE}}(\mathbf{m}) = T^0$ et $d(\mathbf{m}) < \delta$

retourner $(u^e, v^e) = \text{Barycentre}(P)$;

Algorithme 1 : Algorithme de sélection d'illuminant. En pratique, les paramètres suivants procurent de bons résultats sur les différentes bases de données testées : $\delta = 0.0125$, $T_{\min} = 2000K$, $T_{\max} = 20,000K$, $N = 30$, et $n = 3$.

particulier lorsque plusieurs illuminants sont présents dans la scène. De plus, même lorsqu'un unique illuminant est présent, l'argmax n'est pas toujours le moyen le plus approprié puisque le mode peut s'étendre sur une large gamme de températures. Dans cette section, nous proposons un algorithme de sélection de modes, inspiré des méthodes *a contrario* [33]. Cet algorithme permet de raffiner la détection et de s'adapter au cas où plusieurs illuminants de couleurs différentes éclairent la scène.

Comme nous n'avons aucun *a priori* sur le nombre d'illuminants présents dans la scène, les approches de segmentation comme l'algorithme des k -moyennes ne peuvent être utilisées directement. Un mélange de gaussiennes avec un critère MDL (*minimum description length*) peut éventuellement être utilisé pour résoudre le problème, mais cela amène à introduire un *a priori* sur la forme des modes à détecter qui n'est pas souhaitable.

Nous proposons ici un algorithme générique pour la détection de modes dans les histogrammes. Cet algorithme est détaillé dans l'Annexe A. Il est inspiré de la détection de modes *a contrario* présentée dans [32, 33] et possède l'avantage d'être adapté à la procédure spécifique de vote pondéré utilisée pour construire l'histogramme des températures de couleur proximales.

Nous précisons ici simplement les hypothèses qui sont faites sur la construction de l'histogramme. Soit H un histogramme de N classes, correspondant chacune à un intervalle de températures de couleur proximales de couleur. Cet histogramme est décrit par l'algorithme 1 : la contribution de chaque pixel \mathbf{m}_i $i \in \{1, \dots, M\}$ est pondérée par une fonction puissance de sa luminance. Suivant l'idée générale des méthodes *a contrario*, les modes sont détectés comme étant des intervalles dont la masse totale contredit une hypothèse nulle \mathcal{H}_0 . Décrivons plus précisément cette hypothèse \mathcal{H}_0 .

Hypothèse \mathcal{H}_0 Pour un intervalle $H[a, b]$ et un pixel \mathbf{m}_i , nous définissons deux variables aléatoires C_i et W_i . La première, C_i , correspond à l'évènement "la température de couleur proximale de \mathbf{m}_i appartient à l'intervalle $H[a, b]$ ". La seconde variable aléatoire, W_i , correspond à la masse w_i de \mathbf{m}_i . Les variables C_i et W_i $i \in \{1, \dots, M\}$ suivent l'hypothèse nulle \mathcal{H}_0 si :

- elles sont mutuellement indépendantes,
- toutes les variables C_i suivent une loi de Bernoulli de paramètre $p_{[a,b]}$ ($p_{[a,b]}$ étant la longueur relative de l'intervalle $H[a, b]$),
- toutes les variables W_i suivent une loi exponentielle d'espérance λ .

La distribution exponentielle a été choisie parce qu'elle constitue une approximation raisonnable de la loi suivie par les poids w_i dans la plupart des images naturelles (rappelons que les poids w_i sont égaux à une puissance de la luminance). En pratique, le paramètre λ est appris pour chaque image.

Pour chaque intervalle maximal significatif, nous appliquons la procédure finale décrite dans l'algorithme 1 (les deux dernières lignes) ce qui revient à moyenniser les pixels contribuant à chaque mode de façon à obtenir autant d'estimations d'illuminants que de modes.

3.4 Expériences

Dans cette section, nous étudions les performances de notre algorithme d'estimation d'illuminants. Notre méthode est d'abord évaluée et confrontée aux méthodes standards de l'état de l'art sur des scènes ne contenant qu'un illuminant. Pour s'assurer de la qualité de ces résultats, la comparaison est effectuée sur trois bases d'images différentes, à savoir la base *Colorchecker* [53], la base *Grey-Ball* [25] ainsi que la base *SFU laboratory* [6]. Dans un deuxième temps, nous évaluons la méthode dans un cadre multi-illuminants.

3.4.1 Espaces de couleur

Avant de décrire le protocole expérimental utilisé pour les expériences, il est important de spécifier les espaces de couleur utilisés dans ces expériences, ainsi que la raison de cette utilisation. Comme expliqué précédemment dans la Section 3.2, chaque appareil possède son propre espace RGB, défini par les fonctions de sensibilité de ses capteurs. En conséquence, les mesures RGB de deux appareils ayant des fonctions de sensibilité différentes ne sont pas directement comparables. Leurs composantes couleur peuvent par exemple avoir des gammes de dynamique très variables. Pour ne pas dépendre de ces espaces spécifiques, notre algorithme travaille sur des images fournies dans l'espace sRGB. Cela a plusieurs avantages. Premièrement, les paramètres de l'algorithme n'ont pas à être adaptés à l'espace couleur de l'appareil et peuvent être fixés une fois pour toutes dans sRGB. Deuxièmement, le calcul de distances couleur ainsi que la projection sur le lieu de Planck peuvent être effectués de manière plus satisfaisante dans cet espace que dans un espace dépendant de chaque appareil. En

effet, il est ainsi facile de convertir les couleurs depuis sRGB en chromaticités uv , ce qui est recommandé [155] par la CIE pour le calcul des températures de couleur⁴. Pour terminer, travailler avec un unique espace est le seul moyen d'obtenir des mesures d'erreurs entre estimations d'illuminants qui soient cohérentes d'un appareil à l'autre. Nous pensons que cette cohérence entre images est essentielle puisque ces mesures d'erreurs sont utilisées pour le calcul de statistiques sur différentes bases d'images.

En pratique, lorsque les images sont fournies dans l'espace de leur appareil, il est donc nécessaire pour appliquer notre algorithme de les convertir en sRGB. Cela ne pose pas de réels problèmes puisque la plupart des appareils photographiques modernes permettent la conversion depuis leur propre espace de couleur vers sRGB.

3.4.2 Estimation mono-illuminant

Protocole

Pour l'évaluation mono illuminant, nous suivons un protocole expérimental similaire à celui décrit dans [64]. Chaque image utilisée dans ces expériences est fournie avec une mesure \mathbf{L}^g de la couleur de la source de lumière éclairant la scène. Cette mesure est permise par la présence d'une mire de couleur (*Colorchecker*) dans le champ de l'appareil photographique.

Suivant [65], nous adoptons comme mesure d'erreur entre cette vérité terrain et l'estimation de l'illuminant L^e l'erreur angulaire définie par

$$E_a(\mathbf{L}^e, \mathbf{L}^g) = \cos^{-1} \left(\frac{\mathbf{L}^e \cdot \mathbf{L}^g}{\|\mathbf{L}^e\| \cdot \|\mathbf{L}^g\|} \right). \quad (3.3)$$

Cette mesure d'erreur peut être interprétée dans un espace RGB comme l'angle formé par deux axes « gris », chacun étant défini par un des illuminants \mathbf{L}^e et \mathbf{L}^g . Notons que cette mesure est indépendante de l'intensité de chacun des deux illuminants.

Afin d'obtenir une vision globale des résultats par base de données, nous calculons diverses statistiques sur les erreurs mesurées pour chacune d'entre elles. Comme remarqué par les auteurs de [64], la distribution des erreurs n'est pas symétrique et donc la moyenne des erreurs n'est que faiblement informative. La médiane et la mesure *trimean* (c'est une mesure pondérée du médian et des deux quartiles : $\text{trimean} = \frac{Q_1 + 2Q_2 + Q_3}{4}$) constituent ici des statistiques plus représentatives des performances des différents algorithmes.

Bases d'images

Nous décrivons ici les trois bases d'images utilisées pour les comparaisons dans le cas mono illuminant.

Base *Colorchecker* La base *Colorchecker*, fournie par Gehler *et coll.* [53], est composée de 568 images, capturées à l'aide de deux appareils photographiques différents (Canon EOS-1D et Canon EOS-5D). Comme dit précédemment, les résultats calculés dans l'espace RGB de l'appareil puis moyennés entre différents espaces couleur doivent être pris avec précaution. Cette base d'images est composée de photographies prises en intérieur et en extérieur, sous

4. En pratique, nous utilisons l'approximation proposée par [74] pour calculer cette projection et les températures de couleur.

des couleurs d’illuminants variées. La vérité terrain est livrée sous forme de triplets RGB dans les espaces respectifs des appareils. Les deux appareils utilisés pour créer cette base étant bien connus, nous pouvons facilement convertir ces triplets dans l’espace sRGB⁵ et visualiser leurs projections dans le diagramme de chromaticité xy (voir la Figure 3.2).

Gehler *et coll.* [53] donnent cette base de données sous deux formats : RAW et TIFF. La version TIFF [136] ayant subi des corrections non linéaires telles qu’une correction gamma, et des corrections colorimétriques (dont une éventuelle balance des blancs), nos expériences se basent sur la version RAW, laquelle est convertie en sRGB (cf. paragraphe 2.1.4). Afin d’effectuer cette conversion correctement, nous supprimons le niveau de noir des appareils photographiques (129 pour le Canon 5D et 0 pour the Canon 1D). Malgré son importance, cette correction n’est jamais faite par les auteurs travaillant sur cette base.

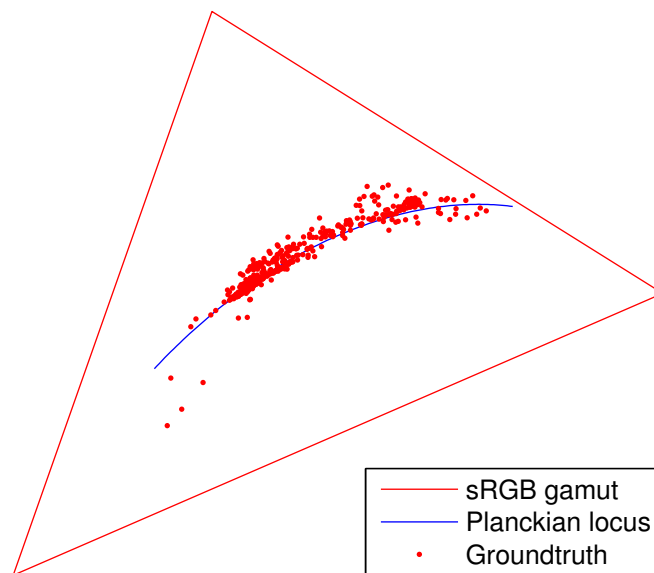


FIG. 3.2 Vérité terrain des illuminants de la base *Colorchecker* dans le diagramme xy .

Base d’images Grey-Ball La base d’images *Grey-Ball* [25] contient 11346 images JPEG tirées de vidéos. Comme recommandé par [64], la correction gamma est enlevée afin de travailler sur des images linéaires pour lesquelles les intensités mesurées sur chacun des canaux RGB est linéairement proportionnelle au nombre de photons reçus par le capteur. L’espace RGB de l’appareil (Sony VX-200) étant inconnu, nous travaillons comme si ces images étaient dans l’espace sRGB.

Pour chaque image, une boule grise est présente dans le champ visuel de l’appareil. Les créateurs de la base de données utilisent les pixels appartenant à cette boule pour établir une vérité terrain des illuminants et fournissent ces illuminants sous la forme de triplets RGB⁶. Comme illustré par la Figure 3.3, cette base de données couvre elle aussi une large gamme de couleurs d’illuminants.

5. Les matrices de conversion sont documentées dans dcraw [26].

6. Les pixels appartenant à cette boule grise sont ignorés lors des expériences.

TAB. 3.1 La première table montre les performances sur la base *colorchecker* [136] dans l'espace RGB de l'appareil photo et la seconde table, celles obtenues dans l'espace sRGB. Les résultats sont comparés à des méthodes sans apprentissage ([6, 16, 89, 145]) ainsi qu'à des méthodes impliquant une phase d'apprentissage ([20, 51, 63]).

Méthode	Moy.	Med.	Trim.	Meill.-25%	Pires-25%
White-Patch [89]	7.5°	5.7°	6.4°	1.5°	16.2°
Grey-World [16]	6.4°	6.3°	6.3°	2.3°	10.6°
general Grey-World [6]	4.7°	3.5°	3.8°	1.0°	10.2°
1st – order Grey-Edge [145]	5.3°	4.5°	4.7°	1.8°	10.2°
Pixel-based Gamut Mapping [51]	4.2°	2.3°	2.9°	0.5°	10.8°
Edge-based Gamut Mapping [63]	6.5°	5.0°	5.4°	1.9°	13.6°
ML (category-wise prior) [20]	3.7°	3.0°	3.1°	1.0°	7.6°
Méthode proposée $\delta = 0.02$	4.1°	2.7°	3.1°	0.8°	9.6°
Méthode proposée (paramètres fixes)	4.5°	3.1°	3.5°	0.8°	10.8°

Méthode	Moy.	Med.	Trim.	Meill.-25%	Pires-25%
White-Patch [89]	9.6°	7.7°	8.6°	2.2°	20.3°
Grey-World [16]	9.1°	8.6°	8.8°	2.9°	16.2°
general Grey-World [6]	6.6°	4.9°	5.4°	1.4°	14.6°
1st – order Grey-Edge [145]	7.0°	6.3°	6.4°	2.8°	12.5°
Pixel-based Gamut Mapping [51]	5.6°	3.4°	4.0°	0.8°	13.8°
Edge-based Gamut Mapping [63]	9.6°	7.4°	8.1°	2.8°	20.0°
ML (category-wise prior) [20]	5.0°	4.1°	4.3°	1.4°	10.1°
Méthode proposée $\delta = 0.02$	5.2°	4.0°	4.3°	1.4°	10.8°
Méthode proposée (paramètres fixes)	6.1°	4.2°	4.8°	1.2°	14.1°

Base de données *SFU laboratory* Cette base de données [6] contient 321 images acquises à partir de 31 scènes, chacune étant illuminée par 11 sources de lumière de couleurs différentes. Les scènes ne contiennent souvent que peu d'objets et donc la gamme de couleur observable dans une image est limitée. Cette base constitue donc un cas difficile pour les algorithmes de constance de couleur.

Pour chacune des images de cette base de données, les auteurs fournissent une mesure du spectre de l'illuminant ainsi que les valeurs RGB correspondantes dans l'espace RGB de l'appareil. En projetant ces spectres dans le diagramme de chromaticité xy , comme illustré sur la partie gauche de la Figure 3.4, on observe que tous ces illuminants sont situés dans un voisinage proche du lieu de Planck.

Afin d'appliquer notre algorithme aux images de cette base, il est nécessaire de les convertir dans l'espace sRGB. Ceci est possible si l'on connaît la matrice de transformation de l'espace RGB de l'appareil vers l'espace sRGB⁷. Pour estimer cette matrice, nous utilisons une image de la base contenant un *colorchecker*.

7. Cette matrice de conversion peut également être estimée à partir des fonctions de sensibilité de l'appareil (qui sont ici connues). Cependant, ces fonctions de sensibilité sont rarement connues en pratique, alors qu'il est toujours possible de prendre une photographie d'un *colorchecker*.

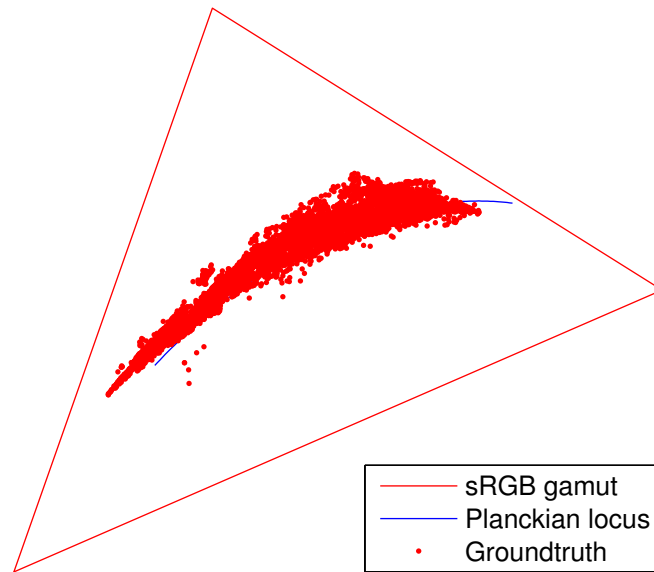


FIG. 3.3 Vérité terrain de la base Grey-World une fois projetée dans le diagramme de chromaticité xy .

TAB. 3.2 Performances sur la base d'images SFU grey-world linéaire, dans l'espace de couleur de l'appareil photographique.

Méthode	Moy.	Méd.	Trim.	Meill.-25%	Pires-25%
White-Patch [89]	12.7°	10.5°	11.3°	2.5°	26.2°
Grey-World [16]	13.0°	11.0°	11.5°	3.1°	26.0°
general Grey-World [6]	11.6°	9.7°	10.2°	3.4°	22.7°
1st – order Grey-Edge [145]	10.6°	8.8°	9.2°	3.0°	21.1°
Pixel-based Gamut Mapping [51]	11.8°	8.9°	10.0°	2.8°	24.9°
Edge-based Gamut Mapping [63]	12.8°	10.9°	11.4°	3.6°	25.0°
ML [20]	10.3°	8.9°	9.2°	2.8°	20.3°
Méthode proposée ($\delta = 0.0075$)	10.5°	8.2°	8.9°	2.2°	22.5°
Méthode proposée (paramètres fixes)	11.1°	9.1°	9.5°	2.6°	23.0°

À ce stade, il est important de noter que les fonctions de sensibilité de l'appareil ne permettent pas une transformation parfaite depuis l'espace RGB de l'appareil vers de l'espace sRGB. Cela est illustré par la partie droite de la Figure 3.4 : Les triplets RGB des illuminants utilisés pour créer la base sont fournis par les auteurs dans l'espace RGB de l'appareil photo. Lorsque cette vérité terrain est convertie dans l'espace sRGB par la transformation précédemment calculée, alors les projections sont significativement différentes de celles obtenues par projection directe des spectres des différents illuminants. Néanmoins ce problème ne devrait pas apparaître avec des appareils photographiques récents offrant la compatibilité avec l'espace sRGB.

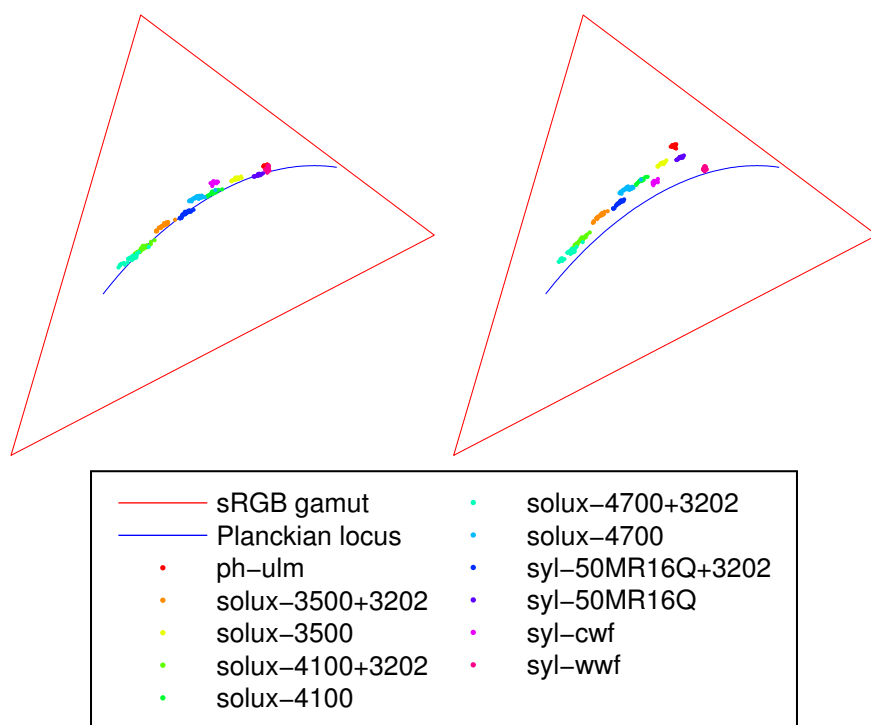


FIG. 3.4 Vérité terrain des illuminants de la base *SFU LAB* après projection dans le diagramme de chromaticité xy . Sur la gauche, la projection directe du spectre des illuminants. Sur la droite, la projection des triplets d'illuminants après conversion en sRGB. La différence entre ces deux projections est expliquée par l'impossibilité d'une conversion parfaite de l'espace RGB de l'appareil vers l'espace sRGB (plus de détails sont fournis dans le texte).

Évaluation comparative

Les Tables 3.1 à 3.5 montrent les résultats obtenus à partir des trois bases d'images décrites ci-dessus. Dans chacune de ces tables, les résultats de notre approche sont fournis avec deux jeux de paramètres différents : un jeu identique pour toutes les expériences (dans le cas mono-illuminant) et un second jeu pour lequel le paramètre δ est optimisé pour la base de données considérée.

Colorchecker Pour la base d'images *Colorchecker*, les résultats sont donnés à la fois dans l'espace RGB de l'appareil ainsi que dans l'espace sRGB (voir la Table 3.1). Les résultats des approches de l'état de l'art sont reproduits à partir de [64]. Pour pouvoir fournir les performances de notre algorithme dans l'espace de l'appareil et les comparer à ces résultats, nous y projetons nos estimations de l'espace sRGB. Pour les raisons données précédemment, ces résultats sont à prendre avec précautions : les composantes RGB des deux appareils n'étant pas comparables, en conséquence, moyenner ces erreurs n'est pas forcément cohérent.

Pour cette raison, nous donnons également les statistiques des erreurs de toutes les méthodes dans sRGB. Pour le calcul de ces statistiques et pour tous les algorithmes de l'état de l'art, les estimations d'illuminants sont projetées depuis l'espace de l'appareil vers de l'espace sRGB. On peut noter que notre approche produit de meilleurs résultats que toutes les approches ne faisant pas appel à l'apprentissage. Nous voyons également, à partir de cette table, que tout en

conservant un coût algorithmique très raisonnable, notre approche permet des performances similaires à celles d’algorithmes plus complexes de l’état de l’art, qui eux font appel à une phase d’apprentissage.

Grey-Ball Sur la base *Grey-Ball* (voir la Table 3.2), Notre approche montre également de meilleurs résultats que les méthodes standards d’estimation de l’illuminant. Sur cette base, les résultats sont même légèrement meilleurs que les approches avec apprentissage dans le cas où le paramètre δ est bien choisi. On peut aussi remarquer que l’algorithme est stable, procurant de bons résultats à la fois sur les bases de données *Colorchecker* et *Grey-Ball* avec le même jeu de paramètres.

SFU Lab De même que pour la base *Colorchecker*, les résultats sur la base *SFU LAB* sont fournis dans l’espace RGB de l’appareil ainsi que dans l’espace sRGB, et ce dans le but de comparer l’algorithme présenté aux méthodes de l’état de l’art.

La Table 3.3 montre les erreurs statistiques mesurées sur l’ensemble de la base. En dépit du fait que notre algorithme nécessite une conversion propre entre l’espace RGB de l’appareil et l’espace sRGB, les performances de notre algorithme sur cette base sont proches de celles des algorithmes récents de l’état de l’art. En regardant la valeur médiane et la mesure *trimean*, nous pouvons observer que les performances de notre approche sont légèrement inférieures à celles de l’état de l’art. Sur la colonne correspondant au 25% des meilleurs, nos résultats sont supérieurs. Cela peut être expliqué par une large variance de la distribution des erreurs sur cette base. Quand une scène satisfait les hypothèses de l’algorithme (la présence d’une surface grise dans la scène, ou d’un ensemble de pixels gris très lumineux), notre algorithme permet d’excellentes performances. Dans le cas inverse, et comme le montre la colonne des 25% les plus mauvais, de piètres performances sont observées. Ceci est illustré par la Table 3.5, qui montre les statistiques calculées pour chaque scène. Comme attendu, les scènes contenant des surfaces achromatiques, montrées en haut de la Figure 3.5 permettent de bons résultats à l’inverse de celles n’en possédant pas, comme le montrent les images reproduites en bas de la Figure 3.5. La Table 3.4 montre que les statistiques sont moins dépendantes du type d’illuminant utilisé que de la scène observée.

3.4.3 Estimation multi-illuminant

Dans cette section, nous évaluons la capacité de notre approche à traiter le cas où une scène est éclairée par plusieurs illuminants. Comme décrit dans la Section 3.3, la sélection de modes opérant à la fin de notre algorithme peut être étendue pour détecter plusieurs illuminants dans l’histogramme des températures H . Permettre la détection de plusieurs sources de lumière revient en pratique à supposer que chacun de ces modes a été généré par une source de lumière.

Observons que supprimer l’hypothèse d’unicité de l’illuminant est potentiellement risqué. En effet, l’ambiguïté inhérente au problème d’estimation d’illuminant augmente quand le nombre de sources lumineuses est inconnu. En effet, une couleur observée peut correspondre à une surface neutre vue sous un illuminant coloré autant qu’à une surface colorée vue sous un illuminant canonique.

Évaluer l’algorithme dans ce contexte soulève deux questions. Premièrement, lorsque seulement une source de lumière est présente dans la scène, dans quelle mesure la détec-

TAB. 3.3 Résultats sur la base SFU LAB. La première Table montre les performances dans l'espace RGB de l'appareil photo, la seconde dans l'espace sRGB.

Méthode	Moy.	Méd.	Trim.	Meill.-25%	Pires-25%
White-Patch [89]	9.1°	6.5°	7.5°	1.8°	21.1°
Grey-World [16]	9.8°	7.0°	7.6°	0.9°	23.6°
general Grey-World [6]	5.4°	3.3°	3.8°	0.5°	13.8°
1st – order Grey-Edge [145]	5.6°	3.2°	3.7°	1.1°	14.1°
Pixel-based Gamut Mapping [51]	3.7°	2.3°	2.5°	0.5°	9.4°
Edge-based Gamut Mapping [63]	3.9°	2.3°	2.7°	0.5°	10.0°
Méthode proposée ($\delta = 0.015$)	6.4°	2.4°	3.6°	0.3°	19.3°
Méthode proposée (paramètres fixes)	6.5°	2.7°	4.0°	0.3°	19.0°

Méthode	Moy.	Méd.	Trim.	Meill.-25%	Pires-25%
White-Patch [89]	8.0°	5.6°	6.3°	1.8°	18.1°
Grey-World [16]	10.8°	7.8°	8.3°	1.2°	26.6°
general Grey-World [6]	5.0°	2.7°	3.5°	0.6°	13.0°
1st – order Grey-Edge [145]	5.4°	3.1°	3.7°	1.0°	13.5°
Pixel-based Gamut Mapping [51]	4.2°	2.5°	2.9°	0.6°	10.6°
Edge-based Gamut Mapping [63]	4.5°	2.9°	3.3°	0.7°	11.4°
Méthode proposée ($\delta = 0.015$)	5.3°	1.9°	2.5°	0.3°	16.4°
Méthode proposée (paramètres fixes)	6.2°	2.6°	3.7°	0.4°	17.7°

TAB. 3.4 Performances sur la base SFU lab en fonction de l'illuminant.

Illuminant	Moy.	Méd.	Pires-25%(μ)
ph-ulm	8.5°	4.6°	21.5°
solux-3500+3202	5.3°	1.9°	16.6°
solux-3500	7.2°	2.7°	22.6°
solux-4100+3202	4.8°	1.6°	15.7°
solux-4100	6.5°	2.9°	20.3°
solux-4700+3202	4.5°	1.5°	15.2°
solux-4700	7.1°	2.7°	21.1°
syl-50MR16Q+3202	5.9°	2.1°	18.7°
syl-50MR16Q	7.3°	2.9°	22.4°
syl-cwf	6.9°	3.5°	20.7°
syl-wwf	5.9°	4.6°	14.6°



FIG. 3.5 Exemples de scènes pour lesquelles l'hypothèse de l'algorithme sur la présence d'une surface grise dans la scène est satisfaite (haut) ou non (bas).

TAB. 3.5 Performances sur la base SFU lab en fonction du type de scène.

Scène	Moy.	Méd.	Pires-25% (μ)
Apples	21.4°	22.5°	29.7°
Blocks	27.4°	29.3°	31.2°
Fruit	20.0°	20.4°	25.7°
Paper	14.0°	12.5°	20.2°
Books	4.2°	3.3°	7.3°
Jersey	1.7°	1.6°	2.5°
Monkey	0.8°	0.8°	1.7°
Plastic	2.0°	1.0°	4.9°

tion de plusieurs modes détériore-t-elle les performances? Deuxièmement, lorsque plusieurs illuminants sont en effet impliqués dans la scène, dans quelle proportion les résultats sont-ils améliorés par la détection de ces modes secondaires? Pour chacune de ces questions, une expérience illustre le comportement de l'algorithme.

Mesure de l'erreur entre deux ensembles d'illuminants

Avant de décrire ces expériences, regardons comment mesurer une erreur entre un nombre arbitraire d'illuminants et d'estimations. Pour cette tâche, nous utilisons la mesure EMD introduite par Rubner *et coll.* dans [128]. Soit N^g le nombre d'illuminants présents dans la scène, qui correspondent donc à la vérité terrain, et soit N^e le nombre estimé d'illuminants. Nous attribuons une masse de $1/N^g$ à chacun des illuminants de la vérité terrain ainsi qu'une masse $1/N^e$ à chacune de nos estimations. En prenant comme fonction de coût la distance euclidienne dans le diagramme de chromaticité rg , nous calculons alors la distance EMD entre ces deux ensembles d'illuminants.

En présence de sources multiples

Il n'existe pas de base d'images standard pour l'évaluation d'algorithmes de détection d'illuminants multiples. Il est donc nécessaire d'en construire une. Pour cela, nous suivons le protocole précédemment proposé par Gijsenij *et coll.* dans [67]. Nous créons une nouvelle base à partir de la base *Colorchecker*. Chaque image est en premier lieu corrigée à partir de la vérité terrain qui lui est associée en suivant le modèle de Von Kries. Ensuite, pour chaque image corrigée, deux illuminants sont sélectionnés aléatoirement⁸ dans l'ensemble des illuminants composant la vérité terrain fournie avec la base d'images. L'image corrigée est ensuite illuminée par ces deux sources de lumière. Les deux images ainsi obtenues sont ensuite fusionnées pour n'en former plus qu'une. Cette fusion est opérée par un masque définissant la proportion locale de chacun des illuminants. Plusieurs masques sont utilisés (verticaux, horizontaux et diagonaux). Afin de simuler différentes transitions entre illuminants, un filtrage gaussien, dont la variance est comprise entre 1 et 5, est utilisé pour flouter le masque binaire. Un exemple de masque est visible sur la Figure 3.6.

En présence d'une source unique

Pour le cas mono illuminant, nous reproduisons partiellement l'expérience sur la base *Colorchecker* décrite dans la Section 3.4.2, à la différence que nous autorisons l'algorithme à détecter plusieurs modes. L'expérience illustre ainsi le comportement de l'algorithme lorsque l'*a priori* sur le nombre de modes est levé. Les performances de l'algorithme ne peuvent alors que décroître, et le but de cette expérience est de mesurer cette décroissance.

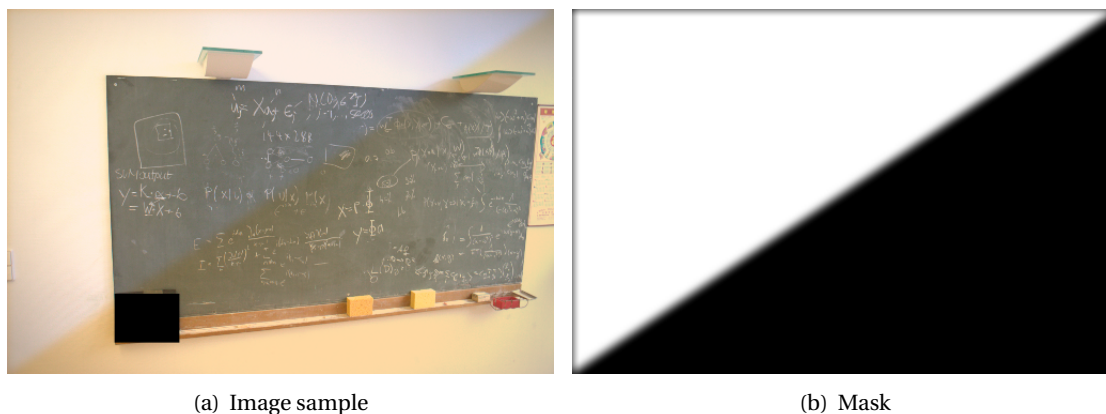


FIG. 3.6 À gauche : un exemple de l'une des images créées pour l'estimation d'illuminants en présence de plusieurs sources lumineuses. À droite : le masque définissant la proportion spatiale de chacun des illuminants.

Résultats

La Table 3.6 montre les performances de notre algorithme sur l'ensemble des images générées en utilisant deux sources lumineuses, comme cela est décrit dans la Section 3.4.3. Notre algorithme est évalué deux fois sur cette base. La première fois, nous choisissons délibérément

8. Nous vérifions que ces deux illuminants ne sont pas trop proches, et empêchons les paires qui ne respectent pas $EMD(L^1, L^2) > 1.11$. Ce seuil correspond grossièrement à celui utilisé dans [67].

TAB. 3.6 Résultats sur des images générées en utilisant deux sources lumineuses (voir la Section 3.4.3). Première ligne : l'algorithme suppose une seule source de lumière. Seconde ligne : l'hypothèse d'unicité est levée. $k = 3, \delta = 0.0150, Tm = 2000, TM = 20000, bins = 300$

Méthode	Crit.	Moy.	Méd.	Trimean	Meilleurs-25%	Pires-25%	Max.
Mono	5.8	6.2	5.3	5.6	1.3	12.6	26.1
Multi	3.7	4.5	2.9	3.4	0.3	11.0	20.3

TAB. 3.7 Résultats sur la base Colorchecker lorsqu'aucune information *a priori* sur le nombre d'illuminants n'est utilisée. $k = 3, \delta = 0.0150, Tm = 2000, TM = 20000, bins = 100$

Méthode	Crit.	Moy.	Méd.	Trimean	Meilleurs-25%	Pires-25%	Max.
Multi modes	3.9	4.7	3.1	3.6	0.6	11.2	23.1
Mono	2.8	3.7	2.0	2.4	0.5	9.6	25.6

de ne garder que le mode le plus significatif de l'histogramme. Dans la seconde nous relâchons l'hypothèse d'unicité et conservons tous les modes maximaux significatifs. Dans cette expérience, autoriser la détection multiple de modes permet d'améliorer significativement la qualité de la détection d'illuminants.

La Table 3.7 montre la perte de performances résultant du relâchement de l'hypothèse d'unicité lorsque l'algorithme est appliqué aux images originales de la base *Colorchecker*. La première ligne donne l'erreur statistique quand le nombre d'illuminants à détecter n'est pas imposé, tandis que la seconde montre les performances de la détection en ne conservant qu'un seul mode. On peut observer que la perte de performance entre les deux configurations est très raisonnable. Nous insistons sur le fait qu'à notre connaissance, aucun algorithme existant n'estime à la fois le nombre d'illuminant et leurs couleurs. Les Figures 3.7, 3.9, 3.8, 3.10, 3.11 et 3.12 présentent quelques exemples de détection multi-illuminant sur des images de la base *colorchecker*.

3.5 Conclusion

Dans ce chapitre, une nouvelle procédure d'estimation de la couleur du ou des illuminants pour les images couleurs a été proposée. Une procédure de vote a été introduite et combinée avec des contraintes physiques, dans le but de sélectionner le ou les illuminants les plus probables dans un espace de chromaticité.

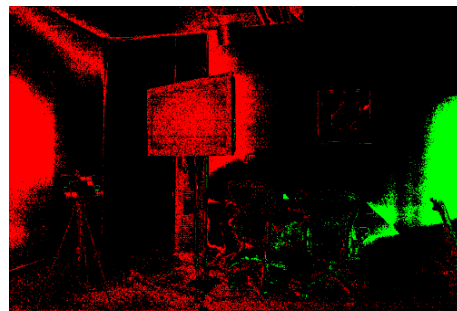
Il a été montré que cet algorithme permet d'atteindre un niveau de performances équivalent à l'état de l'art sur différentes bases d'images calibrées, sans avoir recours à une procédure d'apprentissage. La capacité d'adaptation de la procédure aux situations où le nombre de sources lumineuses de couleurs différentes est inconnue a également été démontrée. À notre connaissance, une telle capacité n'a jamais été présentée auparavant dans la littérature. La méthode proposée travaille globalement sur les images, même dans le cas multi-illuminant. Cela diffère fortement des approches récentes travaillant localement sur des *patches* ou des superpixels [11, 67]. Ces méthodes permettent des corrections locales mais ne peuvent détecter le nombre de sources lumineuses. L'étude du compromis entre notre approche et des approches locales constitue une piste pour de futurs travaux.



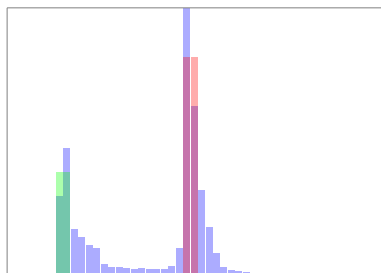
(a) Image non corrigée



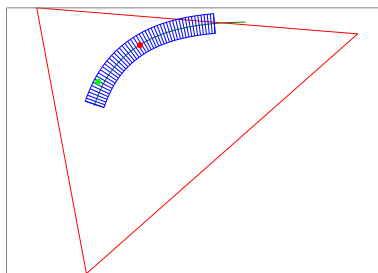
(b) CCT associée à chaque pixel



(c) Pixels votant pour chacun des modes



(d) Histogramme des votes



(e) Illuminants sélectionnés



(f) Correction basée sur le mode rouge.

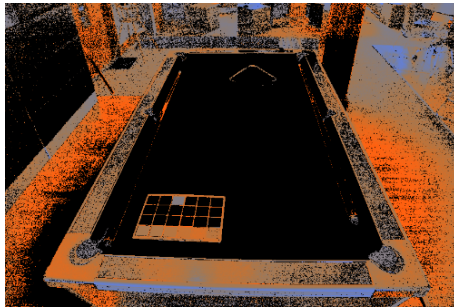


(g) Correction basée sur le mode vert.

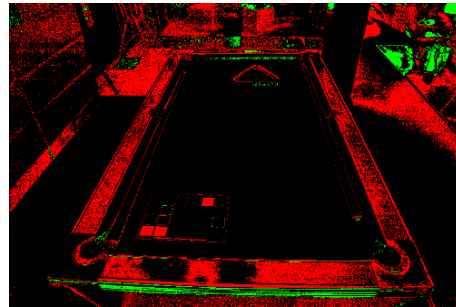
FIG. 3.7 Sur cet exemple, deux sources lumineuses sont présentes. Une source artificielle éclaire la partie gauche de l'image et une source extérieure éclaire la droite de l'image. L'influence de ces illuminants est bien visible sur l'image des projections sur le lieu de Planck 3.7(b). Les pixels contribuant aux modes détectés dans l'histogramme 3.7(d) sont visibles sur l'image 3.7(c), la pondération n'apparaît pas sur cette image.



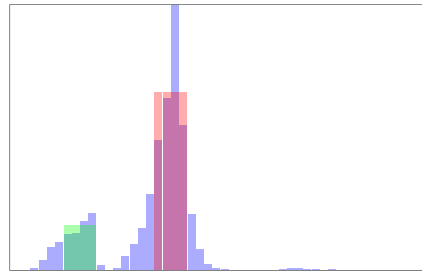
(a) Image non corrigée



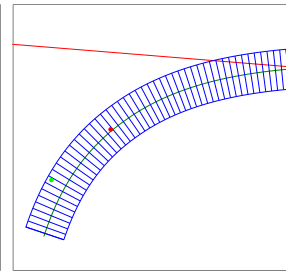
(b) CCT associée à chaque pixel



(c) Pixels votant pour chacun des modes



(d) Histogramme des votes



(e) Illuminants sélectionnés



(f) Correction basée sur le mode rouge.



(g) Correction basée sur le mode vert.

FIG. 3.8 Cette image est illuminée d'un coté par une source extérieure dont les effets sont visibles en haut à droite de l'image et d'un autre coté par une source intérieure qui éclaire le billard. On voit ici l'intérêt de la pondération pour aboutir à un compromis entre la quantité et la luminosité des pixels contribuant. Ce compromis permet d'équilibrer les masses des modes de l'histogramme et ainsi de faciliter la segmentation. L'image 3.8(c) met en évidence des pixels sur le devant du billard, votant pour le mode vert. On peut supposer que la couleur de ces pixels est liée à un illuminant extérieur provenant de l'arrière du billard. On voit ici, l'intérêt que peut avoir la détection multi-illuminants pour la compréhension de scènes.

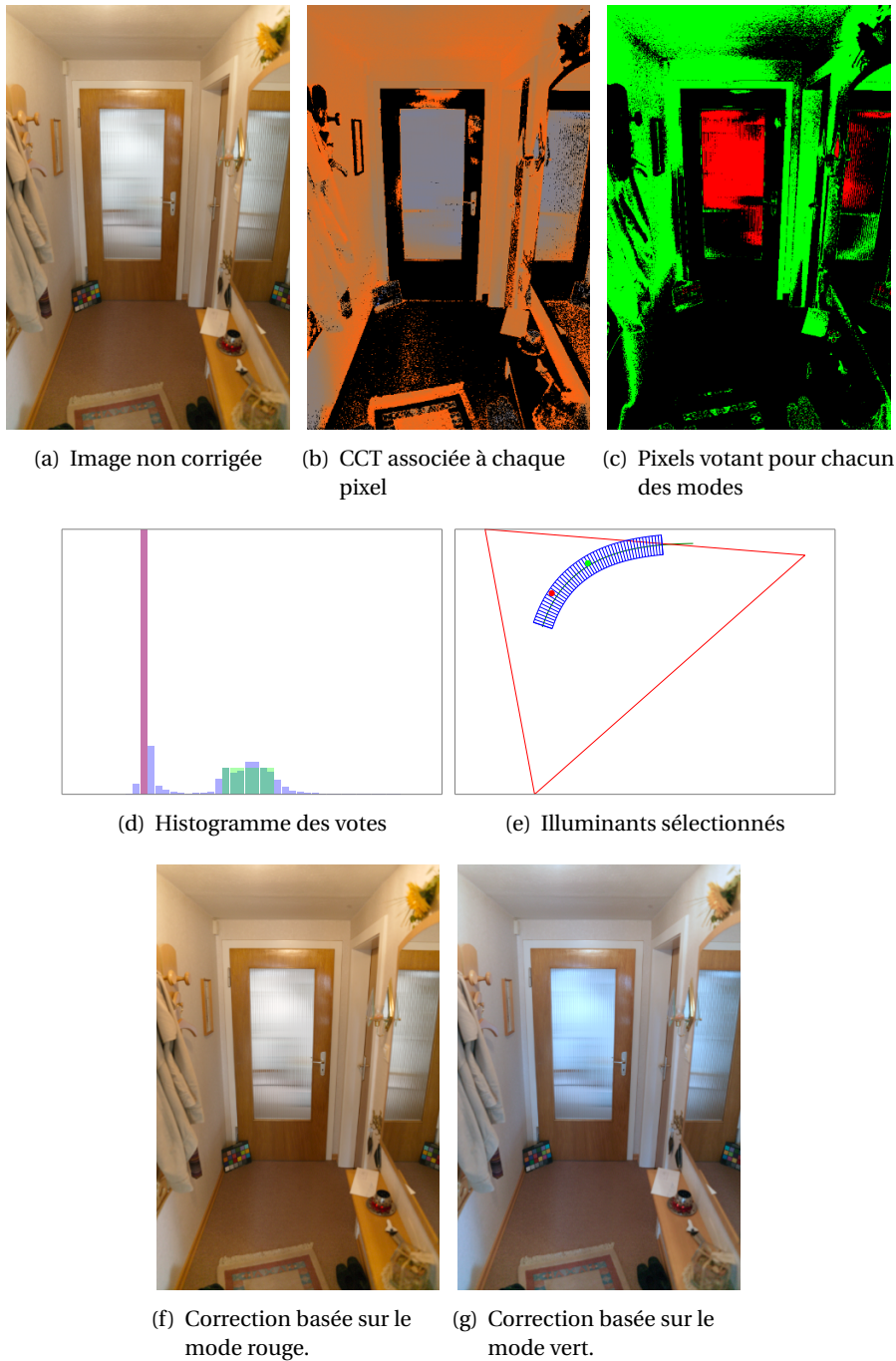
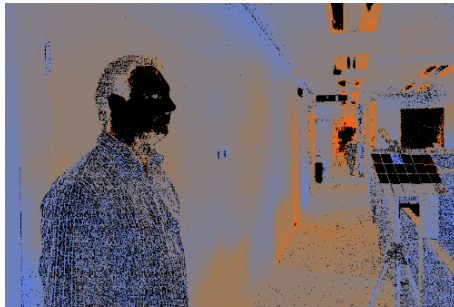


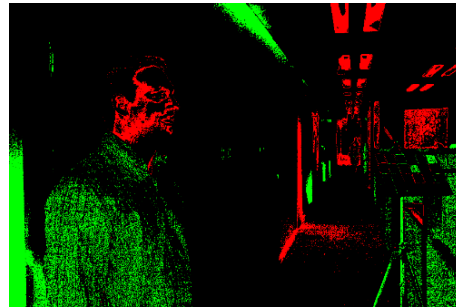
FIG. 3.9 Sur cet exemple également, deux sources lumineuses sont présentes. Une source extérieure éclaire la vitre de la porte tandis qu'une source extérieure illumine le reste de l'image. Les masses des deux modes détectés montrent les différences entre les intensités des illuminants. Ces disparités de masse rendent difficile la segmentation de l'histogramme. Cet exemple montre à la fois la robustesse de l'algorithme ainsi que les limites posées par une approche globale.



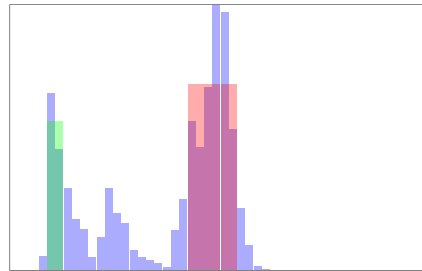
(a) Image non corrigée



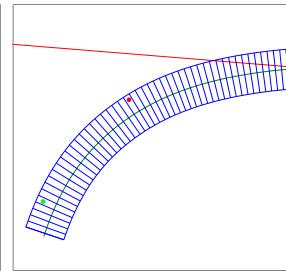
(b) CCT associée à chaque pixel



(c) Pixels votant pour chacun des modes



(d) Histogramme des votes



(e) Illuminants sélectionnés



(f) Correction basée sur le mode rouge.



(g) Correction basée sur le mode vert.

FIG. 3.10 Deux illuminants se mélangent dans cette image : un illuminant extérieur sur le côté droit de l'image et un illuminant intérieur provenant des néons au plafond. L'histogramme de la Figure 3.10(d) possède trois modes dont deux sont détectés. Le mode non détecté, situé au milieu des deux autres, est dû aux pixels du mur, qui sont simultanément éclairés par les deux illuminants. Quand plusieurs illuminants sont présents, la délimitation spatiale entre les zones éclairées n'est pas nécessairement franche. Cela tend à créer un seul mode regroupant l'ensemble des illuminants et il devient alors très difficile de les séparer.

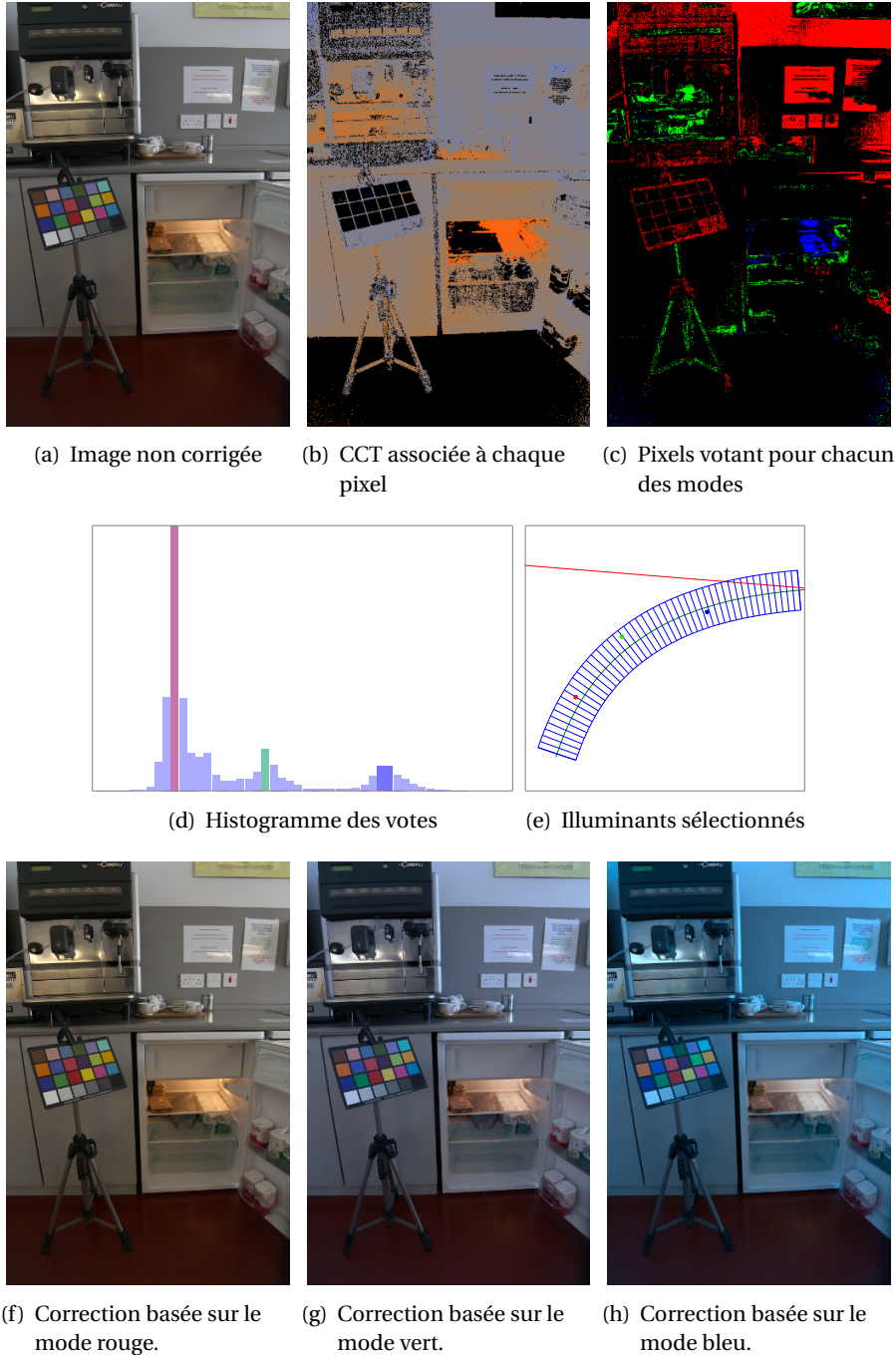
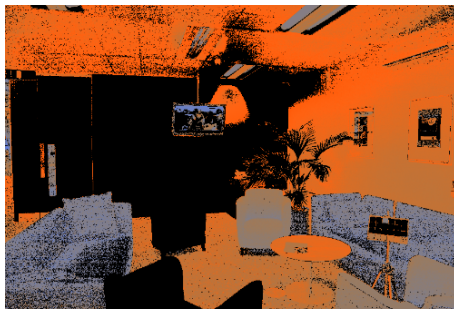


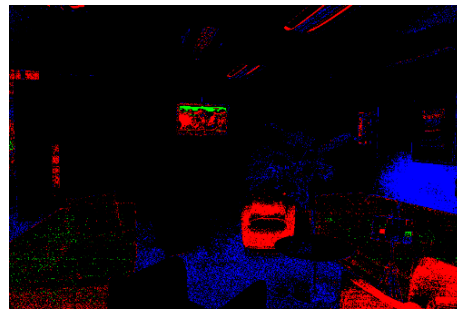
FIG. 3.11 Dans cet exemple, trois modes sont détectés. Le premier correspond à un illuminant extérieur, le deuxième à l'ampoule du réfrigérateur, et le troisième à des reflets sur une machine à café. Le mode correspondant aux reflets métalliques est une détection ambiguë car elle ne correspond pas à une source lumineuse directe.



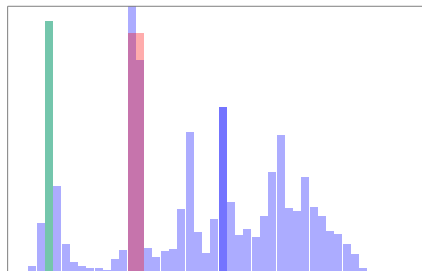
(a) Image non corrigée



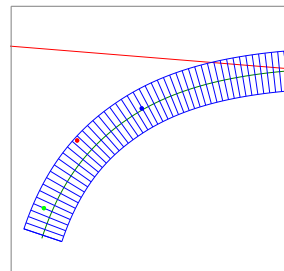
(b) CCT associée à chaque pixel



(c) Pixels votant pour chacun des modes



(d) Histogramme des votes



(e) Illuminants sélectionnés



(f) Correction basée sur le mode rouge.



(g) Correction basée sur le mode vert.



(h) Correction basée sur le mode bleu.

FIG. 3.12 Trois modes sont également détectés ici, le mode bleu correspond aux pixels du mur, le mode rouge aux fauteuils gris-bleu et le mode vert correspond au ciel de l'image diffusée par la télévision. La détection des fauteuils montre l'ambiguïté intrinsèque au problème qui consiste à faire la différence entre une surface bleue éclairée par un illuminant blanc et une surface blanche éclairée par un illuminant bleu.

Couleur et comparaison d'images : état de l'art

Dès les débuts du traitement d'image et du graphisme par ordinateur, la couleur a été utilisée comme source d'information. Tout d'abord pour résoudre des problèmes liés à la détection de contours et à la segmentation [115–117] puis pour des problématiques plus complexes nécessitant la description du contenu couleur d'images et d'objets, telles que la classification de scènes, la reconnaissance et le suivi d'objet.

En premier lieu, les solutions apportées à ces problématiques se sont basées sur des statistiques globales mesurées sur les images, avant de voir apparaître ces dix dernières années des outils permettant l'utilisation de statistiques locales [94]. Ces descriptions sont rapidement devenues des outils incontournables de la vision par ordinateur.

Puisque l'essentiel de l'énergie du signal visuel est contenu dans le canal de luminance, ces descriptions s'appuient essentiellement sur l'information que porte ce canal [17]. De plus, la sensibilité de la couleur aux conditions d'acquisition complique son utilisation dans des conditions non contrôlées (capteurs différents, couleur de l'illuminant et matériaux utilisés). Néanmoins, dans l'optique d'améliorer les performances de ces descriptions, l'utilisation conjointe de représentations basées sur la luminance et sur la couleur a été proposée dans divers travaux.

Nous présentons ici les différents usages de la couleur qui ont été faits pour la description et la comparaison d'images. Pour cela, nous partons d'une modélisation du processus d'acquisition de l'image qui nous permet de mettre en évidence les sources de perturbations entrant en jeu. Nous présentons ensuite les principaux attributs couleur ainsi que leurs propriétés d'invariance. Nous introduisons enfin les différents modes de description du contenu couleur utilisés pour les comparaisons globales et locales d'images.

4.1 Attributs couleur invariants

Les capteurs numériques des appareils photo numériques classiques sont constitués de cellules photosensibles permettant de compter les photons qui les atteignent. Pour la plupart des appareils photographiques, un filtre permet de ne laisser passer que les photons correspondant à une certaine plage de longueurs d'ondes du domaine du visible. Les trois types de filtres utilisés correspondent grossièrement aux plages de rouge, de vert, et de bleu.

Ces mesures sont sensibles à de nombreuses variations des conditions d'illumination ou, plus généralement, des conditions d'acquisition. Par conditions d'illumination, nous comprenons les sources de lumière et leurs directions d'éclairage tandis que les conditions d'acquisition regroupent les conditions d'illumination ainsi que les caractéristiques de l'appareil photo et du point de vue. D'autres variations peuvent également survenir, comme un changement dans les propriétés de la surface observée.

Il est ici important de remarquer que ces mesures sont dépendantes de l'appareil photographique utilisé et particulièrement de ses fonctions de sensibilité. Or, l'étude d'attributs couleur invariants se limite aux perturbations liées aux conditions d'illumination ou aux propriétés des surfaces. La plupart des attributs couleur dépendent donc du capteur employé.

Dans le but de comprendre comment ces conditions d'illumination affectent les mesures effectuées par l'appareil photographique, nous présentons, à partir d'un modèle de formation des images couleur, les invariances possibles pour la comparaison couleur. Le modèle de formation des images couleur que nous utilisons est celui de Shaffer [134] dans sa version incluant l'influence de la lumière diffuse¹. Ce modèle peut être simplifié en faisant l'hypothèse que les fonctions de sensibilité de l'appareil peuvent être approchées par des fonctions de Dirac. Sous cette hypothèse, les valeurs RGB d'un pixel $\mathbf{m} = \{m_r, m_g, m_b\}$ sont données par

$$\underbrace{\begin{pmatrix} m_R \\ m_G \\ m_B \end{pmatrix}}_{\text{pixel}} = k \underbrace{\begin{pmatrix} e_R & 0 & 0 \\ 0 & e_G & 0 \\ 0 & 0 & e_B \end{pmatrix}}_{\text{illuminant}} \underbrace{\begin{pmatrix} s_R + \alpha_s/k \\ s_G + \alpha_s/k \\ s_B + \alpha_s/k \end{pmatrix}}_{\text{surface}} + \underbrace{\begin{pmatrix} a_R \\ a_G \\ a_B \end{pmatrix}}_{\text{lumière diffuse}} \quad (4.1)$$

Dans ce modèle, la réflectance de la surface est définie par $\mathbf{s} = \{s_R, s_G, s_B\}$, l'illuminant par $\mathbf{e} = \{e_R, e_G, e_B\}$ et la composante diffuse de la lumière par $\mathbf{a} = \{a_R, a_G, a_B\}$. La variable k modélise un changement d'intensité de l'illuminant et la variable α contrôle l'influence des réflexions spéculaires. Sous ce modèle, les variations des conditions d'illumination peuvent engendrer différentes perturbations sur \mathbf{m} , qui sont énumérées dans les points suivants :

- Un changement d'intensité peut être la conséquence d'une ombre, d'un dégradé ou encore d'une variation de la source lumineuse. C'est alors le paramètre k qui se trouve modifié.
- Un changement de la couleur de l'illuminant peut arriver lorsque la luminosité d'une pièce baisse, et que la lumière d'une lampe vient remplacer la lumière naturelle. Ces modifications sont prises en compte conjointement par k pour l'intensité et par \mathbf{e} pour le changement de couleur.
- Une modification de la couleur de la surface observée se manifeste par une modification de \mathbf{s} .
- Un changement du point de vue du photographe ou de la direction de l'éclairage va modifier ou faire apparaître des reflets. Cela correspond à un changement du paramètre α .
- Lorsqu'un objet est dans l'ombre, il n'est pas éclairé directement par la source de lumière mais par les rayons réfléchis par des objets voisins. La couleur de la lumière éclairant l'objet est alors une moyenne de la couleur de ces réflexions. Ces réflexions constituent la source de lumière diffuse \mathbf{a} dont une variation modélise un changement du contexte de l'objet observé. L'hypothèse de monde gris (voir la section 3.1) suppose

1. Ce modèle est présenté à la section 2.1.1.

qu'en moyenne le monde est achromatique. Cette hypothèse, quand elle est appliquée sur une scène, implique que cette source de lumière diffuse est de même couleur que l'illuminant.

Pour décrire la couleur d'une surface, on souhaite généralement mesurer \mathbf{s} , qui est une caractéristique de la surface. Elle est donc indépendante des conditions d'illumination. Or, tous les paramètres décrits ci-dessus influent sur la couleur mesurée par le capteur et empêchent ainsi de remonter de la mesure (\mathbf{m}) à cette propriété (\mathbf{s}) sans une connaissance suffisante des conditions d'illumination. À défaut, il est possible de calculer des quantités invariantes ou robustes au changement de conditions d'illumination. Le problème posé par ces mesures est qu'elles sont alors moins aptes à discerner deux surfaces et qu'un compromis entre invariance et pouvoir discriminant devient nécessaire. Le réglage de ce compromis est difficile et dépend de l'application visée.

4.1.1 Espaces de couleur avec propriétés d'invariance

Les espaces de couleur ont pour but de représenter une information spectrale après capture. La représentation choisie dépend de son objectif. Elle peut, par exemple, être guidée par la volonté de séparer les informations de luminance et de chrominance ou par le besoin de décorréler les composantes couleur ou encore par la recherche d'un système intuitif. Les espaces dont nous décrivons les propriétés d'invariance ont été présentés dans le chapitre 2.1.

Espace de chromaticité rg [155] Les espaces de chromaticité visent à la représentation de l'information couleur de manière indépendante de l'intensité lumineuse. Pour cela, la couleur est projetée sur un plan de luminance constante (tel que la somme $R + G + B$ soit constante). Par exemple, pour un triplet RGB, les composantes r , g et b sont

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}, \quad b = \frac{B}{R + G + B}. \quad (4.2)$$

Ainsi, $r + g + b = 1$. L'information chromatique peut alors être représentée par les seules composantes r et g , qui décrivent une proportion des couleurs primaires dont la somme est constante. Dans le cas où, dans l'équation (4.1), les termes \mathbf{a} et α sont négligés, ces composantes chromatiques sont invariantes à un changement d'intensité de l'illuminant.

Espaces antagonistes [155] Ce type d'espace représente l'information sous la forme d'une composante d'intensité et de deux composantes chromatiques exprimant des contrastes de couleur. La vision humaine, par exemple, utilise une représentation de ce type, qui peut être approchée par les formules suivantes

$$O_1 = \frac{1}{\sqrt{2}}(R - G), \quad O_2 = \frac{1}{\sqrt{6}}(R + G - 2B), \quad O_3 = \frac{1}{\sqrt{3}}(R + G + B). \quad (4.3)$$

En présence d'un illuminant neutre, c'est à dire vérifiant $e_R = e_G = e_B$, les composantes O_1 et O_2 sont invariantes aux reflets spéculaires. Elles sont de plus insensibles à une composante diffuse neutre ($a_R = a_G = a_B$). Elles restent néanmoins linéairement dépendantes d'un changement d'intensité de l'illuminant.

Pour obtenir l'invariance à un changement d'intensité, il est possible de normaliser les composantes O_1 et O_2 par la composante d'intensité O_3 , mais il faut alors renoncer à la propriété d'invariance à l'égard de la composante diffuse. Une telle normalisation est de plus instable aux faibles intensités.

4.1.2 Invariants pixéliques et leurs liens avec la représentation de la couleur

Certains canaux des espaces de couleur possèdent des propriétés d'invariance, mais de nombreux autres invariants ont été proposés dans la littérature. Nous présentons ici différents invariants calculés en chaque pixel de l'image.

Gevers *et coll.* [61] précisent les propriétés d'invariance des caractéristiques couleur habituelles dont les chromaticités rg , la teinte et la saturation. D'autres invariants au point de vue, à la géométrie de l'objet, ou à un changement de l'illumination sont proposés (direction, intensité, et spectre). Ces invariants sont obtenus à partir d'angles calculés entre les composantes RGB ainsi qu'en regardant différents rapports entre les canaux RGB, de même que des ratios entre composantes RGB de pixels voisins.

Van de Weijer *et coll.* [146, 147] suggèrent d'utiliser des invariants pour le calcul d'un tenseur couleur similaire à celui de Di Zenzo [35]. En considérant le modèle dichromatique (cf. Section 2.1.1) ainsi qu'en supposant connue la couleur de l'illuminant, ils montrent qu'il est possible de séparer un gradient RGB en trois vecteurs. Le premier représente les variations dues au changement de réflectance, le deuxième celles dues à une ombre ou à un dégradé et le troisième celles causées par une réflexion spéculaire. Connaissant les trois directions de variation possibles, des invariants orthogonaux aux changements d'intensité et aux réflexions spéculaires sont présentés.

Dans le modèle de couleur gaussien [59], le spectre couleur réfléchi par une surface est estimé à partir de fonctions de sensibilité de l'appareil et des valeurs RGB. Partant de ce modèle, Geusebroek *et coll.* décrivent dans [58] différents invariants parmi lesquels les invariants C' , H' et W' qui ont été repris dans la littérature. Ces invariants sont calculés dans un espace spécifique à l'appareil utilisé qui est très similaire à l'espace antagoniste mesurant des contrastes de couleur bleu/jaune et vert/rouge. En supposant que cet espace spécifique est l'espace antagoniste, alors, l'invariant H' correspond au ratio O_1/O_2 entre les deux composantes chromatiques. Il est donc relatif à la teinte dont il partage les propriétés d'invariance. L'invariant C' est, quant à lui, une normalisation des canaux antagonistes O_1 et O_2 par le canal d'intensité, soit O_1/O_3 et O_2/O_3 . Ces deux ratios possèdent ainsi les mêmes propriétés d'invariance que celles décrites dans la section 4.1.1 pour les canaux antagonistes normalisés. L'invariant W' s'interprète comme les fonctions dérivées des canaux couleur antagonistes normalisés par l'intensité. Selon Chu *et coll.* [24], l'invariant H' est trop peu discriminant car du fait de ses fortes propriétés d'invariance il élimine trop de structures dans l'image. À l'inverse, l'invariant W' ne l'est pas assez. Ils situent donc leur compromis entre invariance et pouvoir discriminant au niveau de la caractéristique C' . Il est important de noter que le modèle sur lequel sont construits ces invariants requiert une transformation linéaire à partir des valeurs RGB mesurées par l'appareil pour permettre l'estimation du spectre. Or, cette transformation repose sur la connaissance des fonctions de sensibilité de l'appareil, qui sont rarement connues. Cependant, les auteurs proposent une transformation moyenne et notent qu'en pratique les invariants proposés sont faiblement sensibles à cette transformation.

4.2 Descriptions globales d'images basées sur des statistiques couleur

Les invariants présentés au paragraphe 4.1 correspondent à différentes manières de représenter l'information couleur. Ces représentations peuvent être utilisées dans le but de décrire

le contenu couleur d'une image ou d'une imagerie. Nous nous intéressons ici aux descriptions globales d'images ou d'objets de la littérature.

4.2.1 Représentations normalisées des couleurs de l'image

Afin d'être robuste aux modifications de la couleur de l'illuminant, une alternative au calcul d'invariants pixéliques est de normaliser les couleurs de l'image préalablement à l'extraction de caractéristiques. Cette normalisation étant effectuée à un niveau plus global qu'à l'échelle du pixel, il est possible d'utiliser des *a priori* sur la distribution des couleurs.

Par exemple, cette idée est proposée en 1987 par Gershon *et coll.* [55] dans un article où ils utilisent un modèle linéaire de dimension finie pour décrire l'information mesurée par le capteur. Dans ce modèle, les réflectances et les sources lumineuses s'écrivent sous la forme d'une somme pondérée d'un ensemble de fonctions de base. Une estimation de l'illuminant par une approche de type monde gris (*Grey-World*) est alors adaptée à la description des surfaces et des illuminants pour permettre d'obtenir une image normalisée.

Une dizaine d'années plus tard, Finlayson [49] a proposé une normalisation de l'image indépendante de l'intensité ainsi que de la couleur de l'éclairage. Pour cela, une approche itérative composée de deux étapes est présentée. La première est une étape de normalisation vis-à-vis de l'intensité de l'illuminant et la seconde est une étape de normalisation à l'égard de la couleur de l'illuminant par une approche de type *Grey-World*. Ces deux étapes sont répétées jusqu'à convergence.

Un changement d'illuminant peut être modélisé par un changement de contraste et donc par l'application d'une fonction monotone croissante à chacun des canaux RGB. Remarquant cela, Finlayson *et coll.* [41] utilisent une mesure d'ordre effectuée sur chacun des canaux RGB comme invariant à ce type de transformation. Ils proposent alors d'appliquer une égalisation d'histogramme indépendamment sur chacun des canaux RGB pour obtenir une représentation invariante au regard de l'illuminant et robuste par rapport aux fonctions de sensibilité de l'appareil.

4.2.2 Statistiques simples et distributions de couleurs

Approches par histogramme

Les approches par histogramme visent à résumer une image par des distributions de couleurs. Cela permet d'obtenir à moindre coût des propriétés d'invariance vis-à-vis de modifications géométriques dans les images telles que les translations, les rotations ou encore les changements d'échelle.

Parmi les pionniers de ces approches, Swain et Ballard [141] ont utilisé des histogrammes 3D construits à partir des valeurs de pixels prises dans l'espace de couleurs antagonistes afin de décrire un objet. Ces approches par histogramme sont invariantes à de nombreuses modifications géométriques mais ne sont pas invariantes vis-à-vis des conditions d'illumination notamment par rapport à la couleur et à l'intensité de l'éclairage.

D'après la théorie Retinex [89], en supposant que les conditions d'illumination varient lentement et que le modèle de Von Kries [151] (cf. Équation (2.26)) est applicable, les ratios entre composantes RGB de pixels voisins sont indépendants de l'illuminant. Pour obtenir une description plus robuste aux variations de couleur de l'illuminant, Funt *et coll.* [52] utilisent

ainsi ces ratios, ou alternativement leurs dérivées, pour la construction des histogrammes, qui sont alors utilisés comme descripteurs d'images. De plus, il est remarqué que ces méthodes améliorent les performances de reconnaissance d'objet dans le cas de changement d'illumination tout en ne dégradant que peu les performances lorsque l'illuminant est identique. Cette technique est toutefois plus coûteuse en calcul et les ratios mesurés dans les zones de faibles intensités sont sujets au bruit. Une approche similaire est proposée par Nayar *et coll.* dans [114].

Les problèmes posés par les ratios de couleurs sont relevés par Stricker [140] qui note de plus que ces ratios ne peuvent être utilisés dans le cas d'images possédant des pixels saturés. Il remarque également que ces histogrammes mesurent à la fois des variations de couleur ainsi qu'une longueur des contours couleur. Comme alternative, il suggère l'emploi d'histogrammes des longueurs de contours couleur pour la description et la reconnaissance d'images. Chaque classe de l'histogramme contient alors une estimation de la longueur totale de la frontière entre deux couleurs. Pour obtenir une invariance à l'égard de la couleur de l'illuminant, il suggère d'appliquer un algorithme de constance de couleur sur l'image, préalablement à l'extraction des histogrammes.

Moments couleur

Les moments couleur permettent de résumer l'information de luminance et de couleur de manière compacte. Plusieurs invariants basés sur ces statistiques ont été proposés.

Healey *et coll.* [73] utilisent des moments invariants calculés à partir d'histogrammes de couleurs pour la reconnaissance d'objet en présence de changements d'illumination. Ces moments sont établis d'après un modèle linéaire de dimension finie, décrivant les fonctions de réflectance des surfaces à la manière de [55]. Ce modèle est vérifié en pratique [98].

Lenz *et coll.* [91] estiment la transformation couleur séparant deux images par le biais d'une matrice. Cette matrice est celle qui normalise un ensemble donné de moments calculés à partir de la distribution des couleurs d'une image. Mindru *et coll.* [107] proposent quant à eux des invariants basés sur la combinaison de moments couleur généralisés permettant ainsi une combinaison simple de la couleur et de la géométrie.

Autres statistiques

Variation couleur perpendiculaire au lieu de Planck Finlayson *et coll.* [46], utilisant à la fois l'équation de Planck comme modélisation de l'illuminant et l'hypothèse de capteurs à bandes étroites, montrent que dans l'espace log-chromaticité, un changement de température de couleur de l'illuminant se manifeste par une translation identique de tous les pixels. Sous l'hypothèse de capteurs à bande étroite, la relation liant les valeurs RGB mesurées de deux pixels \mathbf{m}_1 et \mathbf{m}_2 sous deux illuminants est : $(R_2 = \alpha R_1, G_2 = \beta G, B_2 = \gamma B)$

$$\left(\log\left(\frac{R_2}{G_2}\right), \log\left(\frac{B_2}{G_2}\right) \right) = \left(\log\left(\frac{R_1}{G_1}\right), \log\left(\frac{B_1}{G_1}\right) \right) + (a, b), \quad (4.4)$$

où $(a, b) = (\log(\alpha/\beta), \log(\gamma/\beta))$. Sous l'hypothèse que le spectre de l'illuminant suit l'équation de Planck, il est alors montré que le coefficient directeur de la translation (a, b) est fixé. Comme

cela a été présenté dans la section 2.2.1, l'équation de Planck peut être simplifiée. Le spectre d'un radiateur de Planck chauffé à une température T est alors donné par

$$L(T, \lambda) = c_1 \lambda^{-5} \exp\left(-\frac{c_2}{\lambda T}\right).$$

Sous l'hypothèse de capteurs RGB monochromatiques, le logarithme des valeurs mesurées d'un pixel p_k avec $k \in \{R, G, B\}$ est donné par

$$\log(p_k) = \log(S(\lambda_k)) + \log(L(T, \lambda_k)) = \log(c_1 \lambda_k^{-5}) - \frac{c_2}{\lambda_k T}. \quad (4.5)$$

En posant $S_k = \log(c_1 \lambda_k^{-5})$ et $E_k = -\frac{c_2}{\lambda_k}$ alors on obtient les rapports suivants

$$\begin{cases} p'_R = \log(p_R) - \log(p_G) = S_R - S_G + \frac{1}{T}(E_R - E_G) \\ p'_B = \log(p_R) - \log(p_B) = S_R - S_B + \frac{1}{T}(E_R - E_B) \end{cases} \quad (4.6)$$

Ainsi, un changement d'illuminant dans l'espace log-chromaticité déplace les coordonnées des pixels sur une droite dont le coefficient directeur ne dépend que des fonctions de sensibilité de l'appareil. Une caractéristique invariante au changement d'illuminant (et donc orthogonale à cette translation) est obtenue en projetant dans l'espace log-chromaticité les coordonnées des pixels sur la droite perpendiculaire à la translation. Cette caractéristique invariante est ensuite utilisée dans une expérience de reconnaissance d'objet.

4.3 Description locale d'images

Plus récemment, les approches locales ont connu un essor important car elles permettent une forte robustesse à une plus grande classe de transformations géométriques que les approches globales. Elles sont également plus robustes aux phénomènes d'occultation et de changement de contexte. Parmi ces approches, la méthode SIFT [94] a connu un succès important. Cette méthode de description n'utilise à l'origine que l'information de luminance et ne fait pas appel à la couleur. Diverses démarches ont été proposées pour prendre en compte cette information couleur.

Avant de regarder les réponses apportées par la littérature, nous décrivons ici la méthode SIFT. Puis nous détaillons les approches existantes utilisant les SIFT conjointement à une description couleur. Nous abordons ensuite les approches appliquant les descripteurs SIFT directement sur des invariants couleur.

4.3.1 La description SIFT et ses variantes

Nous présentons ici la méthode SIFT originale publiée par Lowe [94], ainsi que le principe de certaines variantes. Nous détaillons plus précisément celle introduite dans [125], qui sera utilisée par la suite.

L'algorithme SIFT original

La description SIFT repose sur la détection de points d'intérêt dans un espace échelle. À partir du contexte de chacun de ces points d'intérêt, une orientation principale ainsi qu'un descripteur est calculé. Une description sectorisée du contexte qui repose sur des histogrammes d'orientation du gradient est alors mise en œuvre.

Points d'intérêt L'espace échelle est défini pour une image I par l'équation suivante

$$I_\sigma(x, y) = G_\sigma * I(x, y), \quad (4.7)$$

où (x, y) sont les coordonnées du pixel considéré et σ l'échelle d'observation. Lindeberg [93] a montré que, dans cet espace, les extrema d'opérateurs différentiels normalisés constituent des points caractéristiques stables par changement d'échelle de l'image. En particulier, l'opérateur du Laplacien normalisé est défini par l'équation

$$\Delta_\sigma I(x, y) = \sigma^2 \Delta I_\sigma(x, y), \quad (4.8)$$

où Δ est l'opérateur Laplacien. En vertu de l'équation de la chaleur, il est possible d'approcher cet opérateur par une différence de gaussienne. Cette approximation est appelée DOG (Difference Of Gaussian). Ainsi, une pyramide d'images de plus en plus floues est construite et l'opérateur du laplacien est obtenu à partir de différences entre couches successives de cette pyramide. Les points correspondant à des maximums en espace et en échelle sont alors retenus. À chaque point est associée une échelle. Ces points, très nombreux, sont fréquemment situés à proximité de contours : cela les rend peu discriminants. Pour pallier ce problème, Lowe propose d'utiliser le critère de Harris [70].

Orientation principale Pour obtenir un descripteur invariant par rotation, l'idée utilisée par Lowe [94] est de décrire une région d'intérêt relativement à une orientation principale. Pour cela, à chaque point clé est associé un histogramme d'orientation du gradient pondéré par la magnitude (norme) du gradient. Cet histogramme est calculé sur voisinage du point clé dont la taille est proportionnelle à l'échelle du point clé. Dans la méthode de Lowe, le maximum de cet histogramme est utilisé comme orientation principale. De plus, chaque maximum secondaire dont l'intensité correspond au moins à 80% du maximum aboutit à la création d'un nouveau point clé dont ce maximum définit l'orientation principale.

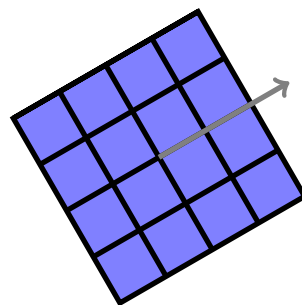


FIG. 4.1 Masque utilisé pour la description locale SIFT. La flèche représente l'orientation principale et la taille de la grille est déterminée par l'échelle de détection. Un histogramme d'orientation du gradient est calculé en chaque secteur de la grille.

Description Pour la description des régions d'intérêt, Lowe utilise une grille carrée de seize secteurs qui est représentée sur la Figure 4.1. Ce masque est orienté suivant la direction principale. En chacun des secteurs de ce masque est calculé un histogramme d'orientation du gradient pondéré par la norme ainsi que par une fonction de la distance au point clé. Une orientation contribue ainsi plus fortement à l'historgramme si elle est proche du point d'intérêt et si elle est associée à une forte norme du gradient. L'orientation du gradient est calculée

relativement à l'orientation principale pour assurer l'invariance par rotation. La pondération par la norme permet quant à elle de donner plus d'importance aux fortes transitions, ainsi que d'être robuste au bruit. Le descripteur final est obtenu en concaténant l'ensemble des histogrammes puis en normalisant cette concaténation de façon à ce que la somme du descripteur soit unitaire. Il est ainsi invariant par un changement de contraste affine sur l'ensemble de la région d'intérêt. Afin de comparer les descripteurs ainsi construits, la métrique euclidienne est utilisée.

Variantes et alternatives

En plus des variantes couleur détaillées de manière plus précise dans la suite, de nombreuses variations ont été proposées autour de cette méthode. Sans prétendre à l'exhaustivité, nous en présentons ici certaines.

Pour la détection de points clés, il a été proposé un critère de Harris multi-échelle [103]. Un détecteur invariant par transformation affine a été décrit dans [104]. Comme alternative à la détection de point clé, un détecteur de région d'intérêt a également été proposé par Matas *et coll.* [99].

Du côté des descriptions, des auteurs ont proposé d'améliorer le pouvoir discriminant du descripteur SIFT en appliquant une analyse en composantes principales [86] sur un ensemble de régions d'intérêt afin d'en apprendre l'espace propre. Chaque patch est ensuite décomposé dans cet espace et cette décomposition tient lieu de description. Le descripteur ASIFT a été développé afin d'obtenir une description invariante même pour de fortes transformations affines [110]. Des descripteurs binaires adaptés aux textures ont été introduits [118]. D'autres descripteurs binaires ont été utilisés plus récemment [19]. Les descripteurs SURF constituent une alternative complète et rapide à la méthode SIFT [7]. Récemment, il a été présenté un descripteur pour l'appariement dense [142].

4.3.2 Variante SIFT utilisée

Nous présentons ici la chaîne d'appariement de descripteurs locaux présentée par Rabin *et coll.* [124, 125]. Cet algorithme (noté SIFT dans la suite) est une variante de l'algorithme SIFT original décrit ci-dessus, nous en présentons les différences.

Ici, le calcul de la pyramide Laplacienne ne se fait pas en utilisant l'approche DOG mais en utilisant un schéma discrétisé du Laplacien. Les points clés retenus correspondent toujours aux extrema du Laplacien normalisé mais, le raffinement permettant d'éliminer les points de bords est réalisé à l'aide du critère de Harris multi-échelle proposé par Mikolajczyk *et coll.* [103]. Cette approche revient à appliquer le critère de Harris [70] à l'échelle considérée tout en tenant compte de cette échelle

$$C_{\sigma}(x, y) = \sigma^2 G_{\sigma} * \begin{bmatrix} \left(\frac{\partial I_{\sigma}}{\partial x}\right)^2 & \frac{\partial I_{\sigma}}{\partial x} \frac{\partial I_{\sigma}}{\partial y} \\ \frac{\partial I_{\sigma}}{\partial x} \frac{\partial I_{\sigma}}{\partial y} & \left(\frac{\partial I_{\sigma}}{\partial y}\right)^2 \end{bmatrix}. \quad (4.9)$$

Les valeurs propres de la matrice C décrivent la courbure au point (x, y) . Pour éviter les points de bords, le critère suivant est alors défini

$$\mathcal{C}(x, y) = \text{Det}(C_{\sigma}(x, y)) - l \cdot \text{Tr}(C_{\sigma}(x, y))^2 > t. \quad (4.10)$$

Dans le critère \mathcal{C} , Det est le déterminant de C et Tr la trace de C . En pratique, les valeurs suivantes sont utilisées : $l = 0.04$ et $t = 10^5$.

Pour détecter les orientations principales, l'algorithme de sélection de modes *a contrario* décrit dans l'annexe A est utilisé. Cet algorithme permet à la fois une sélection plus robuste des modes pour chaque point clé et rejette de plus l'utilisation de zones d'intérêt qui ne possèdent pas d'orientation privilégiée.

Pour la description des régions d'intérêt, Lowe utilise une grille carrée de 4×4 secteurs, nous utilisons ici un masque circulaire [9, 125] tel que représenté par la Figure 5.2.

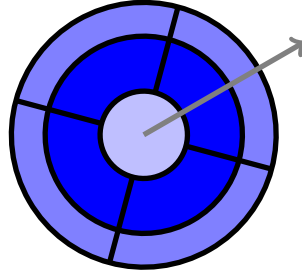


FIG. 4.2 Masque circulaire à neuf secteurs de tailles identiques.

Le masque est composé d'anneaux concentriques qui sont eux-mêmes subdivisés en sous-régions. Pour obtenir M régions de même taille, le diamètre de l'anneau r_i est donné par

$$r_i = \sqrt{\frac{s_i}{M} + r_{i-1}^2}, \quad (4.11)$$

où s_i est le nombre de régions composant l'anneau i et $r_0 = 0$. Augmenter le nombre de secteurs donne un descripteur plus discriminant mais aussi moins compact et moins robuste aux forts changements de point de vue. Pour la construction des histogrammes, la pondération n'est pas proportionnelle à la distance au point clé mais uniquement à la norme du gradient.

La métrique euclidienne ainsi que les autres distances bin à bin ne sont pas robustes à la quantification des histogrammes. Pour remédier à ce problème, une mesure de similarité alternative basée sur le calcul de coût de transport entre histogrammes a été proposée par Rubner *et coll.* [129]. Cette mesure est désignée par l'acronyme EMD pour Earth Mover's Distance. Elle n'est cependant pas adaptée aux histogrammes circulaires. Une variante tenant compte de cette spécificité et améliorant significativement les performances de la comparaison de descripteurs locaux a été proposée par Rabin *et coll.* [124].

4.3.3 Descripteurs SIFT combinés à des statistiques couleur

Van de Weijer *et coll.* [148] proposent d'utiliser différents descripteurs locaux couleur conjointement avec le descripteur SIFT. Ils utilisent en particulier, un histogramme de teinte pondéré par la saturation, ainsi que des invariants proposés dans [146]. En plus de leurs descripteurs, ils proposent de normaliser le *patch* à partir duquel les descripteurs couleur sont calculés en utilisant les normalisations de type *grey-world* et *grey-edge* [16, 145]. Trois évaluations sont proposées. L'une d'elle est basée sur l'appariement des descripteurs locaux, la seconde est basée sur la recherche d'images et la dernière sur la classification. Ils concluent que l'histogramme de teinte constitue un bon compromis pour les images saturées tandis que le descripteur basé sur les angles entre canaux antagonistes est conseillé pour les images aux couleurs moins saturées.

Parmi les approches faisant usage de la teinte, Zhang *et coll.* [156] emploient une agrégation de descripteurs couleur proposée par Gevers *et coll.* [60] dont un descripteur de trois histogrammes de teinte calculés à partir d'anneaux concentriques dans la région d'intérêt.

Une approche originale a été proposée par Ancuti *et coll.*. Ils suggèrent d'enrichir le descripteur SIFT à l'aide des histogrammes de cooccurrence de teinte proposée par Chang *et coll.* [22]. Les histogrammes de cooccurrence sont ici des histogrammes à trois dimensions. Deux d'entre elles servent à indexer la couleur et l'autre à indexer la distance. Est alors comptabilisé le nombre de fois qu'une distance est observée entre deux teintes données. Un problème évident de cette approche est le choix de la quantification pour les couleurs et la teinte.

Dans l'objectif d'obtenir un descripteur compact, Quelhas *et coll.* [123] ont proposé de combiner un descripteur SIFT ayant subi une analyse en composantes principales (ACP) et un descripteur concaténant la moyenne et la variance de chacun des canaux de l'espace de couleurs Luv calculés sur la région d'intérêt. Sur le même principe, il a été proposé de combiner un descripteur SIFT et un descripteur couleur contenant la moyenne de chaque canal RGB calculé sur la région d'intérêt [28]. L'image à partir de laquelle est extraite cette description couleur a subi au préalable une égalisation d'histogramme appliquée indépendamment sur chaque canal.

4.3.4 Descripteurs SIFT appliqués sur des canaux couleur

L'application directe de la méthode de description SIFT sur des canaux couleur est une alternative à la concaténation de descripteurs de SIFT et de descripteurs couleur. Ces canaux couleur peuvent être issus d'invariants couleur proposés dans la littérature ou bien, simplement correspondre aux canaux d'espaces de couleurs.

Parmi les approches proposant l'utilisation d'invariants couleur, Abdel-Hakim et Farag [1] ont les premiers employé l'invariant H proposé par Geusebroek *et coll.* comme base pour le calcul de descripteurs SIFT. Une évaluation systématique de l'ensemble des invariants proposés par Geusebroek *et coll.* [58] pour l'extraction de descripteurs locaux de type SIFT a ensuite été effectuée par Burghouts *et coll.* [18].

Reprenant cette idée, Bosch *et coll.* [13] ont proposé de calculer des descripteurs SIFT directement à des fins de classification de scène. Ils proposent d'utiliser des SIFT calculés sur chacun des canaux HSV. Le problème de cette approche est qu'elle ne prend en compte ni l'aspect circulaire de la teinte ni son instabilité pour les faibles valeurs de circulation. Un léger changement de couleur peut ainsi modifier très fortement la mesure du gradient de teinte.

Préférant le descripteur SURF [7] au descripteur SIFT, Chu *et coll.* [24], appliquent la description SURF sur les canaux de l'espace antagoniste ainsi que sur la caractéristique C [58]. Comme noté dans [144], cette caractéristique C peut être vue comme une normalisation des canaux couleur O_1 et O_2 par le canal d'intensité O_3 . Ils concluent en accord avec Van de Sande *et coll.* [144] que l'espace antagoniste, normalisé ou non par l'intensité, est le meilleur choix pour une description combinée de la couleur et de la géométrie.

Van de Sande *et coll.* [144] ont réalisé un comparatif de descriptions locales utilisant la couleur. Ils proposent de comparer la combinaison de descripteurs SIFT avec divers histogrammes de couleurs ainsi que l'application de la description SIFT sur des canaux couleur. Une étude détaillée des propriétés d'invariance est proposée. De manière intéressante, ils ne

s'intéressent pas aux invariances vis-à-vis des reflets spéculaires et considèrent un modèle lambertien auquel ils ajoutent une composante continue modélisant l'influence des réflexions diffuses.

Leurs évaluations portent sur des expériences de classification ainsi que sur des expériences d'appariement. Les expériences de mise en correspondance leur permettent notamment d'évaluer la sensibilité des descripteurs à une large gamme de variations. En conclusion, ils retiennent comme meilleur descripteur pour une application généraliste le descripteur SIFT appliqué sur les canaux antagonistes de l'espace *Opponent* (descripteurs notés Opponent-SIFT). D'autres descripteurs, issus de la méthode SIFT appliquée sur des canaux RGB (méthode notée RGB-SIFT) ou appliquée aux canaux antagonistes normalisés O_1/O_3 et O_2/O_3 se montrent également performants.

4.3.5 Autres approches

Certaines approches se distinguent de celles décrites jusqu'à présent. Par exemple, Matas *et coll.* [100] introduisent un descripteur basé sur la détection de zones contenant plusieurs couleurs. Cette détection est réalisée à l'aide de l'algorithme *mean shift* [23]. À partir de ces zones, ils calculent un ensemble d'invariants locaux dépendant des changements d'illumination considérés. Suivant ces changements, les descripteurs sont obtenus à partir de ratios de chromaticité, de valeurs brutes des canaux couleurs ou des combinaisons de ces informations. Une description de la couleur similaire et appliquée à la description d'objet est proposée par Naik *et coll.* [113].

Dans une approche couleur au sens large, Brown *et coll.* [15] proposent l'utilisation de SIFT couleur en intégrant le proche infrarouge. Pour cela, se basant sur une base image contenant un canal infrarouge, ils déterminent la transformation permettant de décorréler les quatre composantes (R, G, B, et infrarouge) à la manière de ce qui est réalisé pour obtenir l'espace antagoniste.

Geodeme *et coll.* [54] utilisent des moments couleur [107] pour réaliser un filtrage des correspondances aberrantes.

Khan *et coll.* [135] proposent d'utiliser la couleur comme carte d'attention afin de guider l'extraction de caractéristiques géométriques dans les zones qui contiennent probablement l'objet.

Dans le contexte de l'appariement dense de paires stéréo, des auteurs montrent que la couleur permet d'améliorer la qualité de la mise en correspondance [21, 112] tandis que récemment, d'autres auteurs relèvent que la couleur n'apporte que de faibles améliorations [75] ou même, détériore les résultats [12]. Dans une optique de suivi d'objet, des concaténations variées de descripteurs couleur ont été suggérées [27, 139, 153].

Bien entendu, la concaténation de descripteurs implique l'usage de descripteurs plus longs qui induisent un coût plus élevé en calculs. Une méthode couramment utilisée est la réduction de dimension par ACP. Cette approche a initialement été proposée dans ce contexte pour les descripteurs SIFT [86] puis a été fréquemment reprise pour les descripteurs couleur [28, 123, 144, 156]. Certaines descriptions compactes binaires ont également été étendues à la couleur. Par exemple, Zhu *et coll.* [157] proposent de calculer des descripteurs compacts de type LBP

(Local Binary Pattern) sur le canal de teinte ainsi que sur trois canaux antagonistes normalisés ou non par l'intensité.

Conclusion Deux catégories émergent parmi les approches proposées utilisant la couleur pour la comparaison d'images. La première applique la méthode SIFT sur des canaux couleur et la seconde concatène un descripteur couleur au descripteur SIFT. La première catégorie a l'avantage d'une forte robustesse, mais l'information contenue dans les différents canaux est souvent redondante. De plus, ces méthodes ne tirent pas pleinement parti de l'information couleur. En effet, en utilisant les canaux couleur séparément, ces descriptions ignorent la nature bidimensionnelle de la couleur. À l'inverse, la seconde propose des descripteurs avec un pouvoir discriminant réel mais ceux-ci sont souvent jugés peu robustes. Dans le chapitre suivant, nous décrivons une méthode robuste pour concilier ces deux approches.

Descripteurs locaux couleur : choix et appariement

Nous introduisons dans ce chapitre une méthode générique de mise en correspondance d'images exploitant une information couleur locale. Cette méthode repose sur de nouveaux descripteurs couleur possédant la particularité d'être complémentaires tant entre eux que par rapport à la description SIFT décrite dans la section 4.3.2.

Le principe de cette description est de coder d'une part une information couleur d'ordre zéro par l'intermédiaire de distributions couleur, et d'autre part, une information sur les transitions couleur par le biais d'opérateurs différentiels sur canaux couleur. Nous présentons également un critère d'appariement *a contrario* permettant de combiner de manière générique les différentes descriptions présentées.

5.1 Un nouveau descripteur de teinte

La couleur est une information bidimensionnelle et il est important de ne pas négliger cet aspect lors de la comparaison du contenu couleur de deux régions d'intérêt. De plus, la couleur brute est une information en soi complémentaire de celle portée par les transitions couleur.

Pour décrire une couleur, les attributs de teinte et de saturation sont les plus communs. Ainsi, ils sont largement utilisés en vision par ordinateur pour des applications telles que le suivi d'objet [121] ou comme décrit dans le chapitre 4 pour la reconnaissance d'objet à l'aide de descripteurs locaux [148, 156]. Nous présentons ici un descripteur couleur simple et performant basé sur une représentation locale de la teinte. Cette représentation locale est effectuée par l'intermédiaire d'histogrammes de teinte pondérés par la saturation. Ce descripteur possède des similitudes avec les descripteurs [148, 156].

Le descripteur proposé ici dispose également de nombreuses similarités avec le descripteur de type SIFT présenté à la section 4.3.2. Il est notamment calculé à partir des mêmes points clés et son masque est orienté suivant la même direction principale que celle détectée pour le descripteur SIFT.

Définition de la teinte Les définitions utilisées ici pour la teinte et la saturation sont celles données par les Équations (2.19) qui sont rappelées ici

$$H = \arctan\left(\frac{\sqrt{3}(R-G)}{R+G-2B}\right), \quad S = \sqrt{\frac{(R-G)^2}{2} + \frac{(R+G-2B)^2}{6}}, \quad L = \frac{R+G+B}{\sqrt{3}}.$$

D'autres définitions sont possibles, notamment pour la saturation. Néanmoins celle-ci est la plus robuste du fait que la dynamique de la saturation dépende de la luminosité. Une autre définition souvent utilisée est celle de l'espace HSV, dans lequel la saturation est définie relativement à l'intensité (cf. paragraphe 2.1.5). Ainsi, la dynamique du canal de saturation est indépendante de l'intensité. Cette définition retire donc l'information résiduelle d'intensité présente dans le canal de saturation (sans cette normalisation, une faible saturation est plus probable aux faibles intensités), ce qui engendre une instabilité dans les zones de faible luminosité. La définition donnée par l'Équation (2.19) permet d'éviter ce type de problème, illustré par la Figure 2.5.

Taille de la zone d'intérêt Le descripteur a pour but d'encoder l'information de teinte présente dans le voisinage du point clé. La taille de la région considérée est linéairement proportionnelle à l'échelle de détection du point clé, ce qui permet l'invariance vis-à-vis d'un changement d'échelle. Le facteur de taille utilisé ici est 12.

Normalisation des couleurs du support du descripteur La teinte étant particulièrement sensible aux changements de couleur de l'illuminant, il est nécessaire de la normaliser localement sur la zone d'intérêt servant de support à l'extraction du descripteur afin de rendre la description robuste. Différentes normalisations ont été proposées [148]. Ces normalisations ont pour but d'imposer certaines statistiques à la zone d'intérêt. Ici, les distributions de valeur de chacun des canaux RGB sont normalisées de sorte que leurs valeurs moyennes soient identiques. Considérons l'ensemble des M pixels m_i appartenant à un patch \mathcal{P} . À chacun de ces pixels correspond le vecteur constitué des valeurs RGB $\mathbf{p}(m_i) = (p_R(m_i), p_G(m_i), p_B(m_i))$. La composante $\bar{p}_c(m_i)$ avec $c \in \{R, G, B\}$ d'un pixel normalisé est donnée par

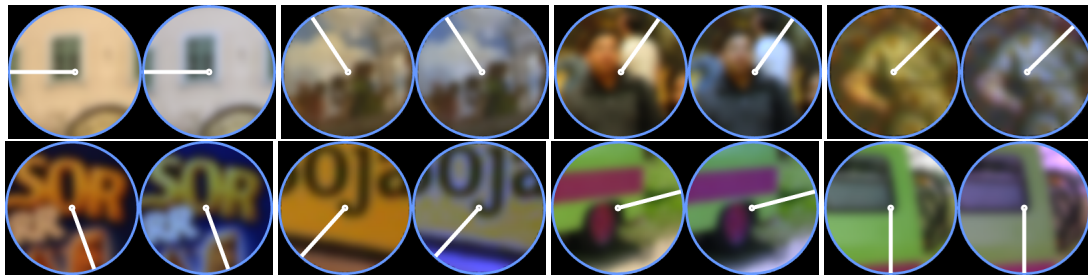
$$\bar{p}_c(m_i) = p_c(m_i) \frac{M}{\sum_{i=1}^M p_c(m_i)}. \quad (5.1)$$

On voit que la multiplication par une constante de l'ensemble des $p_c(m_i)$ n'affecte pas la composante normalisée $\bar{p}_c(m_i)$. Cela permet l'invariance par rapport à une transformation diagonale des valeurs RGB, ce qui constitue une approximation raisonnable pour un changement d'illuminant. L'intérêt d'effectuer cette normalisation localement est d'être robuste à la présence de multiples sources lumineuses dont la couleur est variable. Les effets de la normalisation sur le patch ainsi que sur le canal de teinte sont visibles sur les Figures 5.1 et 5.3.

Cette normalisation revient à appliquer l'algorithme dit de « *Grey-World* » [16] sur le support du descripteur. À cause des changements de points de vue ou des occultations éventuelles, les régions excentrées du patch sont plus sujettes aux changements de couleur. Pour améliorer la robustesse de la normalisation, il est possible de pondérer le calcul de la moyenne par une fonction de la distance au centre du patch.

Il est légitime de supposer que ce traitement affecte le pouvoir discriminant du descripteur. Nous verrons dans la partie expérimentale que la perte en pouvoir discriminant qui en résulte est en fait très faible, alors même que cette normalisation permet un fort gain en robustesse à l'égard des conditions d'illumination.

Histogrammes de valeurs de teinte pondérées par la saturation À chacun des pixels d'un des secteurs du masque (voir la Figure 5.2) correspond une teinte ainsi qu'une valeur de saturation. Un histogramme est construit pour chacun des secteurs. Chaque pixel contribue



(a) Patches isolés sans et avec normalisation



(b) Images originales

(c) Images reconstruites : hypothèse *Grey-World*(d) Images reconstruites : hypothèse *Grey-Edge*

FIG. 5.1 Sur la Figure 5.1(a), on peut observer les effets de la normalisation locale des couleurs sur quelques patches. Pour chaque exemple, le patch de gauche correspond au patch non normalisé et le patch de droite à sa version normalisée. Les images des Figures 5.1(b) ont été reconstruites à partir des patches extraits des images de la Figure 5.1(c) et 5.1(d). Ces patches ont été normalisés à l'aide de l'Équation (5.1) ou à l'aide d'une hypothèse de type *Grey-Edge* [145], qui suppose que la moyenne de la dérivée de l'image est achromatique (ici du patch). On peut observer que certains patches se retrouvent désaturés, tandis que dans d'autres de nouvelles couleurs apparaissent. Le flou introduit dépend de l'échelle de détection du point clé associé.

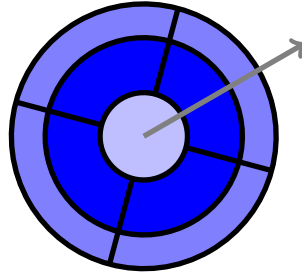


FIG. 5.2 Masque circulaire à neuf secteurs de tailles identiques.

à la classe de l'histogramme correspondant à sa teinte avec un poids correspondant à sa saturation. Ce vote permet ainsi une représentation conservant en partie l'aspect vectoriel de la couleur par un signal unidimensionnel compact. Une amélioration possible pour la construction d'histogrammes pondérés a été proposée dans [62]. Cette construction repose sur une propagation de l'erreur liée au bruit mesuré par le capteur.

Normalisation finale du descripteur La définition de la saturation par l'Équation (2.19) n'est pas indépendante de l'intensité. Ainsi, lors d'un changement de contraste linéaire d'un facteur k sur la zone d'intérêt, on peut observer que la saturation est elle-même multipliée par un facteur k qui augmente d'autant la masse du descripteur.

Le descripteur est normalisé de façon à ce que la somme de l'ensemble des histogrammes le composant soit unitaire. D'après le modèle de formation des images décrit par l'Équation (4.1) et la définition de l'espace HSL, la teinte et la saturation sont invariantes par l'ajout d'une constante sur les canaux RGB ($a_R = a_G = a_B$). Le descripteur final ainsi normalisé est donc invariant à un changement affine de l'intensité lumineuse sur la zone d'intérêt. Une illustration de l'information utilisée par ces descripteurs est donnée sur la Figure 5.3.

5.2 Descripteurs reposant sur l'espace couleur d'Ohta

Le descripteur de teinte que l'on vient de présenter permet de prendre en compte l'information de nature vectorielle propre à la couleur. Il ne décrit cependant ni la direction ni l'intensité des transitions couleur. Appliquer une méthode robuste telle que l'approche SIFT décrite dans la section 4.3.2 permet d'obtenir un descripteur encodant cette information.

Nous décrivons ici une variante des descripteurs Opponent-SIFT présentés à la section 4.3.4. Ces descripteurs ont récemment été présentés comme étant les plus efficaces des descripteurs locaux intégrant la couleur, et ce dans plusieurs opérations de classification [144]. Ces descripteurs sont obtenus en appliquant la description SIFT non seulement au canal de luminance mais également sur les canaux couleur d'un espace antagoniste. Rappelons que l'espace couleur antagoniste est obtenu à partir des composantes RGB de la manière suivante :

$$O_1 = \frac{1}{\sqrt{2}}(R - G), \quad O_2 = \frac{1}{\sqrt{6}}(R + G - 2B), \quad O_3 = \frac{1}{\sqrt{3}}(R + G + B). \quad (5.2)$$

Cet espace trouve son origine dans la vision humaine [17, 131]. Il est un moyen pour le système visuel des êtres humains, et des primates en général, de décorréler l'information couleur portée

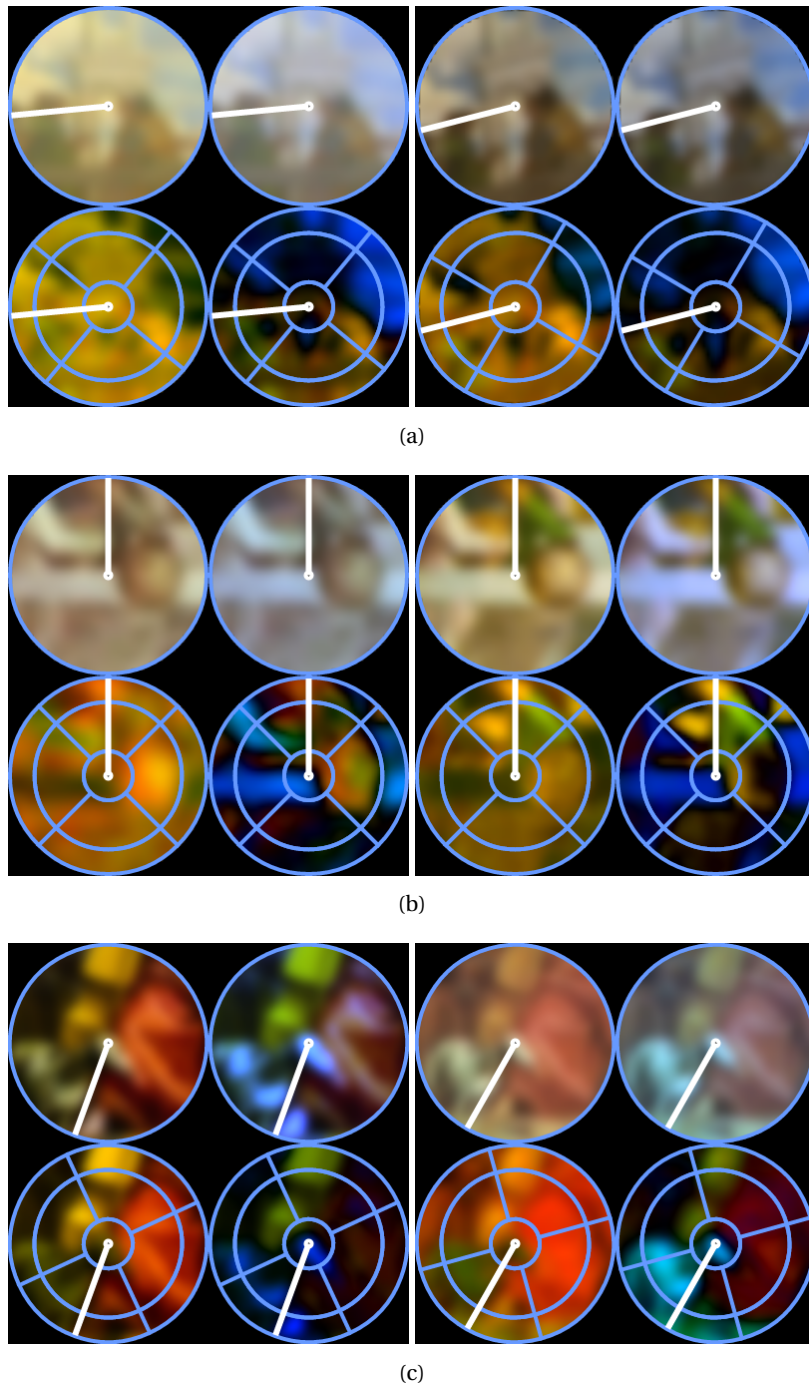


FIG. 5.3 Appariements de descripteurs de teinte extraits de différentes versions du tableau « Les noces de Cana » (cf. Figure 1.1). Pour chaque groupe de quatre patches, on retrouve de gauche à droite et de bas en haut : le patch original puis normalisé, le descripteur de teinte extrait du patch original puis normalisé. Les groupes de gauche et de droite proviennent d'images différentes retrouvées sur internet. Sur les deux premiers exemples, on observe bien l'effet de la normalisation qui supprime la « teinte moyenne ». On peut remarquer que la normalisation ne fonctionne pas pour le dernier appariement : le modèle diagonal n'est pas valide ici ; on peut supposer que les fonctions de sensibilité des deux appareils sont trop différentes.

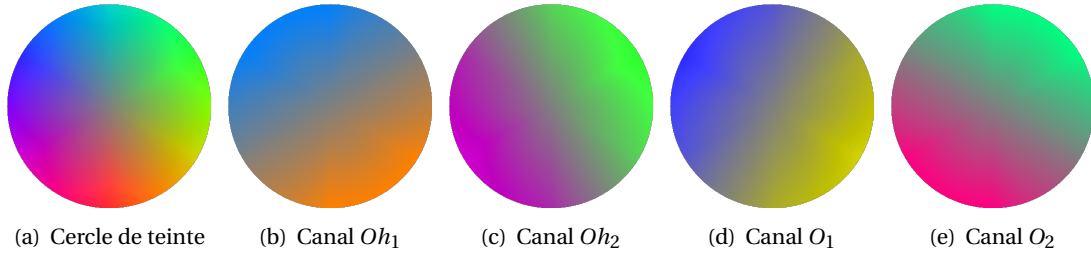


FIG. 5.4 L'image du cercle de teinte possède une saturation nulle en son centre et maximale sur les bords tandis que la teinte varie circulairement. Elle est décomposée soit sur les canaux antagonistes, soit sur les canaux de l'espace Ohta. On voit ainsi qu'il existe une rotation de $\pi/6$ entre les axes des deux espaces.

par les composantes RGB. Néanmoins, cet espace repose sur les fonctions de sensibilité de l'oeil humain. Celles d'un appareil photographique peuvent être très différentes [81].

Ici, nous proposons d'utiliser à la place l'espace Ohta mesurant des contrastes de couleur. Ohta *et coll.* [117] proposent d'utiliser les valeurs RGB moyennes de 100 régions obtenues par la segmentation de 8 images comme base pour une analyse en composantes principales. La transformation, qui décorrèle les canaux RGB, issue de ces expériences est la suivante :

$$Oh_1 = \frac{1}{\sqrt{2}}(R - B), \quad Oh_2 = \frac{1}{\sqrt{6}}(R + B - 2G), \quad Oh_3 = \frac{1}{\sqrt{3}}(R + G + B). \quad (5.3)$$

Nous appelons Ohta-SIFT les descripteurs obtenus à partir de cet espace en appliquant la méthode décrite à la section 4.3.2 sur les canaux Oh_1 et Oh_2 . Le descripteur calculé sur Oh_3 est lui identique au descripteur de type SIFT décrit précédemment. Ces descripteurs sont le résultat de la concaténation d'histogrammes d'orientation du gradient. Ils encodent ainsi une information de nature géométrique, qu'elle soit basée sur le canal de luminance ou sur les canaux couleur.

Bien que calculée sur des canaux couleur, l'information des canaux Oh_1 et Oh_2 comporte une part de redondance avec l'information portée par le descripteur SIFT (Oh_3). En effet, il est très rare que dans une image, une transition entre deux couleurs se fasse à luminance constante. Des exemples de descripteurs sont donnés sur la Figure 5.5.

Ainsi, l'information de transition est souvent déjà contenue dans le descripteur SIFT calculé sur le canal de luminance. L'information contenue dans les canaux Oh_1 et Oh_2 vient donc *renforcer* l'information du descripteur Oh_3 . L'information de phase étant fortement corrélée à l'information portée par le descripteur SIFT, lors d'une transition de luminance, les valeurs des phases calculées sur les trois canaux sont souvent identiques à plus ou moins π près. Cette réflexion s'applique bien entendu également tant aux descripteurs Ohta-SIFT qu'aux descripteurs Opponent-SIFT.

Avantages de l'espace Ohta L'espace antagoniste classique est basé sur la vision humaine et dépend donc des fonctions de sensibilité de l'oeil humain. Or, ces fonctions sont en général différentes de celles d'un appareil photographique. L'espace Ohta est quant à lui obtenu à partir

des fonctions de sensibilité de l'appareil [117]. Une étude récente a montré que les fonctions de sensibilité des appareils étaient peu variables [81].

De plus, les deux canaux couleur de cet espace encodent les contrastes rouge/bleu et vert/magenta (contre bleu/jaune et rouge/vert pour l'espace antagoniste). Or, le contraste rouge/bleu correspond à la direction du lieu de Planck tandis que le contraste vert/magenta correspond à une variation orthogonale au lieu de Planck. Utiliser une telle représentation est utile lors d'un changement de couleur de l'illuminant. Nous reviendrons sur ce point dans la partie expérimentale.

5.3 Validation d'appariements *a contrario*

Nous cherchons dans cette partie à appairer des points clés caractéristiques de la présence de structures d'intérêt dans l'image. Ces points clés sont choisis pour être invariants aux changements d'échelle et robustes à d'autres phénomènes tels que le bruit, les variations de point de vue ou les variations des conditions d'illumination. Dans le cas général, l'appariement de points clés consiste à établir une ou plusieurs correspondances entre un point clé requête et un ensemble de points clés candidats. Dans le cas de la comparaison entre deux images, le but est alors de trouver des appariements entre les points clés des deux images.

La mise en correspondance de descripteurs locaux se déroule en deux étapes. Premièrement, une mesure de dissimilarité est calculée pour chaque paire descripteur requête-descripteur candidat. Puis, étant données ces mesures, un critère de décision est appliqué pour décider si certains descripteurs sont suffisamment similaires pour que leurs points clés soient appariés.

Au paragraphe 5.3.1, nous rappelons les différents critères de décision existant dans la littérature. L'un d'entre eux, proposé par Rabin *et coll.* [125] repose sur une estimation de la distribution des distances entre le descripteur requête et la base de descripteurs candidats. Nous introduisons dans la section 5.3.2 une généralisation de cette procédure aux cas où plusieurs sous-descripteurs sont utilisés.

5.3.1 Critères existants

Un critère d'appariement a pour but d'attribuer une mesure de qualité à chaque appariement testé. En fixant un seuil, il est alors possible de valider les correspondances en comparant leur mesure de qualité à un seuil fixé.

Le critère le plus simple consiste à utiliser la distance mesurée entre deux descripteurs comme mesure de qualité. Ce choix pose problème, car la distribution des distances à un descripteur est fortement dépendante du descripteur en question. Un seuil fixe sur la distance est donc incapable de s'adapter à la région requête comme aux régions candidates. Du fait de cette incapacité, le seuil est très instable et donc difficile à régler. Il est possible d'améliorer quelque peu cela en faisant l'hypothèse que la base de descripteurs candidats ne contient tout au plus qu'une version modifiée du descripteur requête. Il est ainsi possible de ne garder que le plus proche voisin (candidat ayant la plus faible distance au descripteur requête) ce qui permet d'éliminer les correspondances multiples. Ce critère est appelé NN-DT (pour Nearest Neighbor - Distance Threshold).

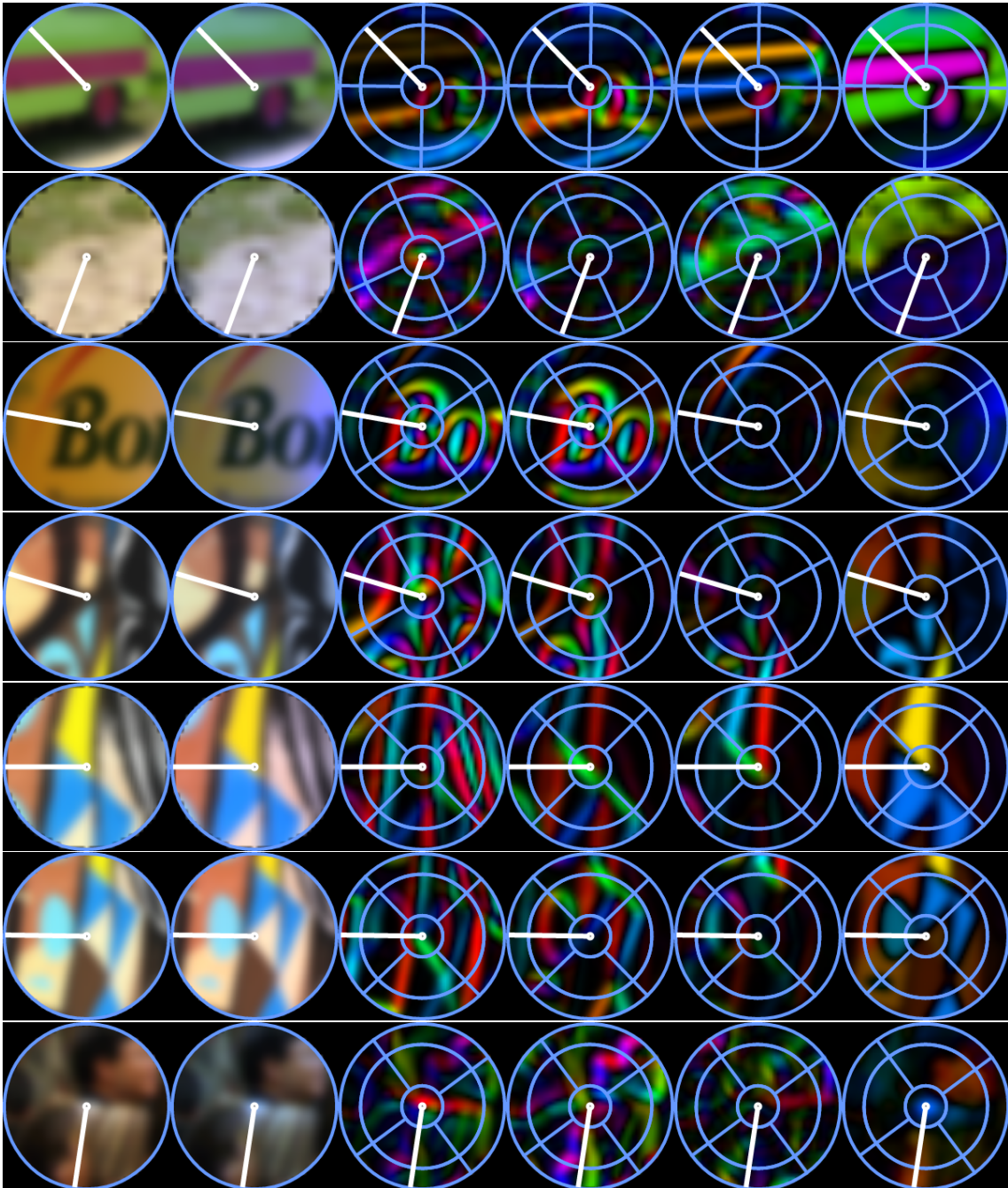


FIG. 5.5 Exemples de descripteurs extraits. Pour chaque exemple, on retrouve de gauche à droite : le patch RGB, sa version normalisée, le descripteur SIFT, les descripteurs OHTA 1 et OHTA 2 et à la fin le descripteur HUE-NORM. Pour les descripteurs de gradient, la teinte correspond à la phase du gradient relativement à l'orientation principale et l'intensité à la norme du gradient. On peut observer la corrélation existante entre les différentes descriptions basées sur le gradient.

Un critère astucieux s'adaptant à la distribution entre la requête et les candidats a été proposé par Lowe [94]. Ce critère consiste à utiliser le ratio entre la distance au plus proche voisin et la distance au deuxième plus proche voisin comme mesure de qualité entre deux correspondances. Ainsi, la qualité du plus proche voisin repose sur sa capacité à se détacher des autres candidats. Cette approche peut être vue comme un test élémentaire sur la distribution des distances. En pratique, cela permet d'améliorer grandement la stabilité du seuil ainsi que de rejeter les structures répétitives ou peu discriminantes. Ce critère est donc limité aux occurrences uniques de la requête. Il est par conséquent inadapté à la mise en correspondance multiple. Nous le notons NN-DR (pour Nearest Neighbor - Distance Ratio).

5.3.2 Critère *a contrario*

L'utilisation du critère NN-DR décrit précédemment, implique d'ignorer beaucoup de bonnes correspondances potentielles en rejetant les structures répétitives (telles que les fenêtres d'un bâtiment), ce qui rend difficile la détection d'objets apparaissant plusieurs fois. Pour remédier à cela, Rabin *et coll.* [125] proposent une approche statistique s'appuyant sur une estimation fine de la distribution des distances au descripteur requête. Une connaissance précise de cette distribution alliée à un test d'hypothèse permet alors de mesurer le nombre de fausses alarmes associé à chaque correspondance. C'est ce nombre de fausses alarmes qui sert de mesure de qualité. Il a été montré que cette mesure est plus robuste que la mesure obtenue en regardant le ratio des deux plus proches voisins, notamment lors de la comparaison d'une image contre une base de données de taille croissante [125]. De plus, ce critère, en regardant la distribution des distances dans sa globalité, dispense de la restriction au plus proche voisin. La détection de plusieurs occurrences du même objet dans une image en est ainsi facilitée. Dans le cas où seul le plus proche voisin est recherché, un appariement correspondant à une structure répétitive n'est plus systématiquement rejeté.

Critère d'appariement *a contrario*

Dans cette section, nous détaillons la procédure de mise en correspondance introduite dans [125] pour la comparaison de descripteurs de type SIFT. Nous l'étendons ensuite au cas où plusieurs descripteurs de différentes natures (tels que les descripteurs couleur décrits précédemment) sont utilisés.

Cette approche s'inspire de la méthodologie *a contrario* [33] pour déterminer une mesure de qualité des appariements permettant de sélectionner automatiquement les descripteurs à appairer. Cela permet d'utiliser exactement la même procédure avec une large variété d'images, d'une manière plus stable que le critère NN-DR classique et sans restriction au plus proche voisin.

Pour présenter le principe de la procédure, nous nous plaçons dans le cas où l'on cherche à appairer un descripteur \mathbf{a} associé au point clé requête à un second descripteur présent dans une base de données de N_C descripteurs candidats (chacun étant également associé à un point clé). Ces candidats sont notés \mathbf{b}^j avec $j \in \{1, \dots, N_C\}$.

Afin de déterminer automatiquement les seuils de mise en correspondance, l'idée est de calculer la loi de probabilité de la distance entre un descripteur requête \mathbf{a} et un descripteur aléatoire générique \mathbf{b} . Le seuil de mise en correspondance T peut alors être choisi tel que la probabilité $P(D(\mathbf{a}, \mathbf{b}) \leq T)$ soit petite. Ainsi, seuls les descripteurs dont la proximité est diffici-

lement attribuable à la chance sont appariés. Précisons maintenant la notion de descripteur générique aléatoire.

Descripteur générique aléatoire et Hypothèse nulle \mathcal{H}_0 Suivant [125], nous supposons que la distance $D(\mathbf{a}, \mathbf{b})$ entre un descripteur requête et un descripteur candidat peut s'écrire comme une somme de L distances

$$D(\mathbf{a}, \mathbf{b}) = \sum_{\ell=1}^L d_{\ell}(\mathbf{a}, \mathbf{b}). \quad (5.4)$$

Dans cette expression, les $\{d_{\ell}(\mathbf{a}, \mathbf{b})\}_{1 \leq \ell \leq L}$ désignent un ensemble de mesures de dissimilarités entre les descripteurs \mathbf{a} et \mathbf{b} .

Pour un descripteur requête \mathbf{a} donné, un descripteur générique aléatoire \mathbf{b} suit l'hypothèse nulle \mathcal{H}_0 si les distances $d_{\ell}(\mathbf{a}, \mathbf{b})$, pour $\ell = 1, \dots, L$, sont des variables aléatoires mutuellement indépendantes.

Nombre de fausses alarmes (NFA) L'hypothèse \mathcal{H}_0 définit le comportement des distances entre une correspondance aléatoire et un descripteur requête \mathbf{a} . Cela permet ainsi de définir *a contrario* une correspondance pertinente par la mesure du rejet de cette hypothèse. Cette mesure de rejet est appelée significativité. Sous l'hypothèse \mathcal{H}_0 , la probabilité d'observer une distance $D(\mathbf{a}, \mathbf{b})$ plus petite qu'un seuil T est définie par

$$P(D(\mathbf{a}, \mathbf{b}) \leq T) = \int_0^T \star_{\ell=1}^L p_{\ell}(x) dx, \quad (5.5)$$

où p_{ℓ} est la densité de probabilité de la variable $d_{\ell}(\mathbf{a}, \mathbf{b})$ et \star désigne le produit de convolution. Pour calculer cette probabilité, les densités de probabilité p_{ℓ} sont apprises empiriquement par les distributions des sous distances $d(\mathbf{a}_k, \mathbf{b}_k^j)_{j \in \{1, \dots, N_C\}}$ avec \mathbf{b}_k^j parcourant les descripteurs candidats.

Il est possible de seuiller cette probabilité pour séparer les bonnes correspondances des mauvaises. Cependant, cette probabilité étant indépendante du nombre de descripteurs candidats (N_C), il suffit d'utiliser une base de descripteurs candidats suffisamment grande pour trouver une distance ayant une probabilité d'occurrence satisfaisant le seuil.

Pour pallier ce problème, une solution classique des méthodes *a contrario* est de pondérer cette probabilité par le nombre de descripteurs testés contre l'hypothèse \mathcal{H}_0 . Cette pondération mène à définir le *nombre de fausses alarmes*. Il est donné pour deux descripteurs \mathbf{a} et \mathbf{b} par

$$\text{NFA}(\mathbf{a}, \mathbf{b}) = N_{\text{Test}} \cdot P(D(\mathbf{a}, \mathbf{b}) \leq T), \quad (5.6)$$

où N_{Test} est le nombre de mises en correspondances possibles pour l'expérience souhaitée.

Nombre de tests et choix du seuil sur le NFA Si N_Q descripteurs requêtes sont recherchés dans une base de N_C candidats, alors le nombre de test est $N_{\text{Test}} = N_Q \cdot N_C$.

Un seuil de validation ϵ est choisi de façon à contrôler le nombre de fausses alarmes. Ainsi les descripteurs \mathbf{a} et \mathbf{b} sont appariés si

$$\text{NFA}(\mathbf{a}, \mathbf{b}) < \epsilon. \quad (5.7)$$

De cette manière, et si l'hypothèse nulle est vérifiée, on peut alors montrer que l'espérance du nombre de faux appariements obtenus en testant un descripteur requête contre N_{Test} candidats est borné par ϵ

$$\mathbb{E}[\#\{(\mathbf{a}, \mathbf{b}), \text{NFA}(\mathbf{a}, \mathbf{b}) < \epsilon \mid \mathcal{H}_0\}] \leq \epsilon. \quad (5.8)$$

5.3.3 Regroupement des distances pour le respect de \mathcal{H}_0

Dans notre cas, chaque descripteur requête \mathbf{a} (resp. candidat \mathbf{b}) est composé de plusieurs sous-descripteurs ($a^\theta, \theta \in \Theta$), obtenus à partir de différents canaux couleur. Par exemple, on peut avoir $\Theta = \{\text{oh}_1, \text{oh}_2, \text{oh}_3, \text{hue}\}$. Chaque sous-descripteur a^θ est lui-même composé de M histogrammes $\{a_m^\theta, m = 1, \dots, M\}$, l'indice m correspondant à la sous région du support sur lequel a_m^θ est calculé¹. Pour une gestion fine de l'importance à accorder à la dissimilarité entre histogrammes homologues, une pondération w_m^θ est associée à chaque dissimilarité $\delta_m^\theta = d(a_m^\theta, b_m^\theta)$. La distance entre deux descripteurs, notée $D(\mathbf{a}, \mathbf{b})$, peut donc s'écrire comme la somme des dissimilarités entre histogrammes homologues, soit

$$D(\mathbf{a}, \mathbf{b}) = \sum_{\theta \in \Theta} \sum_{m=1}^M w_m^\theta \delta_m^\theta. \quad (5.9)$$

Les différentes possibilités de regroupement des distances

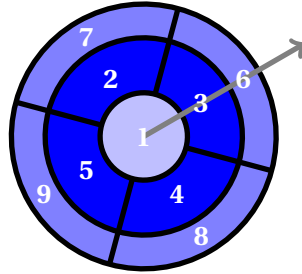


FIG. 5.6 Masque circulaire à neuf secteurs de tailles identiques.

Dans la partie précédente, nous avons décrit le critère d'appariement *a contrario*. Nous allons maintenant présenter les différentes possibilités pour regrouper les distances provenant des descripteurs décrits dans les sections 5.2 et 5.1.

On souhaite écrire la distance $D(\mathbf{a}, \mathbf{b})$ comme la somme de L distances $d_\ell(\mathbf{a}, \mathbf{b})$ vues comme des variables aléatoires indépendantes, afin de satisfaire \mathcal{H}_0 et de pouvoir écrire l'Équation 5.5. Remarquons que si l'on choisissait $d_\ell(\mathbf{a}, \mathbf{b}) = w_m^\theta \delta(a_m^\theta, b_m^\theta)$, on supposerait que les distances $\delta(a_m^\theta, b_m^\theta)$ sont indépendantes, ce qui est manifestement faux (*a priori* les distances associées à des secteurs voisins sont corrélées ainsi que les distances associées au même secteur mais correspondant à deux canaux couleur différents). On décrit dans ce qui suit différentes combinaisons possibles pour regrouper les distances $\delta_m^\theta, m = 1, \dots, M, \theta \in \Theta$.

Une distance par secteur Une première option est de regrouper les distances suivant leurs dispositions spatiales, c'est-à-dire en moyennant les distances issues de la comparaison des

1. Si $M = 9$ et $\Theta = \{\text{oh}_1, \text{oh}_2, \text{oh}_3, \text{hue}\}$, alors chaque descripteur est composé de 36 histogrammes.

différents descripteurs θ mais associées au même secteur géographique m . Les expressions des distances d sont ainsi :

$$d_m(\mathbf{a}, \mathbf{b}) = \sum_{\theta \in \Theta} w_m^\theta \delta_m^\theta, \quad m \in [1, \dots, M]. \quad (5.10)$$

Cela revient à supposer que les distances entre secteurs sont indépendantes. C'est la supposition proposée dans [125] pour l'appariement de descripteurs locaux SIFT. Si cette hypothèse est acceptable pour les descripteurs oh_i , elle ne l'est pas pour le descripteur de teinte. En effet, il est fréquent que les secteurs d'un descripteur soient d'une couleur identique. Dans ce cas, la corrélation entre les distances sera renforcée.

Une distance par descripteur Une seconde option est de regrouper les distances δ_m^θ non plus suivant le secteur m mais suivant le descripteur θ . Dans ce cas chaque distance d est donnée par la somme des distances calculées sur chacun des $L = 4$ sous-descripteurs soit

$$d_\theta(\mathbf{a}, \mathbf{b}) = \sum_{m=1}^M w_m^\theta \delta_m^\theta, \quad \theta \in \{oh_1, oh_2, oh_3, hue\}. \quad (5.11)$$

L'inconvénient de cette approche est que les trois distances finales basées sur des descripteurs $\theta \in \{oh_1, oh_2, oh_3\}$ d'orientation du gradient sont corrélées. C'est en revanche un choix raisonnable dans le cas où $\Theta = \{oh_3, hue\}$ (avec donc $L = 2$).

Approches hybrides Nous proposons ici deux approches reposant sur la séparation des distances provenant de la géométrie, qu'elle soit couleur ou non, ainsi que de la teinte.

Pour la première méthode, on moyenne d'une part les distances provenant d'histogrammes d'orientation du gradient pondéré suivant leur localisation (une distance par secteur) afin de créer M variables d_m et d'autre part, on moyenne les distances provenant du descripteur de teinte afin de créer une distance d_{M+1} . Ainsi, les variables $d_m(\mathbf{a}, \mathbf{b})$ correspondent à 9 distances décrivant la géométrie issue du canal de luminance et des canaux couleur, et une distance couleur provenant du descripteur de teinte. Les distances d sont alors :

$$d_m(\mathbf{a}, \mathbf{b}) = \sum_{\theta \in \{oh_1, oh_2, oh_3\}} w_m^\theta \delta_m^\theta, \quad m \in [1, \dots, M], \quad (5.12a)$$

$$d_{M+1}(\mathbf{a}, \mathbf{b}) = \sum_{m=1}^M w_m^{hue} \delta_m^{hue}. \quad (5.12b)$$

Cette approche reprend l'hypothèse d'indépendance des distances provenant de localisations spatiales différentes. Elle suppose que ces $L = M + 1$ distances sont indépendantes de la somme des distances calculées à partir du descripteur de teinte.

Pour la seconde méthode, on moyenne toutes les distances provenant de sous-descripteurs géométriques de façon à former une première distance d et comme dans l'approche précédente, les distances provenant du descripteur de teinte sont moyennées entre elles. Les distances sont alors :

$$d_1(\mathbf{a}, \mathbf{b}) = \sum_{\theta \in \{oh_1, oh_2, oh_3\}} \sum_{m=1}^M w_m^\theta \delta_m^\theta. \quad (5.13a)$$

$$d_2(\mathbf{a}, \mathbf{b}) = \sum_{m=1}^M w_m^{hue} \delta_m^{hue}. \quad (5.13b)$$

L'hypothèse d'indépendance faite ici est sans doute la plus réaliste puisqu'elle ne suppose aucune indépendance entre les secteurs mais seulement une indépendance entre la distance géométrique totale et la distance entre descripteurs de teinte (ici, $L = 2$).

Dans toutes les combinaisons précédentes, si les poids w_m^θ sont tous égaux, alors toutes les distances auront le même poids dans le calcul final de la distance entre les descripteurs **a** et **b** (voir l'Équation (5.4)).

Expérimentations sur les descripteurs locaux couleur

Dans ce chapitre, nous évaluons les comportements des descripteurs locaux couleur ainsi que celui du critère de validation présentés au chapitre précédent. Cette évaluation est réalisée au travers d'expériences synthétiques et réelles. Nous présentons d'abord les protocoles expérimentaux et les bases de données utilisées, puis des expériences illustrant le comportement des différentes contributions.

6.1 Création de la vérité terrain

Le but est la mesure des performances des descripteurs locaux et du critère proposés, dans le cadre de l'appariement de descripteurs locaux. Pour cela, il est nécessaire de construire des expériences pour lesquelles nous disposons d'une vérité terrain, c'est à dire dans notre cas, des mises en correspondance considérées comme correctes. Cette vérité terrain peut être obtenue à partir de couples d'images réelles ou bien à l'aide d'une image réelle et d'une version transformée de cette image. Ces deux types de vérité terrain sont utilisées dans les expériences présentées ici. Nous définissons également une mesure permettant d'évaluer dans quelles proportions les descripteurs et le critère proposés permettent un résultat proche de celui donné par la vérité terrain.

6.1.1 Expériences synthétiques

Les expériences synthétiques ont pour but l'évaluation des performances dans des conditions parfaitement contrôlées. Pour cela, nous utilisons les 100 premières images¹ de la base utilisée par Rabin *et coll.* dans [125]. Ces 100 images sont d'une résolution de 800×600 et reproduisent des scènes variées acquises sous diverses conditions d'illumination, en intérieur comme en extérieur. De ces images sont extraits $25 \cdot 10^4$ points clés soit en moyenne 2500 points clés par image.

Protocole $A \rightarrow A'$

Pour ce protocole, nous cherchons à comparer une image A avec une image A' obtenue à partir de A par une transformation synthétique. Puisque nous cherchons à étudier des descripteurs couleur, les transformations imposées sont à la fois géométriques et colorimétriques.

1. Cette base en contient 732, mais les résultats obtenus sur les 100 premières images sont représentatifs de ceux obtenus sur la base entière.

Ce protocole permet d'évaluer la capacité à retrouver une région d'intérêt dans une image où elle se trouve. Ce protocole est couramment employé pour l'évaluation de descripteurs locaux [105, 109, 125]. Néanmoins il ne permet pas de tester la capacité d'une méthode d'appariement à ne pas appairer dans le cas où la structure d'intérêt est absente.

Protocole $A \rightarrow A', B$

Dans ce protocole, l'image A est comparée successivement à la version transformée A' puis à une image B indépendante de A . L'image A' est obtenue de la même manière que pour le protocole $A \rightarrow A'$. Dans la comparaison $A \rightarrow B$, il n'existe pas de bon appariement. Ce protocole nous permet d'illustrer la capacité du critère ainsi que celle des descripteurs à rejeter des distracteurs.

Ce protocole est différent de celui qui consiste à comparer l'image A contre une image résultant de la concaténation d'une image A' et d'une image B . En effet, dans le cas présenté ici, deux appariements potentiels sont associés à chaque requête. Le premier peut être bon ou faux, le second est nécessairement faux. On évalue ainsi la capacité d'un critère à séparer les bonnes correspondances des mauvaises dans un cas où beaucoup de faux appariements potentiels sont présents.

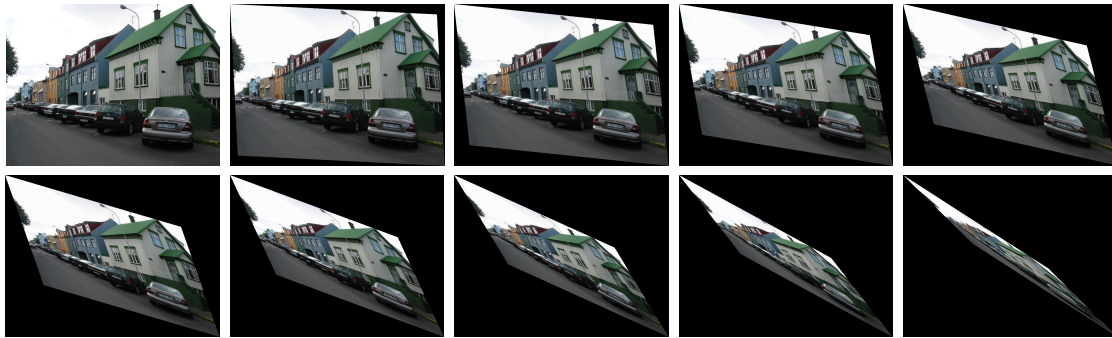
Transformations synthétiques considérées

L'image A' est obtenue à partir d'une transformation géométrique de l'image A suivie d'une éventuelle transformation colorimétrique ainsi que de l'ajout de bruit. La transformation colorimétrique consiste à modifier artificiellement la balance des blancs de l'appareil photographique. Pour cela, l'image est supposée correctement calibrée, c'est-à-dire éclairée par un illuminant ayant une température de couleur de $6500K$. Une dizaine de changements de températures de couleur sont alors appliqués à chacune des images. Les températures de couleur sont choisies dans la gamme $2000K - 20000K$. Les transformations géométriques appliquées aux images sont des transformations affines correspondant à un changement de tilt compris entre 1 et 10. Les descripteurs SIFT sont connus pour être quasiment invariants pour un tilt allant jusqu'à 2 [110]. Le bruit gaussien appliqué possède un écart type $\sigma = 5$. Un exemple des transformations appliquées est donné par la Figure 6.1.

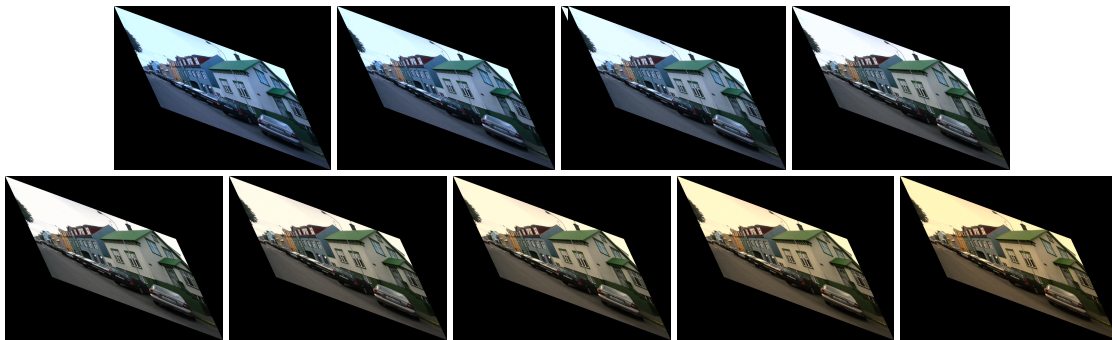
6.1.2 Expériences en conditions réelles

Le protocole décrit permet d'évaluer les descripteurs ainsi que les critères d'appariement présentés dans des conditions réelles. Dans ce paragraphe, le changement de point de vue n'est plus simulé, mais les images sont capturées sous différents angles de vue. Pour permettre d'établir une vérité terrain, les scènes sont choisies pour contenir une structure plane couvrant en grande partie l'image. De cette façon, la transformation séparant les deux scènes peut être convenablement approchée par une homographie. Une fois cette transformation estimée, il est alors possible de définir une correspondance valide ou rejetée.

Base d'images réelles Pour les besoins des expériences réelles, treize jeux de photographies ont été acquis. Toutes les images ont été prises à l'aide d'un appareil photo Canon 7D. Chacun des jeux est associé à une scène et contient une dizaine d'images correspondant à différents points de vue sur cette scène. La structure plane nécessaire à la validation correspond généralement à la façade d'un bâtiment ou à un agrandissement de photographie.



(a) Transformations géométriques



(b) Transformations colorimétriques

FIG. 6.1 Images résultant des transformations appliquées. Lors d'un changement géométrique 6.1(a) ainsi que lors d'un changement géométrique fixe couplé à un changement colorimétrique 6.1(b).

La présence de scènes plus ou moins colorées permet d'étudier l'apport de la couleur dans différentes circonstances. Les photographies ayant été acquises à des instants proches, chaque série d'images ne contient *a priori* que de faibles modifications de la couleur de l'illuminant (comme celles causées par le passage d'un nuage). L'appareil a néanmoins été laissé en mode automatique (temps d'exposition) pour provoquer des variations d'intensité lumineuse. Ces images constituent un cadre réaliste pour l'évaluation de l'appariement de descripteurs locaux couleur. La Figure 6.2 illustre ces prises de vue.

Base d'images réelles de Mikolajczyk Mikolajczyk *et coll.* [106], proposent plusieurs jeux d'images. Chaque jeu contient une image de référence ainsi que d'autres images prises sous différentes conditions d'illumination. L'homographie permettant de passer de l'image de référence à une autre image du même jeu est fournie, ce qui permet de créer une vérité terrain. Cette base contient huit jeux d'images et 6 images par jeu. Néanmoins, plusieurs jeux présentent peu d'intérêt pour l'application visée. En particulier, deux jeux seulement, contenant chacun 6 images, possèdent des changements de points de vue affine (les autres sont des rotations ou des zooms). Ces constatations nous ont amené à créer la base décrite précédemment.

Protocole Le protocole adopté est identique pour ces deux bases. Pour chaque jeu, une des images est prise comme référence et successivement comparée aux autres images de la série.

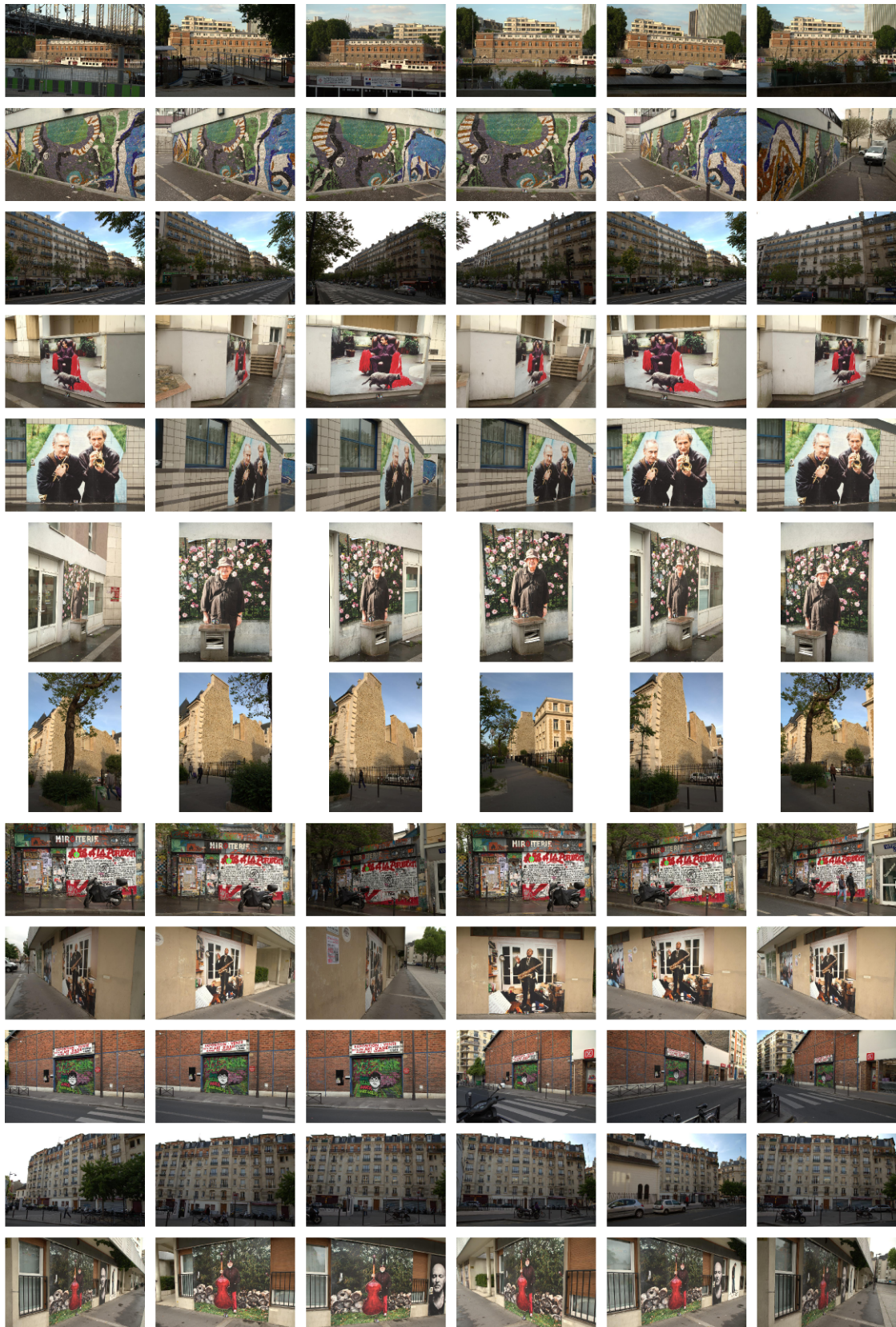


FIG. 6.2 Images utilisés pour les expériences réelles.

Puis, une seconde image est ensuite prise comme référence, comparée aux autres et ainsi de suite jusqu'à épuisement des combinaisons. Pour chaque paire d'images, la transformation est estimée à l'aide de l'algorithme ORSA [108]. Elle est ensuite manuellement vérifiée. Pour la base de Mikolajczyk, quand la transformation entre deux images n'est pas fournie, elle est estimée de la même manière. Pour l'obtention de la vérité terrain, seul les matchs correspondant à la transformation estimée sont considérés valides.

6.1.3 Validation d'une correspondance

Dans les protocoles précédents, la transformation séparant les deux points de vue est connue. Cette transformation permet de définir la projection exacte du point clé requête dans l'image candidate. Par la suite, nous choisissons de valider une mise en correspondance lorsque l'aire du descripteur requête, une fois projeté dans l'image candidate, et l'aire du descripteur du point candidat se recouvrent à plus de 50%. Cette procédure correspond à celles utilisées dans la littérature [105, 125]. Ce critère permet d'établir de manière robuste la validité d'une correspondance retournée par l'algorithme.

6.2 Mesures des performances

Une fois les correspondances sélectionnées et la vérité terrain établie, il nous reste à définir une mesure des performances. Le système que nous souhaitons étudier ici est composé d'un descripteur et d'un critère. C'est un classifieur binaire de mises en correspondance. Les correspondances retournées par l'algorithme sont ordonnées suivant une mesure de qualité. Un seuil ϵ appliqué à cette mesure de qualité permet de séparer les appariements valides des invalides. Dans le premier cas, la correspondance est dite positive et, dans le second cas, elle est dite négative. Au regard de la vérité terrain, ce classement peut alors être vrai ou faux. Quatre statuts peuvent donc être associés à la mise en correspondance, ils sont donnés par la table 6.1.

Vérité terrain / Test	Positive	Négative
Vraie Correspondance	VP	VN
Fausse correspondance	FP	FN

TABLE 6.1 Statuts possibles pour une correspondance.

Pour évaluer les performances de ce type de système, il est usuel de faire usage de courbes ROC (Receiver Operating Characteristic). Ce sont des courbes représentant l'évolution des vrais positifs ainsi que celle des faux positifs, en fonction du seuil de mise en correspondance. Nous utilisons ici le taux de succès comme mesure du nombre des vrais positifs et le taux de fausses alarmes pour mesurer les faux positifs. Pour un seuil ϵ défini sur l'intervalle $[0, 1]$, nous définissons ces taux de la manière suivante

$$\text{Taux de succès}(\epsilon) = \frac{\text{vp}(\epsilon)}{\text{vp}(1) + \text{fp}(1)}, \quad \text{Taux de fausses alarmes}(\epsilon) = \frac{\text{fp}(\epsilon)}{\text{vp}(1) + \text{fp}(1)}.$$

Précisons que $\epsilon = 0$ implique qu'aucune correspondance ne soit sélectionnée et $\epsilon = 1$ implique qu'elles le soient toutes. Dans les expressions précédentes, le terme $\text{vp}(1) + \text{fp}(1)$ correspond au nombre de requêtes effectuées. Un classifieur idéal est un classifieur pour lequel il existe

une valeur ϵ_0 telle $vp(\epsilon_0) = vp(1)$ et que $fp(\epsilon_0) = 0$. Un classifieur aléatoire est lui défini par un rapport $fp(\epsilon)/vp(\epsilon)$ constant.

Cette mesure est intéressante car elle permet de discriminer à la fois des critères et des descripteurs différents. Deux exemples de courbes types sont donnés par la Figure 6.3.

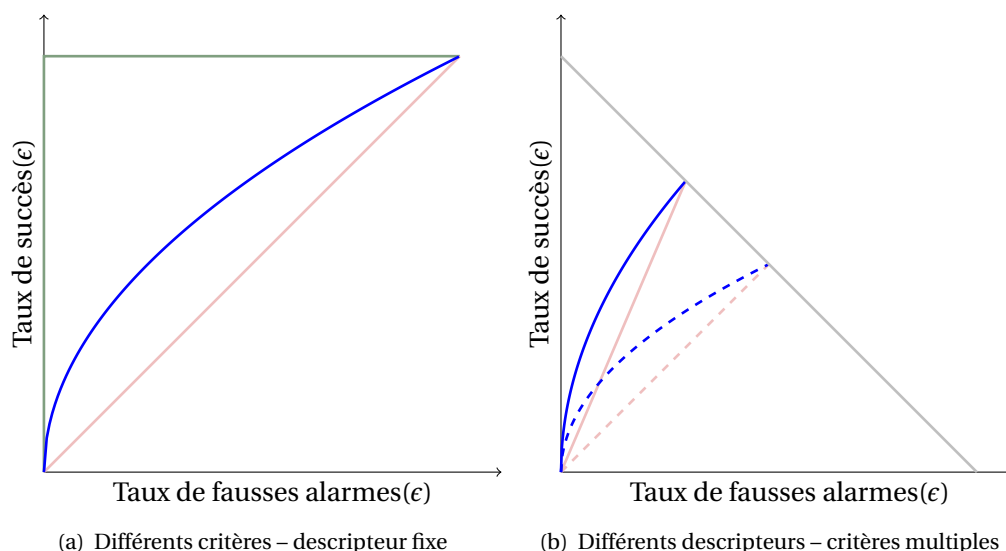


FIG. 6.3 Cette figure illustre les différents comportements possibles pour des critères d'appariements et des descripteurs. Sur la Figure 6.3(a), la courbe verte représente le résultat issu d'une classification parfaite, la courbe rouge donne le résultat d'une classification aléatoire tandis que la courbe bleue illustre un résultat intermédiaire. Pour la Figure 6.3(b), les courbes en traits pleins et pointillés montrent le résultat donné par deux descripteurs différents. La déplacement le long de la courbe grise est la conséquence de l'utilisation des mêmes points clés pour toutes les expériences. Les courbes bleue et rouge illustrent le comportement de ces descripteurs pour deux types de critères. Le point de rencontre entre la ligne grise et une des courbes colorées permet d'apprécier la répartition finale entre le nombre de correspondances valides et invalides.

6.3 Terminologie

Dans les expériences présentées par la suite, nous allons utiliser plusieurs descripteurs nommés de la manière suivante :

- SIFT correspond au descripteur décrit à la Section 4.3.2, c'est à dire à la variante du SIFT classique présentée dans [125],
- Opponent 1, Opponent 2, Ohta 1 et Ohta 2 chacun d'entre eux correspondant à l'application de la méthode SIFT décrite sur un des deux canaux chromatiques, soit de l'espace Opponent (voir la Section 4.3.3) soit de l'espace Ohta (cf. Section 4.3.4),
- HueMono correspondant au descripteur proposé par Van de Weijer [148] qui est basé sur un unique histogramme de teinte (cf. Section 5.2),
- Hue et HueNorm correspondent aux descripteurs de teinte composés de plusieurs secteurs. Pour le descripteur HueNorm, la couleur est normalisée localement, ce qui n'est pas le cas pour le descripteur Hue (cf. Section 5.1).

En plus des descripteurs utilisés séparément, nous considérons leur utilisation conjointe. Les combinaisons basées sur l'application de la méthode SIFT sur les canaux de différents espaces de couleurs sont précisées dans la Table 6.2. Les combinaisons faisant appel au descripteur de teinte (Hue ou HueNorm) sont notées avec le symbole +.

Nom	Combinaison
Opponent	Opponent 1 + Opponent 2
Ohta	Ohta 1 + Ohta 2
Opponent-SIFT	SIFT + Opponent
Ohta-SIFT	SIFT + Ohta
RGB-SIFT	SIFT sur chacun des canaux RGB
HSL-SIFT	SIFT sur chacun des canaux HSL

TAB. 6.2 Combinaisons des différents descripteurs utilisées dans les expériences.

6.4 Validation des descripteurs couleur

Dans le cadre de la mise en correspondance de descripteurs locaux, nous cherchons à étudier le comportement des descripteurs présentés dans les Sections 5.1 et 5.2 afin de déterminer le meilleur moyen de coder l'information couleur localement ainsi que l'apport de la couleur à la méthode SIFT. Plus précisément, nous souhaitons quantifier leur capacité à décrire de manière robuste et discriminante le contexte d'un point clé et cela, dans le cas de changements de point de vue et/ou de modifications des conditions d'illumination.

6.4.1 Robustesse aux changements géométriques

Dans cette partie, les performances des différents descripteurs sont comparées. Les transformations sont illustrées dans le cas de transformations synthétiques par la Figure 6.1(a) et pour les transformations réelles par la Figure 6.2.

Espaces antagonistes

Nous comparons ici les différents espaces antagonistes existants dans la littérature et recherchons le plus adapté au calcul de descripteurs SIFT. Les espaces antagonistes (cf. Section 4.1.1) séparent l'information couleur en trois canaux, un canal de luminance et deux canaux chromatiques encodant des contrastes couleur. Ces espaces se différencient entre eux à la fois par les couleurs utilisées pour évaluer les contrastes mais aussi par une éventuelle normalisation de ces contrastes par rapport à l'intensité lumineuse. Le canal de luminance est identique pour ces trois espaces, qui ne se différencient donc que par leur façon d'encoder la couleur. Dans cette expérience, seuls les deux canaux chromatiques de chaque espace antagoniste sont utilisés et donc, aucun descripteur n'est extrait du canal de luminance. Les trois espaces considérés sont :

Espace Opponent Les canaux chromatiques de l'espace antagoniste « Opponent » mesurent des contrastes jaune/bleu ainsi que des contrastes rouge/vert. Les descripteurs extraits à partir de ces canaux sont notés Opponent.

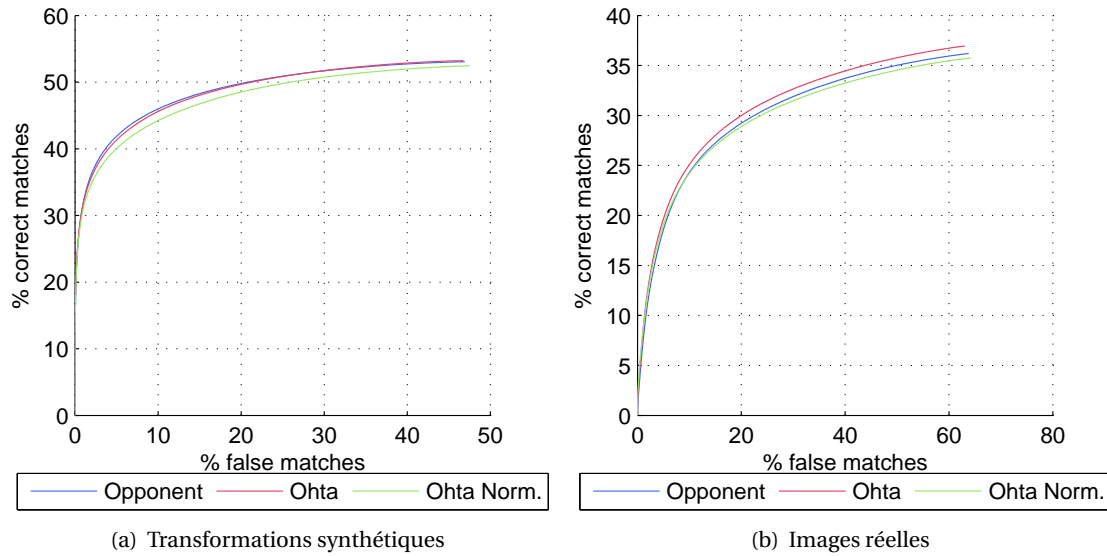


FIG. 6.4 Résultats pour différents descripteurs antagonistes. La méthode SIFT est ici appliquée uniquement sur les canaux chromatiques des espaces antagonistes, Opponent, Ohta et Ohta normalisé. Le critère utilisé pour ordonner les appariements est ici NN-DT.

Espace Ohta Les descripteurs issus des canaux chromatiques de l'espace Ohta, qui encodent des contrastes rouge/bleu et vert/magenta sont notés Ohta.

Espace Ohta normalisé Les descripteurs issus des canaux chromatiques de l'espace Ohta normalisés (cf. Section 4.1.1) par l'intensité sont, quant à eux, notés OhtaNorm. Le suffixe -Norm fait ici référence à la normalisation par le canal d'intensité et non à la normalisation des couleurs du support des descripteurs décrite au paragraphe 5.1.

Cette expérience a pour but de statuer sur l'utilité de la normalisation des canaux chromatiques par le canal d'intensité ainsi que de déterminer l'influence du choix des contrastes couleurs et ce, dans le contexte précis de l'appariement de descripteurs locaux. Il ressort de la littérature [144], que l'apport de la normalisation dépend de l'application. Ils observent par exemple que les descripteurs *OpponentNorm* donnent de meilleurs résultats que les descripteurs *Opponent* dans une expérience de classification d'objet et des résultats opposés dans une expérience de classification à partir de vidéos. Concernant le choix des contrastes couleurs, seul l'espace antagoniste Opponent a été évalué dans cet article.

La Figure 6.4 présente les résultats pour le protocole $A \rightarrow A'$ ainsi que pour le protocole utilisant des transformations non synthétiques. Les correspondances sont triées selon le critère NN-DT. Dans le cas du protocole $A \rightarrow A'$, la courbe est obtenue en agrégeant les résultats des 10 transformations géométriques présentées précédemment pour chacune des 100 images utilisées.

Le choix de l'espace antagoniste n'est pas évident à partir de ces expériences car les courbes de performances associées aux descripteurs Ohta et Opponent sont proches notamment pour les transformations synthétiques. Le descripteur Ohta est légèrement plus performant dans l'expérience basée sur des transformations réelles. Le principal avantage lié à son utilisation

réside dans sa plus grande robustesse face à certaines transformations colorimétriques. Cet avantage est étudié plus en détail au paragraphe 6.4.2. On peut observer que dans les deux expériences proposées, la normalisation des canaux chromatiques (OhtaNorm) dégrade les performances. Il est bien entendu possible de normaliser également les canaux de l'espace Opponent, les résultats sont alors identiques à ceux fournis par les descripteurs OhtaNorm mais ne sont pas présentés ici.

Comparaison des descripteurs Ohta et Hue en présence d'une transformation géométrique

Maintenant qu'un espace antagoniste Ohta a été choisi comme référence, nous proposons ici d'étudier la résistance de différents descripteurs aux transformations géométriques. Nous utilisons pour cela les protocoles synthétiques décrits plus haut et illustrés par la Figure 6.1(a) ainsi que le protocole utilisant des transformations réelles.

Détails expérimentaux Les descripteurs étudiés ici sont au nombre de 6.

- Les descripteurs Hue, HueMono et HueNorm sont basés sur l'information en provenance du canal de teinte. Le suffixe -Norm fait ici référence à la normalisation locale des couleurs (cf. Section 5.1) et non plus à une normalisation vis-à-vis du canal d'intensité comme le descripteur OhtaNorm précédemment.
- Les descripteurs SIFT, Ohta 1 et Ohta 2 font appel à l'orientation du gradient ainsi qu'à la méthode SIFT. Ces descripteurs sont extraits de chacun des trois canaux de l'espace Ohta. Pour faciliter la lecture, le descripteur issu du canal Oh3 est noté SIFT et correspond au descripteur présenté à la section 4.3.2.

Les résultats pour les deux protocoles correspondant aux transformations synthétiques sont donnés par la Figure 6.5. Pour les expériences réelles, les résultats sont donnés sur la Figure 6.6. Pour les protocoles synthétiques, les courbes sont obtenues en agrégeant les résultats correspondant à une valeur du paramètre de tilt fixée. Pour le protocole réel, les courbes sont obtenues par fusion de tous les résultats. C'est-à-dire que pour une scène donnée, toutes les correspondances sont concaténées et ordonnées suivant le seuil de mise en correspondance. Cette liste est utilisée pour calculer les courbes ROC. Sauf lorsque cela est précisé, les correspondances sont ordonnées suivant le critère NN-DT.

Critique des protocoles synthétiques Avant tout, notons que les protocoles synthétiques $A \rightarrow A'$ et $A \rightarrow A', B$ ne sont pas adaptés à la comparaison directe de descripteurs tels que les descripteurs basés sur le canal de teinte et les descripteurs basés sur l'orientation du gradient. En effet, les transformations colorimétriques imposées aux images ne représentent pas la réalité dans son ensemble. Toutefois, ce protocole permet d'évaluer le pouvoir discriminant de chaque descripteur ainsi que sa robustesse à un changement de point de vue simulé. De plus, il permet une comparaison relative des performances des descripteurs en fonction de l'intensité de la transformation géométrique appliquée. Pour obtenir une comparaison directe entre descripteurs de natures différentes, nous proposons également une expérience utilisant des images réelles.

Influence du tilt Dans ces courbes, l'influence du tilt permet de confirmer que la teinte est plus robuste que l'orientation du gradient à une transformation géométrique. Grossièrement, lorsque le tilt passe de 2 à 3.3, le pourcentage de bons appariements retrouvés à l'aide du descripteur SIFT est divisé par 6, pour les descripteurs Hue, HueNorm et HueMono il est divisé

par 3.5, 4.5 et respectivement 2.5. Ces observations indiquent que la représentation par histogrammes de teinte est plus robuste que les représentations basées sur l'orientation du gradient. On observe aussi qu'un découpage du descripteur de teinte permet d'améliorer de manière très sensible le pouvoir discriminant du descripteur, mais rend ce dernier plus vulnérable aux fortes transformations géométriques.

Cas du protocole $A \rightarrow A', B$ Les courbes obtenues avec ce protocole viennent confirmer les résultats précédents. Logiquement, le pourcentage de bonnes correspondances est divisé par deux puisque le nombre de bons appariements possibles est lui-même divisé par deux en moyenne. On observe alors que les descripteurs de teinte se montrent particulièrement efficaces pour séparer les fausses mises en correspondances (venant de l'image B) des bonnes correspondances (venant de l'image A). Ceci illustre l'intérêt de l'information couleur en général et, en particulier, du descripteur de teinte pour filtrer de fausses correspondances.

Normalisation locale de la teinte Rappelons que pour obtenir le descripteur HueNorm, les couleurs sont normalisées localement, chaque valeur R , G et B est divisée par la moyenne du canal correspondant sur le support du descripteur. On pourrait s'attendre à ce que la normalisation locale entraîne une baisse importante des performances. Effectivement, de nombreux supports sont peu colorés et/ou monochromes, et il paraîtrait normal que la normalisation locale des couleurs fasse disparaître cette information. Et en effet, dans les expériences synthétiques, on observe que la normalisation fait baisser le pourcentage de bons appariements retrouvés en utilisant le descripteur de teinte. Ainsi, le nombre de points clés correctement appariés en utilisant le descripteur HueNorm est inférieur à celui permis par l'utilisation du descripteur Hue. Néanmoins, les distances mesurées à partir de descripteurs dont la couleur a été normalisée localement permettent une meilleure hiérarchisation des appariements entre eux. En effet, la normalisation locale permet d'enlever la teinte dominante, celle que l'on suppose due à l'illuminant. Ceci s'effectue au prix d'une perte d'information pour les descripteurs monochromes et explique en partie pourquoi ce descripteur obtient moins de bons matchs que sa variante non normalisée. On peut par exemple imaginer un descripteur rouge qui se retrouvera « gris » après normalisation. Plus rien ne permet alors de le distinguer d'un descripteur extrait d'un autre support monochrome. De plus, les deux supports étant alors peu saturés, les descripteurs extraits sont fortement bruités et la distance les séparant est alors très variable. La normalisation dégrade donc la robustesse aux transformations géométriques puisque, plus cette dernière est importante, plus le support servant à coder le descripteur va varier et avec lui la valeur servant à le normaliser. Cependant, il est tout à fait remarquable que la normalisation ne détériore que très peu l'information couleur présente sur le support du descripteur. Nous allons voir que le comportement des descripteurs face à la normalisation est différent pour le protocole réel.

Cas des transformations réelles La Figure 6.6 illustre les performances de ces descripteurs lors de transformations réelles. On observe ici que le comportement relatif des descripteurs est très différent : les descripteurs basés sur l'orientation du gradient donnent de meilleurs résultats que ceux reposant sur l'information de teinte. Du côté des descripteurs de teinte, l'intérêt de la sectorisation est confirmé. Une différence notable est que, dans ces courbes, la normalisation locale des couleurs du descripteur de teinte permet d'obtenir une proportion plus importante de bonnes correspondances. Les raisons de ce comportement ne sont pas évidentes. Sur cette figure, on peut observer que le critère NN-DR permet de rejeter les

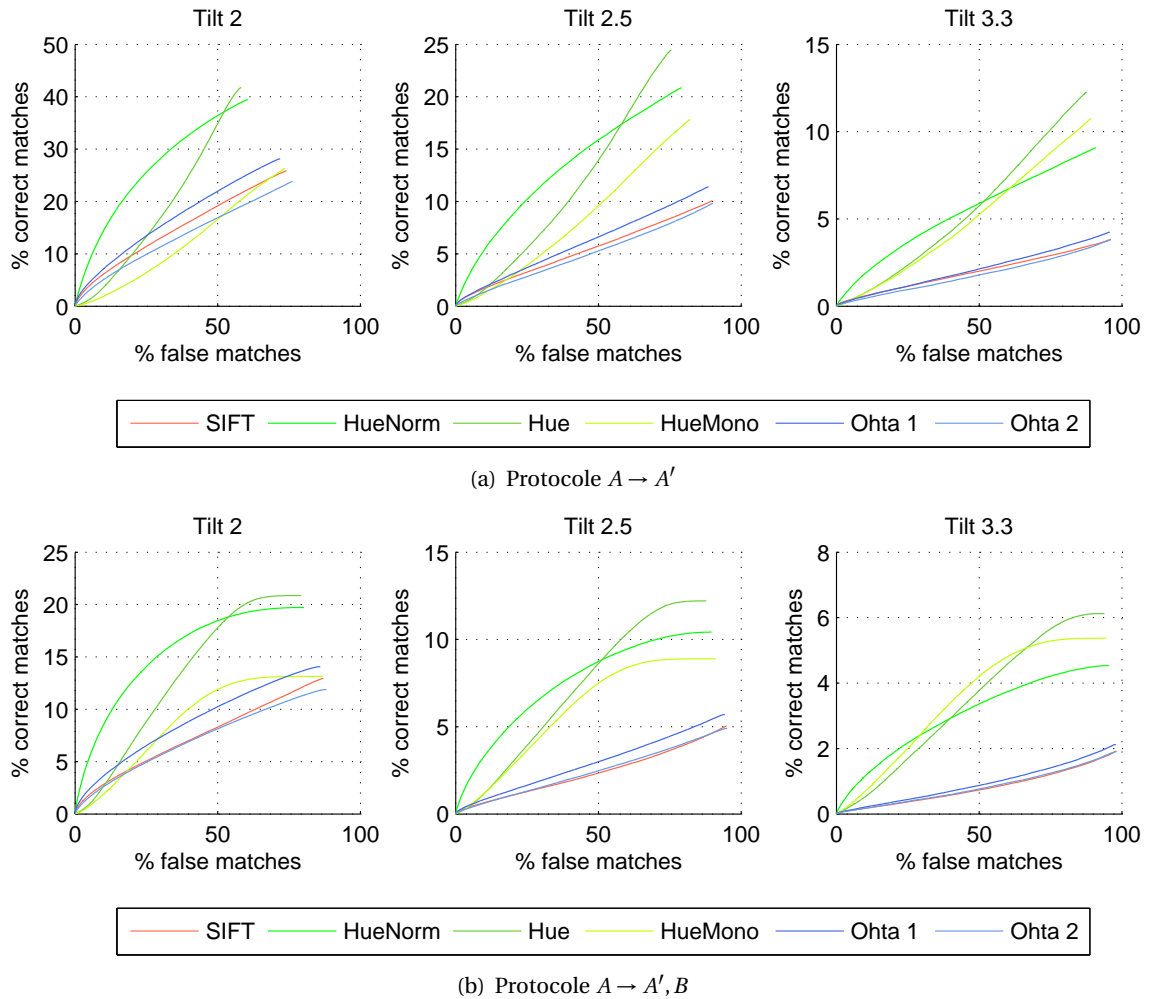


FIG. 6.5 Performances des descripteurs en présence d'un changement géométrique et pour différents protocoles avec le critère NN-DT. On peut observer que les descripteurs utilisant le canal de teinte sont moins affectés par un par le changement de tilt. Dans le cas du protocole $A \rightarrow A', B$, on peut remarquer que ces mêmes descripteurs permettent un rejet efficaces des fausses correspondances induites par le protocole.

descripteurs répétitifs. Ce rejet est particulièrement marqué pour les descripteurs Hue car il est fréquent que plusieurs descripteurs aient le même contenu monochromatique. La normalisation locale des couleurs produit un effet similaire, car elle limite très fortement le nombre de descripteurs monochromes et les rend ainsi plus discriminants. Il faut noter l'excellent comportement du descripteur de teinte, il est en effet à peine moins discriminant que ceux basés sur l'orientation du gradient.

6.4.2 Robustesse au changement d'illuminant

Nous étudions ici la robustesse des descripteurs aux variations colorimétriques. Pour cela, nous utilisons le protocole $A \rightarrow A'$, une transformation correspondant à un tilt de 2.5

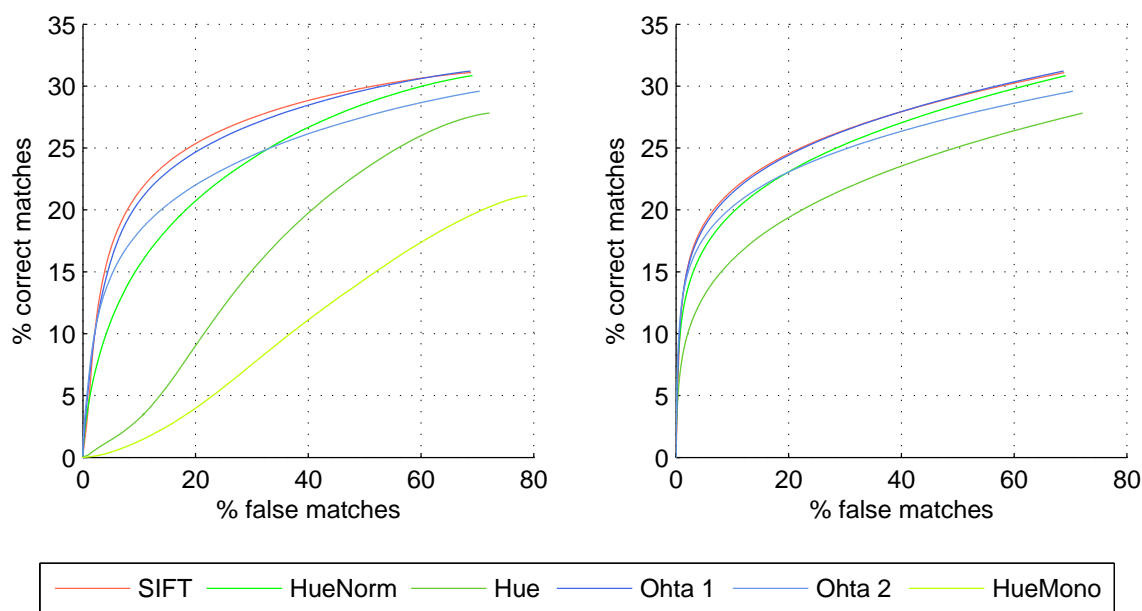


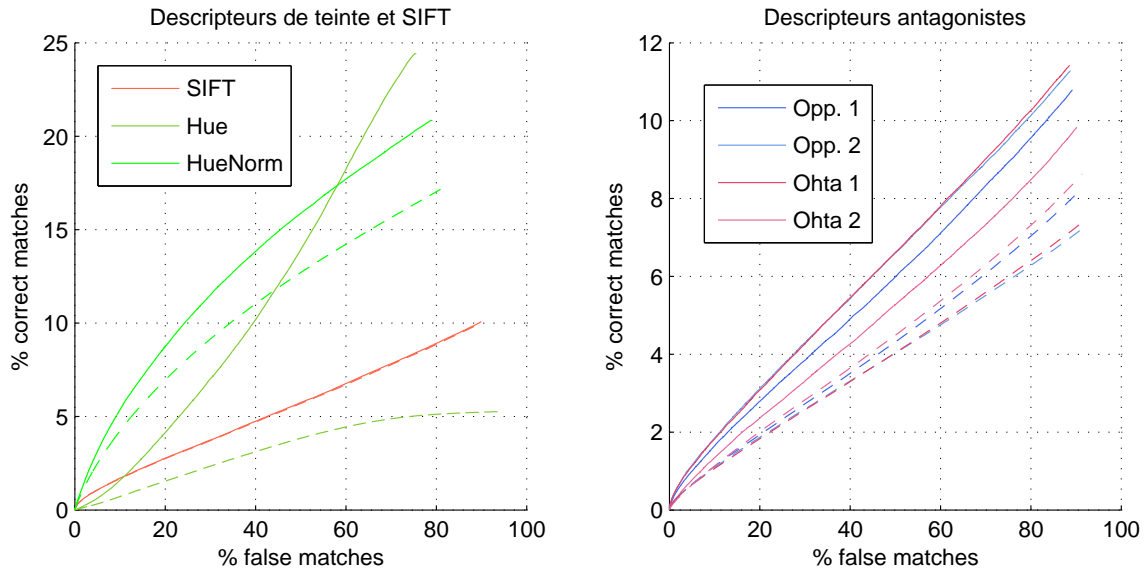
FIG. 6.6 Résultats sur la base d'images réelles. À gauche avec le critère NN-DT et à droite avec le critère NN-DR.

ainsi qu'un changement de température de couleur. Ces modifications sont expliquées au paragraphe 6.1.1 et illustrées par la Figure 6.1(b). Ces expériences ont deux objectifs :

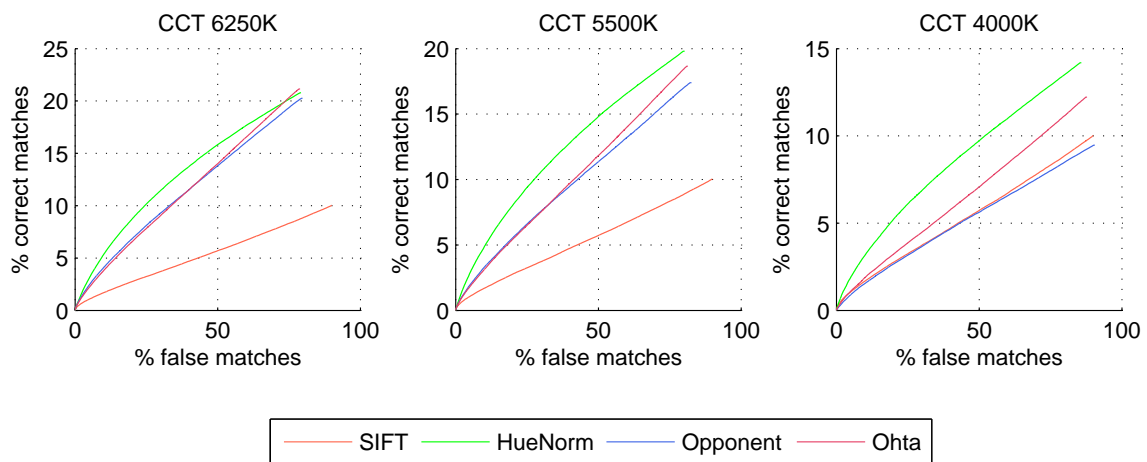
- mesurer l'efficacité de la normalisation locale de la couleur sur la robustesse du descripteur de teinte (Hue et HueNorm) vis-à-vis de ces changements,
- étudier la robustesse des descripteurs présentés précédemment à un changement de couleur d'illuminant.

La Figure 6.7(a) donne les résultats en considérant l'ensemble des changements de température de couleur tandis que la Figure 6.7(b) montre comment se comportent les descripteurs couleur face à une variation progressive de la température de couleur. Les courbes correspondant au descripteur SIFT sont présentes pour servir de référence puisque le descripteur SIFT est invariant aux modifications de la couleur de l'illuminant imposées ici. Il est en effet extrait à partir du canal de luminance qui n'est pas modifié lors d'un changement de la température de couleur.

Normalisation locale de la couleur et descripteur de teinte Dans le cas du descripteur Hue, la différence de résultat induite par une modification colorimétrique illustre à quel point le canal de teinte non normalisé est sensible à un changement de la couleur de l'illuminant. En effet, ses performances s'écroulent complètement, car un changement de température de couleur provoque une modification du canal de teinte difficilement prévisible. Cependant, la normalisation locale de la couleur (descripteur HueNorm) permet de fortement limiter cette perte de performances et même, étonnamment, de les améliorer comme le montrent les expériences réelles. On peut observer sur la Figure 6.7(b) que la teinte est plus robuste que le descripteur Ohta, ses performances déclinant moins pour une même transformation colorimétrique.



(a) Effets de la normalisation (gauche) et influence sur les différents canaux antagonistes (droite). Les courbes en traits pleins sont obtenues sans changement d'illuminant tandis que les courbes en pointillés sont obtenues avec changements de couleur de la source lumineuse. Ces courbes sont issues de la fusion des appariements provenant de l'ensemble des transformations colorimétriques.



(b) Comportement de différentes combinaisons lors d'un changement de couleur de l'illuminant.

FIG. 6.7 Influence d'un changement colorimétrique moyen (Figure 6.7(a)) puis d'un changement colorimétrique croissant (Figure 6.7(b)) pour le protocole $A \rightarrow A'$. Le paramètre de tilt est fixé à 2.5. Le critère utilisé est NN-DT.

Cas des descripteurs antagonistes Le choix de l'espace Ohta par rapport à l'espace Opponent a été justifié par le fait qu'il permettait d'atteindre des performances légèrement meilleures lors d'une transformation géométrique. Le principal avantage des descripteurs issus de cet espace est qu'ils sont également plus robustes à un changement de température de couleur. En effet, les courbes sur la droite de la Figure 6.7(a) montrent qu'en présence de transformations colorimétriques, les descripteurs de l'espace Ohta (Ohta 1 et Ohta 1) sont plus robustes que les descripteurs Opponent (Opponent 1 et Opponent 1). Ceci est particulièrement vrai pour le canal Ohta 2 et peut s'expliquer par le fait que cet axe de couleur

est perpendiculaire au lieu de Planck : même lors d'un faible changement de température de couleur, les descripteurs Ohta se montrent plus performants que les descripteurs Opponent (voir Figure 6.7(b)). Il est bien évidemment possible d'augmenter la robustesse aux variations colorimétriques des descripteurs Ohta et Opponent par une normalisation locale comme c'est le cas pour le descripteur de teinte, mais dans une moindre mesure, ces derniers étant déjà robustes.

6.4.3 Combinaison de descripteurs

Après avoir étudié les caractéristiques individuelles des différents descripteurs décrits, nous proposons maintenant de combiner ces descripteurs. Une première combinaison considérée est Ohta-SIFT. Elle correspond à l'extraction de trois descripteurs SIFT extraits de l'espace antagoniste Ohta. Nous proposons également de combiner le descripteur HueNorm au descripteur SIFT, ce qui donne la combinaison appelée HueNorm+SIFT. Enfin, nous proposons de combiner les descripteurs Ohta-SIFT et HueNorm, combinaison que nous nommons HueNorm+Ohta-SIFT. Nous comparons ces combinaisons à deux méthodes proposées dans la littérature :

- RGB-SIFT (méthode proposée et testée dans [144]),
- HSL-SIFT (méthode proche de celle proposée par Bosch *et coll.* [13]).

Ces deux méthodes consistent à calculer des descripteurs SIFT sur les canaux de l'espace RGB ou sur ceux de l'espace HSL. Les résultats sont donnés par la Figure 6.8. La Figure 6.8(a) présente les résultats avec le protocole $A \rightarrow A'$ tandis que la Figure 6.8(b) présente les résultats sur les images réelles.

Histogrammes de teinte et descripteurs antagonistes Les combinaisons proposées fournissent dans ces deux expériences les meilleurs résultats (HueNorm+Ohta-SIFT, HueNorm+SIFT et Ohta-SIFT). Le descripteur de teinte (HueNorm) utilisé en combinaison avec le descripteur SIFT donne dans les deux expériences des résultats analogues à ceux de la combinaison Ohta-SIFT. Ceci démontre l'intérêt de cette approche car le descripteur HueNorm est plus compact que le descripteur Ohta-SIFT.

Les descripteurs antagonistes codent des contrastes couleur par l'intermédiaire de dérivées alors que le descripteur de teinte décrit la distribution des couleurs à l'aide de valeurs brutes. Signalons que dans le cas de niveaux de gris, la distribution des valeurs brutes (histogrammes de niveaux de gris) est à la fois peu discriminante et peu robuste et ne permet sans doute pas d'enrichir les SIFT. Inversement, les expériences précédentes montrent que pour coder l'information couleur, il est utile de considérer à la fois des transitions et des valeurs brutes. En combinant ces descripteurs, on obtient ainsi une description plus robuste et discriminante du contexte d'un point clé.

Comparaison avec d'autres méthodes de la littérature Le descripteur RGB-SIFT [144] souffre de la redondance d'information des canaux RGB. Les résultats correspondants sur la Figure 6.8 montrent qu'il n'est pas trivial d'augmenter les performances par une simple concaténation de descripteurs et qu'une représentation adaptée de l'information est ici très importante. Néanmoins, ce descripteur a l'avantage d'être invariant par changement d'illuminant et plus robuste dans le cas d'un changement dans les fonctions de sensibilité du capteur. Le descripteur HSL-SIFT similaire à celui proposé dans [13] est moins performant du fait de l'instabilité de la teinte dans les faibles saturations. Précisons que le calcul du gradient n'est ici pas

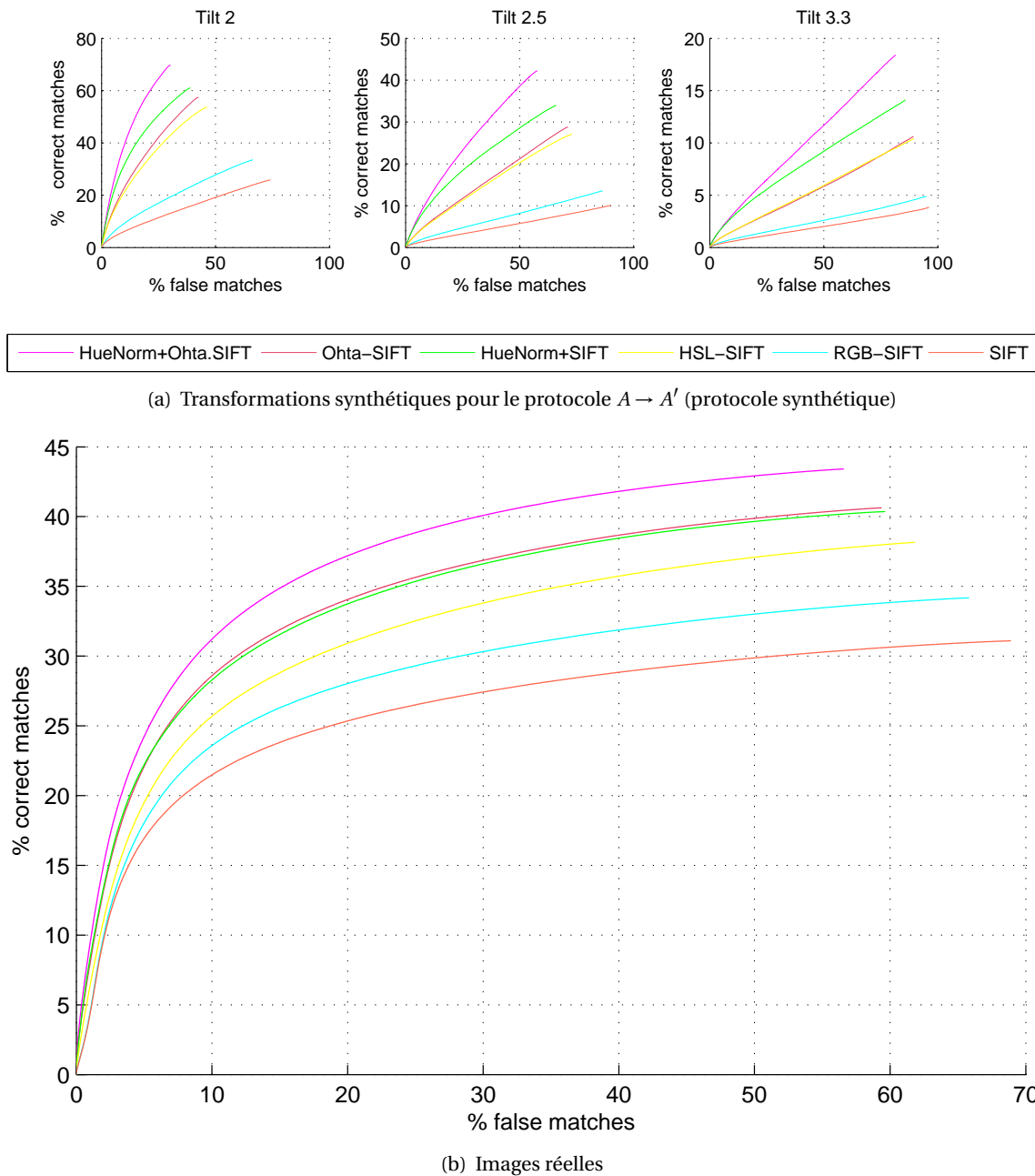


FIG. 6.8 Performances des descripteurs couleur combinés avec le descripteur SIFT en présence d'un changement géométrique et avec le critère NN-DT.

adapté à la nature circulaire de la teinte. Les courbes donnant les résultats pour le descripteur Opponent-Sift ne sont pas tracées ici mais sont quasiment identiques à celles du descripteur Ohta-Sift.

6.4.4 Autres résultats

Découpage du patch en secteur

Nous proposons dans ce chapitre une description locale et sectorisée de la teinte. Comme nous l'avons vu au chapitre 4, VanDeWeijer *et coll.* [148] ont proposé de coder l'information couleur par des histogrammes locaux de valeurs de teinte pondérés par la saturation. Contrairement au descripteur que nous proposons, l'information est codée sur un seul histogramme. L'explication avancée est que le descripteur SIFT contient déjà l'information géométrique, et qu'il n'est donc pas utile de l'encoder une seconde fois. Nous montrons ici que cela est inexact. Dans le cas de la teinte, un encodage sectorisé permet d'avoir une mesure relative

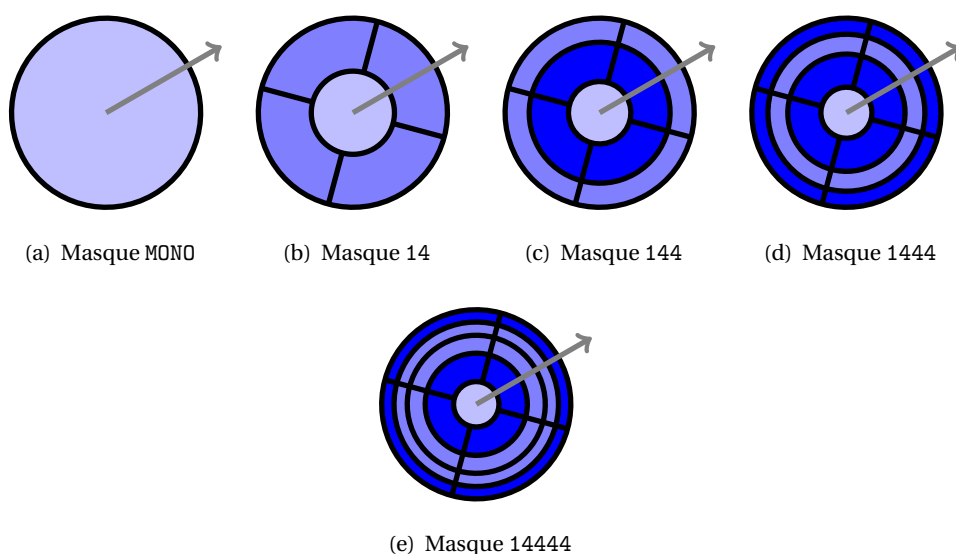


FIG. 6.9 Les différents découpages utilisés.

de la présence d'une teinte dans un secteur du descripteur. La distinction entre descripteurs monochromes est alors faite sur la répartition de la saturation entre les différents secteurs. La sectorisation permet ainsi de différencier des descripteurs ayant la même teinte globale en se basant sur la répartition spatiale de la saturation.

Dans cette expérience, nous testons différents découpages possibles pour chacun des quatre descripteurs proposés. Le facteur de proportionnalité de la taille du patch est conservé constant : l'information est codée sur un disque de rayon 12 fois l'échelle de détection du point clé. La taille des secteurs varie d'un découpage à l'autre. Pour un découpage donné, tous les secteurs ont la même aire. Les différents découpages sont représentés sur la Figure 6.9. Pour les quatre descripteurs, nous testons les performances fournies par les différents découpages. Les courbes résultantes sont visibles sur la Figure 6.10. On observe que pour le descripteur HueNorm, les performances croissent avec un découpage plus fin mais s'écroulent pour le découpage 14444. Pour les descripteurs basés sur l'orientation du gradient, cette dégradation n'est pas observée. Cependant, les performances des découpages 14444 et 1444 sont très proches. En pratique le découpage 144 constitue un bon compromis entre les performances et le temps de calcul. En effet, lors de la recherche d'un descripteur dans une base de données, le coût en calcul est linéairement proportionnel au nombre d'histogrammes utilisés.

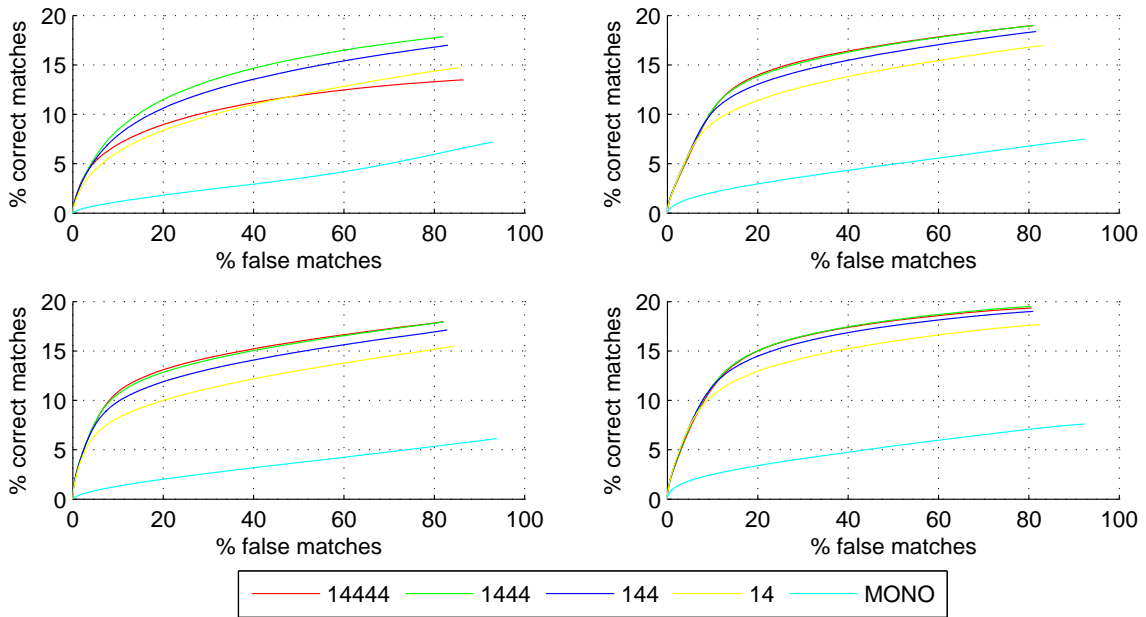


FIG. 6.10 Influence du découpage en secteurs des différents descripteurs. De gauche à droite et de haut en bas les descripteurs de teinte (HueNorm), Ohta (Ohta 1 et Ohta 2) ainsi que le descripteur SIFT SIFT.

Importance des secteurs

Dans le descripteur SIFT proposé par Lowe [94], la contribution de la norme du gradient en un pixel est pondérée par la distance au centre du descripteur. Cela permet d'accorder un poids inférieur aux pixels (et par conséquent aux secteurs) extérieurs du descripteur. Ici, cette pondération n'est pas utilisée. Nous proposons à la place une expérience permettant d'appliquer une pondération sur les distances correspondant aux différents secteurs comme cela est indiqué au paragraphe 5.3.3. De manière à fixer les poids de cette pondération, nous observons le résultat de l'appariement en n'utilisant qu'un seul secteur pour chaque descripteur. Ces résultats sont donnés par la Figure 6.11. Les résultats sont similaires pour les quatre descripteurs. En considérant les performances relatives de chacun des secteurs, on considère les facteurs suivants : 0.16 pour le secteur central, 0.12 pour les secteurs de la première couronne et 0.09 pour les secteurs de la couronne extérieure.

6.4.5 Conclusions

Nous avons montré ici l'utilité des contributions présentées aux Sections 5.1 et 5.2. L'intérêt d'une représentation localisée de la teinte, robuste à de fortes transformations géométriques a été montré. De même, l'utilité de la normalisation des couleurs du support pour l'invariance à un changement de couleur de l'illuminant a été confirmée. Cette normalisation a de plus montré son utilité pour augmenter le pouvoir discriminant du descripteur de teinte. Les descripteurs basés sur les canaux chromatiques de l'espace Ohta ont également été montrés plus robustes aux transformations chromatiques induites par un changement de température de couleur de l'illuminant que les descripteurs utilisant l'espace antagoniste Opponent. Et ce, dans le cas d'expériences réelles ainsi que lors de changements d'illuminants simulés. De plus,

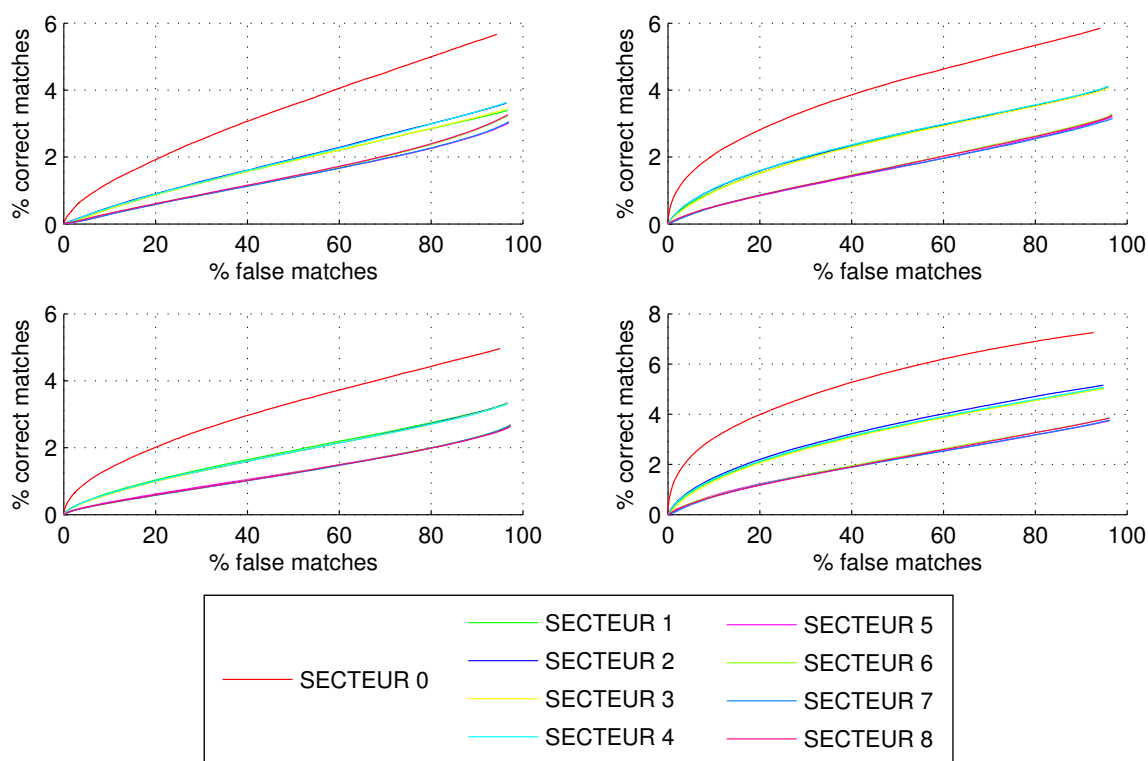


FIG. 6.11 Importance de chaque secteur pour la mise en correspondance. De gauche à droite et de haut en bas les descripteurs de teinte (HueNorm), Ohta (Ohta 1 et Ohta 2) et le descripteur SIFT SIFT. Le découpage en secteurs correspond ici au découpage 144 de la Figure 6.9.

il a été montré que l'utilisation conjointe des descripteurs Ohta-SIFT et HueNorm permet une forte augmentation des performances, tout en créant un descripteur robuste. Cette description repose sur des descripteurs codant des informations de natures différentes : des contrastes couleurs pour le descripteur Ohta-SIFT, et une description de la distribution des couleurs avec le descripteur HueNorm.

6.5 Validation du critère *a contrario*

Nous cherchons ici à évaluer l'apport du critère *a contrario* dans la mise en correspondance couleur. Pour cela, nous utilisons les deux protocoles synthétiques ainsi que le protocole basé sur des transformations réalistes décrit dans la Section 6.1.2. Premièrement, nous observons comment se comporte le critère proposé par rapport aux autres critères de la littérature à savoir les critères NN-DR (ratio entre les distances des deux plus proches voisins) et NN-DT (restriction au plus proche voisin). Puis, nous étudions l'influence des combinaisons de distances sur l'estimation du nombre de fausses alarmes. Nous n'étudions pas ici la possibilité de détecter des mises en correspondance multiples par le critère *a contrario*. Néanmoins, il s'agit d'une possibilité offerte par cette approche [125].

6.5.1 Résultats de classification

Dans le cas du critère d'appariement NN-DR, une bonne correspondance est associée à un descripteur candidat qui possède une distance au descripteur requête petite devant celles associées aux autres descripteurs candidats. Dans le cas de NN-AC, une bonne correspondance est une correspondance pour laquelle la distance entre descripteurs est trop petite pour être due au hasard. Spécifiquement, « au hasard » signifie selon une hypothèse \mathcal{H}_0 . La mesure du rejet de cette hypothèse définit alors *a contrario* une bonne correspondance. Cette différence d'approche permet de mieux comprendre les courbes de la Figure 6.12(a). Une première

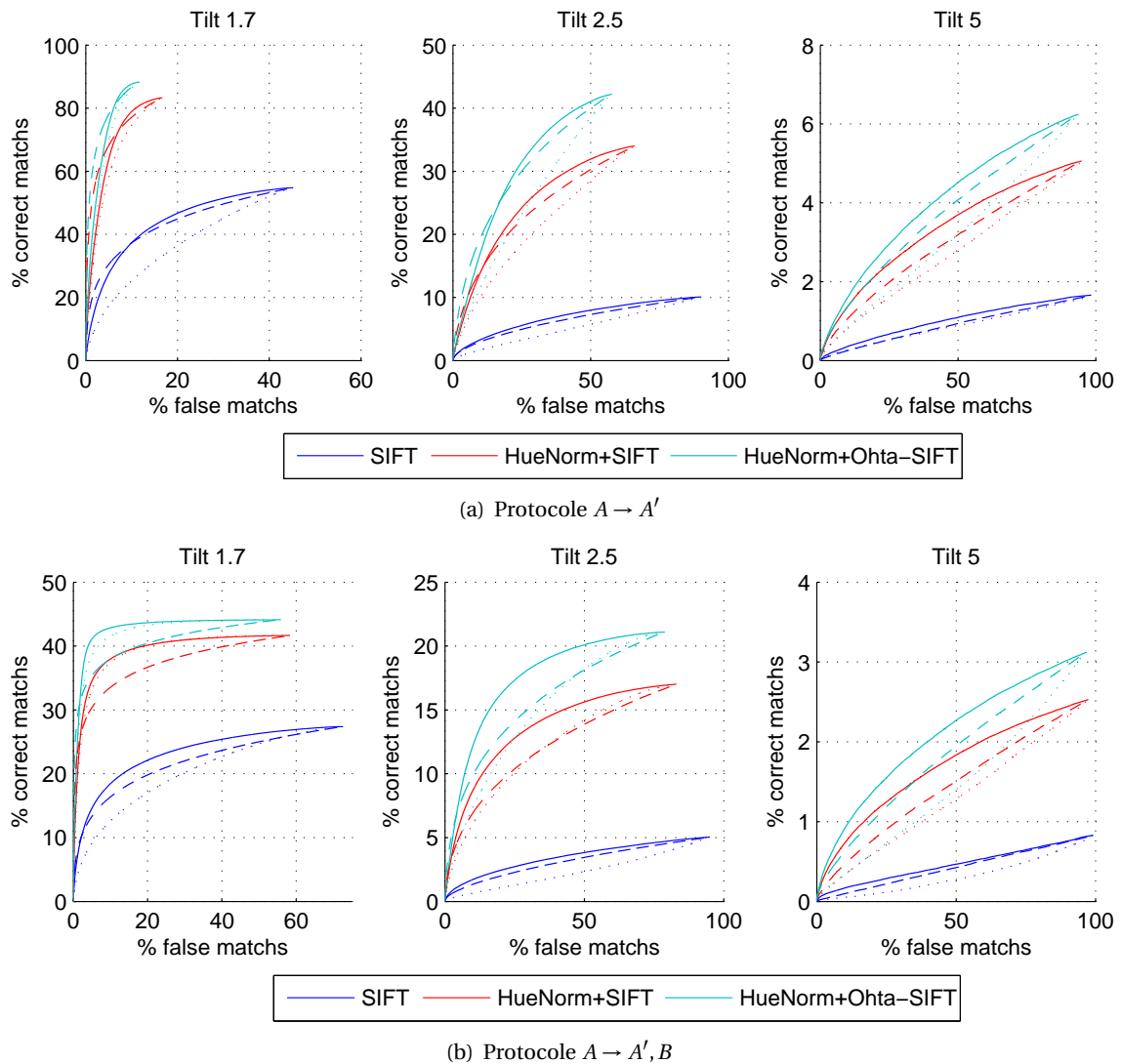


FIG. 6.12 Comportement des différents critères pour différentes valeurs du paramètre de tilt. La courbe correspondant au critère NN-AC est en traits pleins, celle associée au critère NN-DR est hachurée. La courbe du critère NN-DT est en pointillés.

observation est que le critère NN-DT (en pointillé) se montre toujours moins stable que les critères NN-DR (hachures) et NN-AC (trait plein). Cela est dû au fait que le choix du seuil NN-DT ne prend pas en compte les caractéristiques du descripteur requête. L'essentiel de

l'information à tirer de ces courbes vient donc de la comparaison entre le critère NN-AC et le critère NN-DR.

On remarque que, lorsque de nombreux appariements sont validés, le critère NN-AC se montre plus performant que le critère NN-DR. La situation inverse se produit lorsque peu d'appariements sont validés. Trois situations aboutissent à un grand nombre de faux appariements. Premièrement, lorsque la description n'est pas suffisamment discriminante, comme c'est le cas lorsque seul le descripteur SIFT est utilisé et que la couleur est ignorée. Deuxièmement, cela peut arriver lors d'une forte transformation géométrique, par exemple lorsque la transformation appliquée correspond à une valeur de 5 du paramètre de tilt. Troisièmement, cela se produit lorsque de nombreux descripteurs requêtes n'ont pas de correspondant dans la base de descripteurs candidats. Ce dernier cas est illustré par le protocole $A \rightarrow A', B$. Dans chacun de ces trois cas, qui peuvent être combinés, le critère de validation NN-AC est le plus performant.

Dans le cas d'une faible transformation, le descripteur candidat est peu modifié et se retrouve ainsi à une faible distance du descripteur requête, le ratio des deux plus proches voisins est alors plus robuste que le calcul du NFA. À l'inverse, lors d'une forte transformation, le descripteur candidat se retrouve plus éloigné du descripteur requête et le rapport des distances aux deux plus proches voisins s'en trouve dégradé. La mesure du NFA est alors plus robuste pour discriminer les bons appariements des mauvais. Ainsi, si le critère NN-AC permet un bon classement des fausses correspondances, il peut en revanche attribuer ponctuellement une forte mesure de dissimilarité à une bonne correspondance. À l'inverse, le critère NN-DR classe de manière précise les bonnes correspondances mais peut attribuer une faible mesure de dissimilarité à un mauvais appariement. De manière similaire, lorsqu'une image A est comparée consécutivement à une version transformée d'elle-même A' puis à une image indépendante B , le critère *a contrario* se montre plus performant pour rejeter les mauvais appariements. Le critère permet une hiérarchisation efficace des fausses mises en correspondance là où le critère NN-DR permet une hiérarchisation précise des bonnes. Ce comportement avait déjà été observé [125], et il est encore accentué si des images supplémentaires, indépendantes de l'image A , sont introduites dans la comparaison. Ainsi, si le nombre attendu de faux appariements est faible, le critère NN-DR s'avère généralement plus performant. Dans le cas inverse, le critère NN-AC est un meilleur choix. En conséquence, suivant qu'il est préférable d'avoir beaucoup de bons ou peu de faux appariements, le choix se portera respectivement sur le critère NN-AC ou NN-DR.

Contrôle du nombre de fausses alarmes

Dans cette expérience, nous évaluons les conséquences sur la prévision de performances des différents choix possibles présentés à la section 5.3.3 pour combiner les distances. Ici, le nombre de fausses alarmes n'est pas connu par avance. Le but est donc de conserver un critère permettant une classification performante et stable des appariements, tout en ayant, pour un seuil donné, une estimation du nombre de fausses alarmes la plus fiable possible. Les combinaisons évaluées ici sont obtenues à partir des 36 distances correspondant aux distances obtenues à partir des 4 sous-descripteurs Hue, Ohta 1, Ohta 2 et SIFT. Les différentes combinaisons testées sont :

- une distance par secteur (cf. Équation (5.10), label « 9 Dist. » sur la Figure 6.13) : les distances des secteurs homologues sont moyennées entre elles,
- une distance par descripteur (cf. Équation (5.11), label « 4 Dist. » sur la Figure 6.13) : les 9 distances de chaque descripteur sont moyennées,

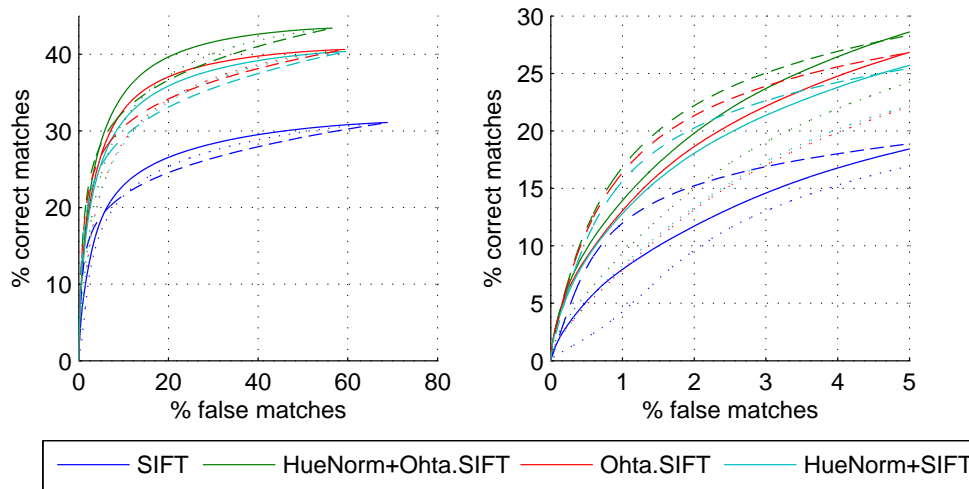


FIG. 6.13 Comparaison des critères NN-AC (traits pleins), NN-DR (traits hachurés) et NN-DT (pointillés) sur les images réelles. La Figure de droite correspond à un zoom sur la Figure de gauche.

- 9 distances pour la géométrie couleur et luminance et une pour le descripteur de teinte (cf. Équation (5.12), label « 10 Dist. » sur la Figure 6.13) : les distances entre secteurs homologues des trois descripteurs Ohta-SIFT sont moyennées, de même que les 9 distances du descripteur HueNorm,
- une distance pour la géométrie et une pour le descripteur de teinte (cf. Équation (5.13), label « 2 Dist. » sur la Figure 6.13) : Les 3×9 distances des descripteurs Ohta-Sift sont moyennées, de même que les 9 distances du descripteur HueNorm.

Toutes les combinaisons testées sont effectuées de manière à ce que les 36 distances aient toutes un poids identique dans la distance finale.

L'évaluation des différentes combinaisons possibles peut être effectuée selon plusieurs critères. Comme précédemment, elle peut porter sur la capacité à ordonner entre eux les plus proches voisins de chaque descripteur requête. Les résultats de cette expérience sont donnés par les courbes ROC de la Figure 6.14. On peut observer que l'influence de la combinaison est minimale, les quatre courbes étant presque parfaitement superposables.

Ce critère ne permettant pas de séparer les différentes possibilités, la sélection peut être effectuée sur la capacité à prédire le nombre de fausses alarmes. Un élément de réponse est donné par les deux graphiques de droite de la Figure 6.14. Sur ces graphiques, on peut observer le ratio de bons matchs ainsi que le nombre de bons et de faux matchs pour une valeur de ϵ donnée. Sur le graphique de droite, on peut observer qu'un contrôle plus juste du nombre de fausses alarmes, s'accompagne d'une plus grande instabilité sur le choix du seuil. C'est-à-dire qu'une variation dans le choix du seuil a plus d'influence sur le nombre de matchs validés. Utiliser plus de distances pour estimer la distribution des distances entre le descripteur requête et les descripteurs candidats, conduit ainsi à une plus grande erreur sur l'estimation du nombre de fausses détections, mais rend moins influent le choix du seuil sur la mise en correspondance.

Le nombre final de distances, utilisé pour l'estimation de la distribution des distances (voir Équation (5.5)) semble ainsi prépondérant sur le choix des combinaisons servant à former ces distances. On peut tout de même remarquer que la courbe verte qui correspond à l'utilisation

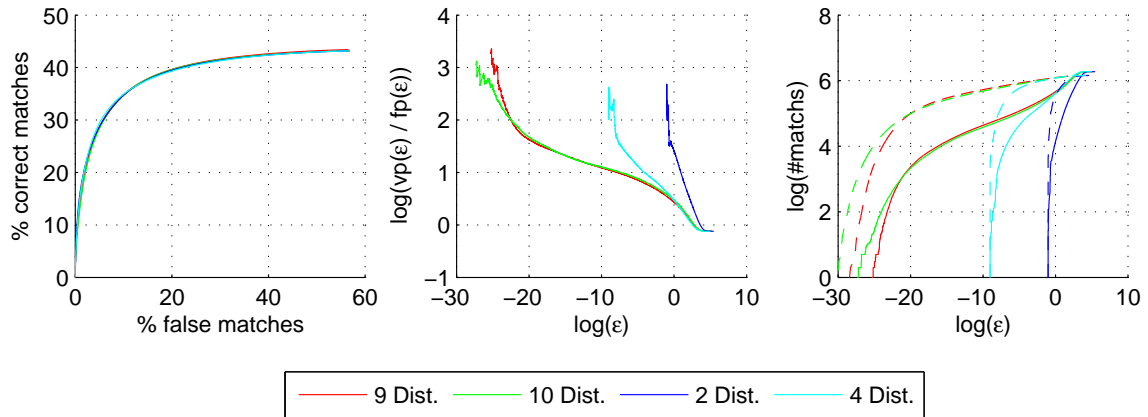


FIG. 6.14 Résultats de mise en correspondance obtenus avec différentes combinaisons de distances pour le critère *a contrario*. Les courbes en rouge correspondent à l'équation (5.10), les courbes en vert à l'équation (5.12), les courbes en bleu à l'équation (5.13) et les courbes en cyan à l'équation (5.11). Le graphique de gauche montre les courbes ROC, celui au centre montre l'évolution du ratio entre bons et mauvais appariements et le graphique de droite le nombre de bons (hachures) et de mauvais (traits pleins) appariements en fonction du seuil ϵ .

de 10 distances ne correspond pas à une augmentation du nombre de fausses alarmes par rapport à la combinaison rouge utilisant 9 distances. Cela accrédite le fait qu'une considération différente du descripteur de teinte par rapport aux descripteurs basés sur l'orientation du gradient a un sens. Un troisième paramètre à prendre en compte pour ce choix est le temps de calcul. Le nombre de distances influe grandement sur la complexité de l'estimation de la distribution des distances. Un grand nombre de distances implique une augmentation du nombre de convolutions nécessaires pour l'estimation de la densité de probabilité, comme le montre la formule (5.5).

Conclusions

Le choix des distances pour la prédiction du nombre de fausses alarmes est avant tout un compromis entre la stabilité recherchée pour le choix du seuil et le temps nécessaire pour l'estimation de la densité de probabilité des distances. En effet, dans nos expériences, le nombre de distances utilisé s'est montré plus influent qu'un choix réfléchi sur la combinaison des distances.

Le critère NN-AC s'est avéré plus robuste que le critère NN-DR en présence d'un grand nombre de faux appariements, confirmant ainsi les résultats précédemment obtenus. Le critère *a contrario* permet d'obtenir un plus grand nombre de bons appariements que le critère NN-DR, mais au prix d'une certaine tolérance sur le nombre de faux. Pour sa part, le critère NN-DR permet davantage de bons appariements lorsque l'on impose un faible nombre de fausses mises en correspondance.

Statistiques des images naturelles couleur

Dans ce chapitre, nous montrons des résultats prospectifs sur des statistiques des images naturelles en couleur. Les statistiques des images naturelles ont été largement étudiées, mais relativement peu de travaux concernent les images en couleur.

L'étude des statistiques des images présente deux intérêts majeurs. Le premier est l'obtention d'une représentation mathématique des images. Cette dernière est requise dans de nombreuses applications comme la reconnaissance de formes, la synthèse de textures ou encore la compression d'images. Le second intérêt est la compréhension de la stratégie utilisée par le cerveau pour représenter des signaux visuels. Le flux d'information visuel que reçoit le cerveau humain est gigantesque, et nécessairement, ce dernier met en œuvre des mécanismes pour réduire la redondance présente dans ce flux. Or l'étude des statistiques des images permet de faire ressortir ces redondances.

Les travaux que nous présentons dans ce chapitre concernent des statistiques d'ordre un et deux, calculées dans des espaces RGB sur un grand nombre d'images. Nous portons un intérêt particulier, d'une part à la stabilité des statistiques observées, et d'autre part à l'influence du processus d'acquisition des images couleur, qui transforme un spectre continu en des données tri-dimensionnelles.

7.1 Travaux connexes

7.1.1 Redondance spatiale

Les premiers travaux se sont intéressés aux statistiques bas niveau, telles que le spectre de puissance ou les covariances spatiales ou encore la décomposition des images dans des bases additives. Cela a permis de mettre en évidence deux caractéristiques majeures des statistiques des images naturelles : leur invariance aux changements d'échelle ainsi que leur non-gaussiannité.

Dans les années 50, les auteurs de travaux sur la compression de signaux télévisuels [30] ont remarqué que le spectre de puissance des images décroissait en $1/f^2$. Il a été montré par Field [39], que cette loi de puissance était directement impliquée par l'invariance de la distribution fréquentielle vis-à-vis du changement d'échelle.

Les transformés en ondelettes ont été utilisés en alternative à l'analyse de Fourier. Les statistiques étudiées concernent alors la distribution des coefficients de ces transformations. Ces distributions, caractérisées par leur forte non-gaussianité [158], peuvent être décrites par une loi Laplacienne généralisée [97].

Plus généralement, la modélisation des statistiques des réponses à des filtres passe-haut a été largement étudiée [90, 130, 137]. Il a été observé par Huang *et coll.* [77] que les distributions résultant de filtres aléatoires de moyenne nulle, respectent la loi de puissance et sont également non gaussiennes à fort kurtosis. Ces travaux ont été utilisés pour la formulation d'*a priori* pour de nombreuses applications, par exemple, pour le débruitage d'images [137], ou encore la détection de contour [34]. La recherche d'une base maximisant la non-gaussianité [120] a mené à l'analyse en composantes indépendantes [8, 78].

Partant du constat que le cerveau ne peut pas traiter directement l'ensemble des informations lui parvenant, Attneave [2], Mackay [96], Barlow [3] ont développé l'idée que le but du codage de source est de supprimer les redondances présentes dans les signaux. Ainsi la compréhension des redondances fournit de précieux renseignements sur la façon dont le cerveau extrait les informations transmises par le système visuel.

Les résultats obtenus par Olshausen *et coll.* [119] montrent que l'analyse en composantes principales fournit une base proche de celle de Fourier, tandis que l'analyse en composantes indépendantes produit des éléments (patches) spatialement orientés. Des résultats similaires ont été observés par Hyvärinen [78].

7.1.2 Redondance chromatique

Les premiers travaux à s'être intéressés à la redondance dans la représentation RGB des couleurs sont ceux d'Ohta [117]. Une des causes plausibles pour la redondance entre les valeurs RGB mesurées est le recouvrement important entre les fonctions de sensibilité [81]. Ces travaux ont été décrits au paragraphe 2.1.5. L'idée est de déterminer la transformation linéaire à appliquer aux canaux RGB afin d'obtenir une matrice de covariance diagonale. Ainsi, l'analyse menée sur une dizaine d'images photographiques dans [117] conduit à la transformation suivante :

$$\begin{cases} Oh_1 = \frac{1}{\sqrt{2}}(R - B) \\ Oh_2 = \frac{1}{\sqrt{6}}(R + B - 2G) \\ Oh_3 = \frac{1}{\sqrt{3}}(R + G + B). \end{cases} \quad (7.1)$$

Les travaux de Buchsbaum [16] abordent le même problème mais du point de vue de la vision humaine et du codage de source. Les fonctions de sensibilité utilisées ne sont alors plus celles d'un appareil photo mais celles des cônes de l'oeil humain. L'espace obtenu par une analyse en composantes principales est alors l'espace antagoniste *Opponent* décrit au paragraphe 2.1.5. Les observations de différents espaces antagonistes faites par Ohta et Buchsbaum *et coll.* posent la question de l'impact des fonctions de sensibilité sur les statistiques observées.

7.1.3 Redondance spatio-chromatique

Les liens entre redondance spatiale et redondance chromatique sont peu abordés dans la littérature. On peut cependant relever que Derrico et Buchsbaum [31] ont développé un

modèle spatio-chromatique pour décrire le codage des images par les cellules photoréceptrices de la rétine. Ce modèle repose sur une décorrélation des canaux RGB permettant de séparer l'information chromatique de l'information de luminance. L'information de luminance est ensuite encodée à l'aide d'un codage prédictif linéaire.

Dans le but de caractériser la redondance de second ordre présente dans les photorécepteurs de la rétine, Ruderman *et coll.* [131], ont utilisé des images multi-spectrales de feuillage, puis ont estimé la réponse des cônes humains à ces images. Ils ont déterminé par ACP la transformation décorrélant ces réponses. Cette transformation est similaire à celle observée par Buchsbaum. En appliquant ensuite une ACP sur des patchs de taille 3×3 , Les auteurs remarquent une décorrélation complète entre les variations spatiales et chromatiques.

Wachtler *et coll.* [152] ont appliqué l'analyse en composantes indépendantes à des spectres provenant d'images multi-spectrales. Ils obtiennent un espace antagoniste similaire à celui obtenu par Ruderman [131]. Ils remarquent cependant que pour des tailles de patchs supérieures à 7×7 , l'hypothèse d'indépendance entre variations spatiales et chromatiques ne tient plus.

7.2 Présentation de la base d'images

Depuis quelques années, les sites de partage de photos en ligne mettent à disposition de leurs utilisateurs de gigantesques bases d'images.

Parallèlement à cela, le développement de la puissance de calcul des ordinateurs permet de traiter de larges volumes de données. Il est désormais possible de faire de la reconnaissance d'objets à partir de millions d'images [143]. Avec une base d'images classées suffisamment grande, une simple recherche de type plus proche voisin permet d'effectuer une classification performante de l'image requête. Mais combien d'images faut-il considérer pour que cette reconnaissance soit robuste? Existe-t-il une relation entre le nombre d'images utilisées et la probabilité de trouver une image similaire ?

D'autres auteurs utilisent plusieurs millions d'images pour la complétion de zones manquantes dans une image [72]. Conjointement avec Torralba *et coll.* [143], les auteurs notent que l'utilisation de plus d'un million d'images permet une augmentation significative des performances. En lien avec ces problématiques, le problème de la recherche efficace d'images similaires dans une base de 10 millions d'images est également abordé par Jegou *et coll.* [80, 154].

Dans ce chapitre, nous travaillons à partir d'une base de 6 millions d'images [72]. De cette base sont extraites 470,000 images ayant une résolution de 1024×768 .

Origine Ces images proviennent du site de partage en ligne de photographies Flickr. La Figure 7.1 indique l'origine géographique de photographies géolocalisées (qui sont différentes de la base). On peut observer que ces photos sont logiquement très concentrées autour de points d'intérêts précis. Les photos de la base sont classées par pays, mais nous les avons réparties aléatoirement de façon à pouvoir établir des statistiques indépendantes du lieu de prise de vue.

Biais éventuels Les photographies proviennent de différents appareils. Malheureusement, les images de la base ne possèdent pas dans leurs données EXIF d'informations sur l'appareil

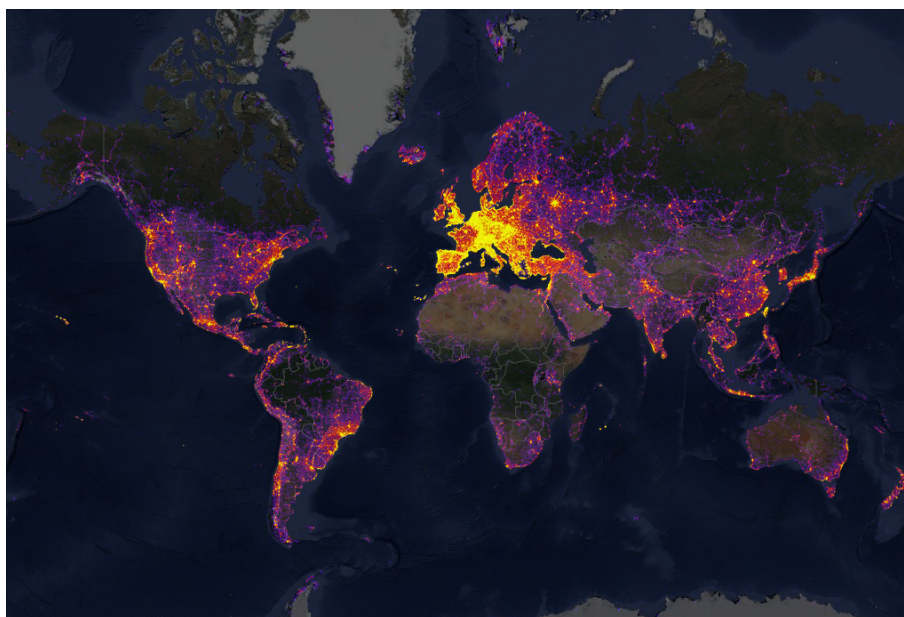


FIG. 7.1 Lieux les plus photographiés dans le monde (<http://www.sightsmap.com/>).

d'origine. La validité des statistiques réalisées à partir de données provenant de différents capteurs n'est donc pas immédiate. Cependant, les fonctions de sensibilité des appareils photographiques sont souvent très similaires [81], on suppose donc que moyenniser les observations a du sens.

Deux phases de traitement font suite à la capture de l'image. La première est celle regroupant les modifications réalisées de manière automatique ou non par l'appareil photographique ou un utilisateur. Ces modifications peuvent concerner un changement dans la balance des blancs, l'application d'une correction de type rehaussement de contours ou bien la compression. La seconde phase est celle appliquée lors de la mise en ligne. L'image est alors redimensionnée et recompressée de façon à correspondre aux standards choisis par l'hébergeur. Ces deux phases peuvent éventuellement introduire des biais sur les statistiques observées.

7.3 Distributions couleur dans les photographies

7.3.1 Distributions de valeurs RGB

Nous proposons ici d'observer les statistiques des lois jointes et marginales de distributions issues des valeurs RGB, comme par exemple, la distribution de la teinte, de la saturation de la luminance et des canaux antagonistes. Pour les expériences, nous considérons les 470,000 images décrites précédemment. Ces images sont redimensionnées à une résolution de 96×128 . Pour chacune de ces images nous calculons plusieurs statistiques élémentaires de type histogramme qui sont ensuite moyennées.

Distribution RGB 3D

La distribution jointe des valeurs RGB, obtenue à partir de tous les pixels de la base, est illustrée par la Figure 7.2. Cette distribution est quantifiée sur 32 classes. Sur ces figures, on peut remarquer que les axes sur lesquels se concentre la masse sont principalement l'axe de luminance et l'axe bleu/rouge.

Lois marginales

Distribution de luminance et lois marginales R , G et B La luminance est ici définie comme la moyenne des canaux RGB. Comme expliqué précédemment, l'histogramme de luminance moyen est obtenu en moyennant les histogrammes de luminance des 470,000 images. Il est visible sur la Figure 7.3(a). Les grandes valeurs correspondant aux fortes intensités sont causées par des images pour lesquelles le ciel est présent et fortement saturé.

Les histogrammes R , G et B sont obtenus de la même manière que l'histogramme de luminance. Ils sont visibles sur la Figure 7.3(b). Les histogrammes de vert et de rouge sont quasiment identiques et possèdent une allure similaire à celle de l'histogramme de luminance. L'histogramme de bleu possède une « bosse » dans les faibles intensités. Cette bosse semble être en partie causée par des images prises la nuit. Dans ces conditions, l'éclairage est alors souvent artificiel avec une faible température de couleur. Ce qui donne une teinte jaunâtre aux images (mélange de rouge et de vert). Une autre cause possible pour cette « bosse » est que certaines images ne possèdent que très peu de couleur bleue. Dans ce cas, la dynamique du canal bleu est sous exploitée et les valeurs se concentrent dans les faibles intensités, ce qui est illustré par la Figure 7.4.

Distribution de la teinte L'histogramme de teinte moyen, visible sur la Figure 7.5(a) permet de voir que les couleurs les plus représentées dans les images sont le bleu et le rouge/orange. Cet histogramme fait aussi ressortir les problèmes de quantification et de compression JPEG.

La compression JPEG est effectuée dans l'espace YCbCr, qui est un espace de type antagoniste. La compression a pour effet de quantifier les valeurs prises par les canaux chromatiques, ce qui limite en conséquence les valeurs de teinte possibles. La conséquence de cette compression s'observe sur l'histogramme de teinte mesurée et se manifeste par des valeurs sur-représentées et des artefacts de quantification importants. Dans le cas d'une compression extrême, les seules valeurs de teinte possibles correspondent aux quatre couleurs définissant les contrastes Cb/Cr. Une illustration des effets de la compression est donnée sur la Figure 7.6.

Histogrammes de saturation Sur la Figure 7.5(b) apparaît l'histogramme de saturation moyen. Il est relativement lisse et semble bien décrit par une loi exponentielle. L'hypothèse de distribution exponentielle pour la saturation est utilisée dans l'Annexe A comme *a priori* pour la détection de modes dans les histogrammes pondérés.

Canaux antagonistes La distribution des valeurs des canaux antagonistes Ohta est donnée sur la Figure 7.7. Tout comme pour la distribution jointe des valeurs RGB, ces distributions mettent en évidence la plus grande variance observée sur l'axe bleu/rouge que sur l'axe magenta/vert. On peut également observer que ces distributions ne sont pas symétriques.

Distribution couleur 2D On regarde ici la distribution des couleurs RGB après projection dans le triangle de Maxwell (cf. Section 2.1.2) c'est à dire dans le diagramme de chromaticité rg , ($r = R/(R+G+B)$, $g = G/(R+G+B)$). Ici encore, on observe logiquement une concentration des valeurs le long de l'axe de couleur rouge/bleu. On observe également de forts effets de quantification. Ces effets sont probablement dus à la compression JPEG. La distribution des couleurs est montrée sur la Figure 7.8

Le monde est-il gris ?

Nous proposons ici de vérifier les hypothèses classiques pour l'estimation de l'illuminant. Les trois hypothèses les plus communes sont :

- la réflectance moyenne est achromatique, ce qui correspond à l'hypothèse de *Grey-World*,
- la dérivée de la réflectance est en moyenne achromatique, ce qui correspond à l'hypothèse *Grey-Edge*,
- la zone de plus forte luminance correspond à l'illuminant, ce qui constitue l'hypothèse *White-Patch*.

Toutes ces hypothèses ont été appliquées successivement à chacune des images de la base. Pour chacune des images, il résulte un triplet de valeurs RGB. Chacun de ces triplets est alors converti en coordonnées rg . Pour chaque hypothèse, un histogramme 2D est alors calculé. Ces distributions peuvent être interprétées comme une estimation de la distribution des illuminants dans le monde. Évidemment, il est très fréquent qu'une balance des blancs automatique soit faite par l'appareil photographique. Si l'on considère que la balance des blancs de l'appareil est parfaite, alors cette distribution renseigne sur la valeur de l'hypothèse. À l'inverse, si l'on suppose que l'erreur de la balance des blancs de l'appareil est isotropique, alors c'est bien une estimation de la distribution des illuminants que fournit cet histogramme.

La Figure 7.9 montre les couleurs moyennes d'illuminants obtenues avec chacune des trois hypothèses. Pour l'hypothèse *Grey-World*, le résultat était en fait déjà connu puisqu'il correspond à la valeur moyenne de chaque histogramme R , G , et B . Ces trois résultats sont la conséquence de la moindre présence de la couleur bleue dans les images.

Sur la Figure 7.10, sont illustrées les distributions d'illuminants obtenues avec chacune des hypothèses. Les trois distributions sont différentes, on peut cependant observer que toutes ces distributions ont une plus forte variance sur l'axe bleu/rouge. Cela est à mettre en relation avec l'hypothèse, faite au chapitre 3, que les illuminants sont situés près du lieu de Planck, présenté à la Section 2.2.1.

Stabilité des distributions

Dans cette partie, nous regroupons les images en paquets de différentes tailles de façon à observer la stabilité des distributions. Pour chaque paquet, des distributions couleur sont calculées et nous observons comment se comportent les distances entre ces distributions en fonction de la taille du paquet. Le choix de la mesure de distance est sans doute très important. Pour ces résultats préliminaires, nous utilisons une distance L_1 entre distributions couleur. Les distances sont calculées sur des distributions de somme unitaire et sont normalisées par le nombre de classe utilisées. D'autres choix de distances devront être considérés, notamment des distances de transport.

Dans l'expérience proposée, 470 paquets de 1,000 et 47 paquets de 10,000 sont considérés. On s'attend à ce que les distances mesurées entre distributions couleurs, de même que la variance de ces distances, décroissent avec la taille du paquet. La Figure 7.11 donne les erreurs mesurées sur les distributions jointes RGB. Pour les marginales, les erreurs sont données sur la figure 7.12. Ces figures confirment les attentes : l'erreur moyenne et la variance des erreurs décroissent avec la taille du paquet. Cette expérience ne permet cependant pas de savoir si l'ordre de grandeur du nombre d'images utilisées est suffisant. Il reste maintenant à caractériser la décroissance de l'erreur.

7.3.2 Liens entre espaces antagonistes et fonctions de sensibilité

Comme expliqué précédemment, deux espaces antagonistes cohabitent dans la littérature. La différence la plus marquante dans les expériences menant à ces résultats réside dans les fonctions de sensibilité utilisées pour échantillonner le spectre. Ohta [117] utilise les fonctions de sensibilité d'un appareil photographique, tandis que Buchsbaum [16] et Ruderman [131] utilisent les fonctions de sensibilité estimées de l'œil humain. On peut donc supposer que l'espace antagoniste obtenu par analyse en composantes principales des triplets RGB dépend des fonctions de sensibilité de l'appareil photo.

Pour apporter un élément de réponse corroborant cette hypothèse, nous réalisons l'expérience suivante : à l'aide de 10^9 triplets RGB provenant de la base, nous appliquons l'analyse en composantes principales. La matrice permettant la décorrélation des canaux RGB est alors :

$$\begin{bmatrix} 0.57515 & 0.58326 & 0.57340 \\ 0.71005 & -0.00774 & -0.70411 \\ -0.40624 & 0.81225 & -0.41860 \end{bmatrix} \quad (7.2)$$

Ce qui correspond tout à fait à l'espace d'Ohta. Cette observation est en accord avec la stabilité des fonctions de sensibilité entre appareils photographiques [81].

En supposant que l'ensemble des images est dans l'espace de couleur sRGB, il est possible d'estimer les réponses des cônes en passant dans le système LMS à l'aide du modèle CIECAM02 [111]. Ce système repose sur une estimation des fonctions de sensibilité de l'œil humain. Premièrement, la correction gamma est retirée en supposant un exposant de 2.2. La matrice suivante est ensuite appliquée de manière à transformer les coordonnées RGB en coordonnées LMS

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \begin{bmatrix} 0.39040 & 0.54994 & 0.00898 \\ 0.07084 & 0.96317 & 0.00136 \\ 0.02311 & 0.12802 & 0.93595 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (7.3)$$

Cette matrice est le résultat d'une transformation de l'espace sRGB vers l'espace XYZ puis d'une transformation de l'espace XYZ vers l'espace LMS. La matrice obtenue en appliquant l'analyse en composantes principales aux pixels dans l'espace LMS est alors :

$$\begin{bmatrix} 0.53761 & 0.58800 & 0.60431 \\ -0.50930 & -0.34472 & 0.78853 \\ -0.67200 & 0.73170 & -0.11416 \end{bmatrix}. \quad (7.4)$$

Cette matrice correspond grossièrement aux contrastes antagonistes *Opponent* et vont dans le sens de l'hypothèse selon laquelle les fonctions de sensibilité jouent un rôle déterminant dans l'espace antagoniste résultant d'une analyse en composantes principales. Cette influence pourra être caractérisée à l'aide d'images multi-spectrales.

7.4 Analyse en composantes principales spatio-chromatiques

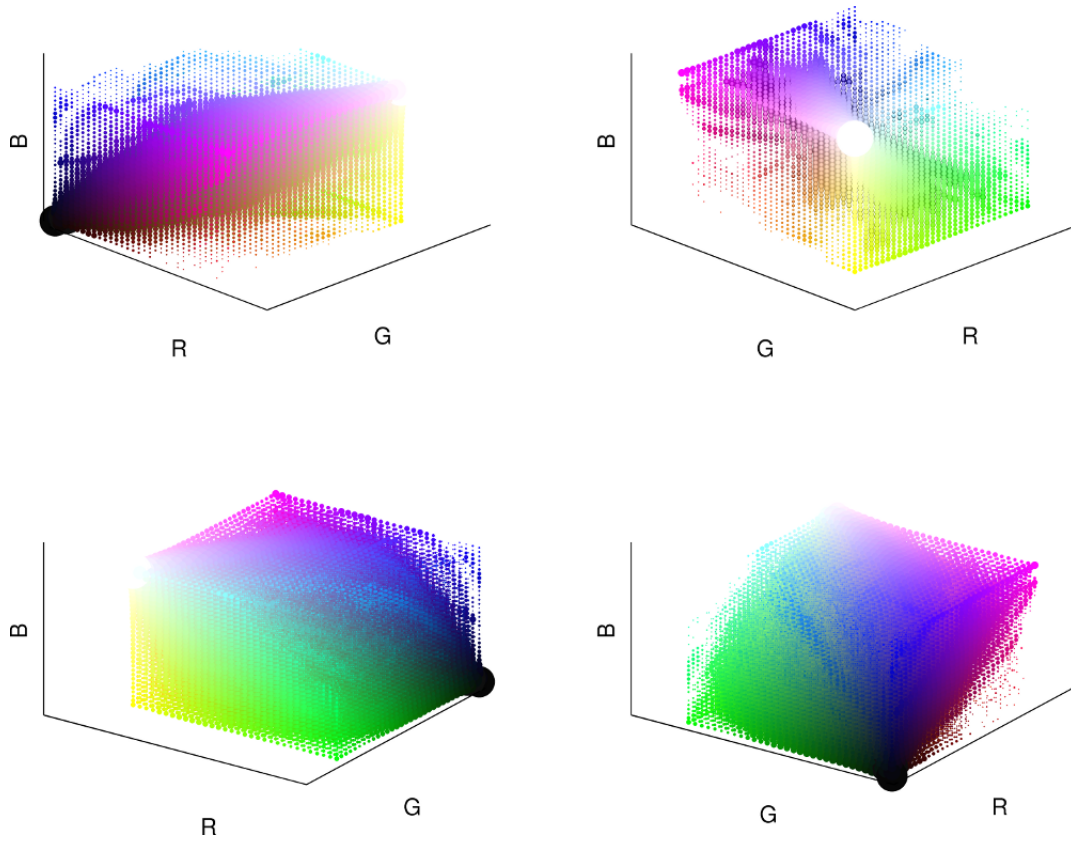
Image moyenne Conformément à l'intuition, l'image moyenne tend vers une image grise uniforme. Néanmoins, on peut observer sur la Figure 7.13, que la partie haute est plus claire. Ceci est visible en normalisant la dynamique de l'image. On peut alors distinguer le ciel bleu, et dans le centre une teinte orangée. Les bords gauche et droit de l'image sont visiblement plus sombres. Une hypothèse permettant d'expliquer partiellement est le vignettage (atténuation lumineuse sur les bords de l'image due à l'appareil photographique). Une autre raison est que le sujet de la photographie est généralement au centre de l'image et est correctement éclairé.

Analyse en composantes principales d'images La Figure 7.14 montre les 20 premières composantes, résultant d'une analyse en composantes principales appliquée à 4 paquets contenant chacun 100,000 images d'une résolution de 96×128 . Pour les niveaux de gris, on voit apparaître des vecteurs semblables à une base de Fourier. Concernant les couleurs, on voit clairement apparaître les canaux antagonistes de l'espace Ohta.

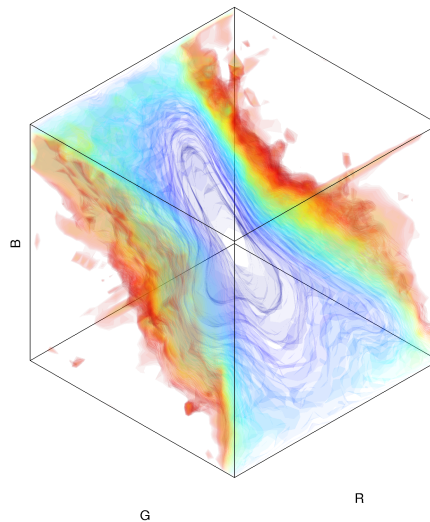
Ces résultats sont à mettre en perspectives avec l'analyse en composantes principales appliquée sur 1×10^6 patchs de dimension 32×32 illustrée par la Figure 7.15. Les Figures 7.17 et 7.16 donnent des exemples d'images possédant un fort coefficient associé à une des premières composantes principales. Une simple expérience de reconstruction d'images à l'aide d'un pourcentage des composantes principales est illustrée par la Figure 7.18.

7.5 Conclusion

Dans ce chapitre ont été présentés quelques résultats de travaux en cours sur les statistiques des images naturelles couleur. Une première piste suivie est l'étude de la covariance couleur et sa possible factorisation avec une base de Fourier. Une seconde piste suivie par ces travaux est la caractérisation de la vitesse de convergence de statistiques couleur. Une autre piste d'intérêt est l'étude de l'influence des fonctions de sensibilité sur les espaces antagonistes résultant de l'ACP. Cette étude sera réalisée à l'aide d'images multi-spectrales.



(a) Sphères



(b) Lignes de niveaux

FIG. 7.2 Vues de la distribution des valeurs RGB obtenues à partir de 470,000 images RGB d'une résolution de 192×256 . Sur les 4 Figures du haut, chaque classe de l'histogramme est représentée par une sphère de la dont la couleur correspond est la couleur moyenne des pixels contribuant à cette classe et dont la taille est proportionnelle au nombre de relatif de pixels contribuant à cette classe. La dernière Figure est une représentation à l'aide de lignes de niveaux.

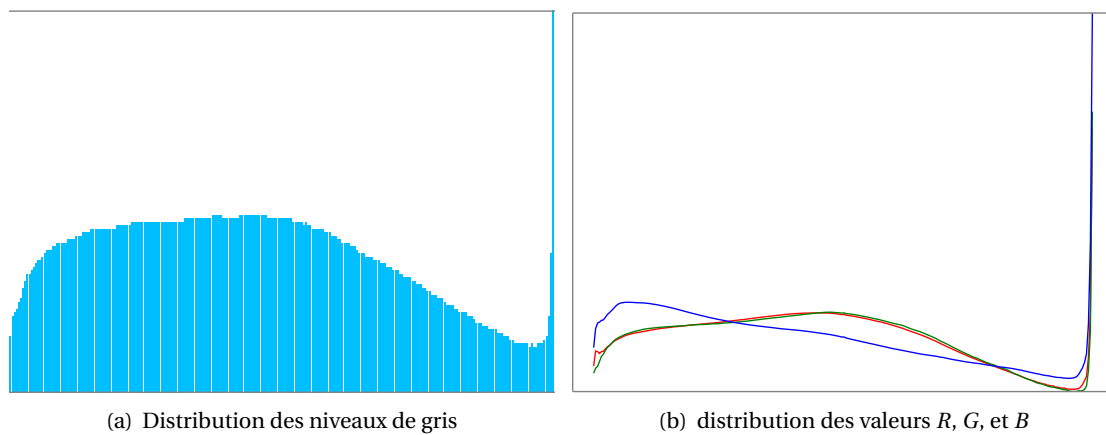


FIG. 7.3 Distributions obtenues à partir de 470,000 images d'une résolution de 96×128 . Les distributions sont quantifiées sur 255 classes.

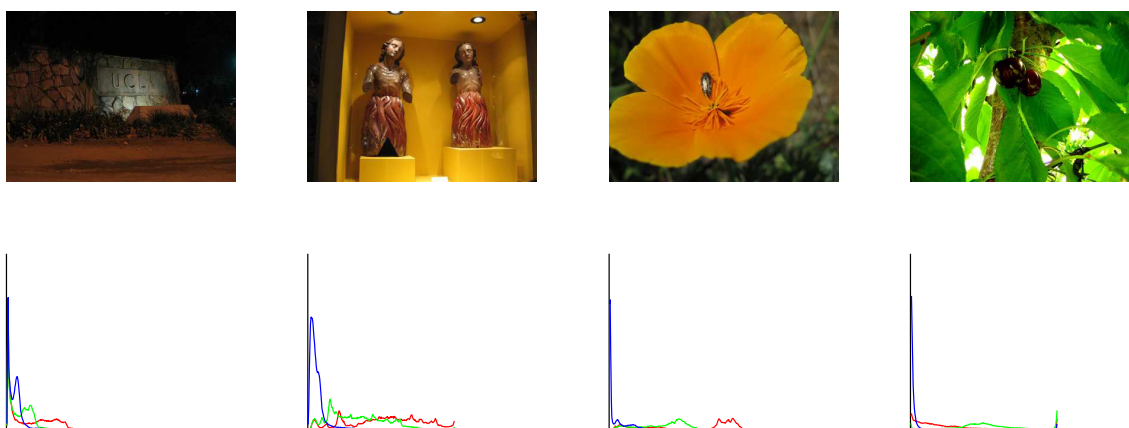


FIG. 7.4 Cause de la « bosse » dans les faibles intensités de l'histogramme du canal bleu. Les distributions sont quantifiées sur 255 classes.

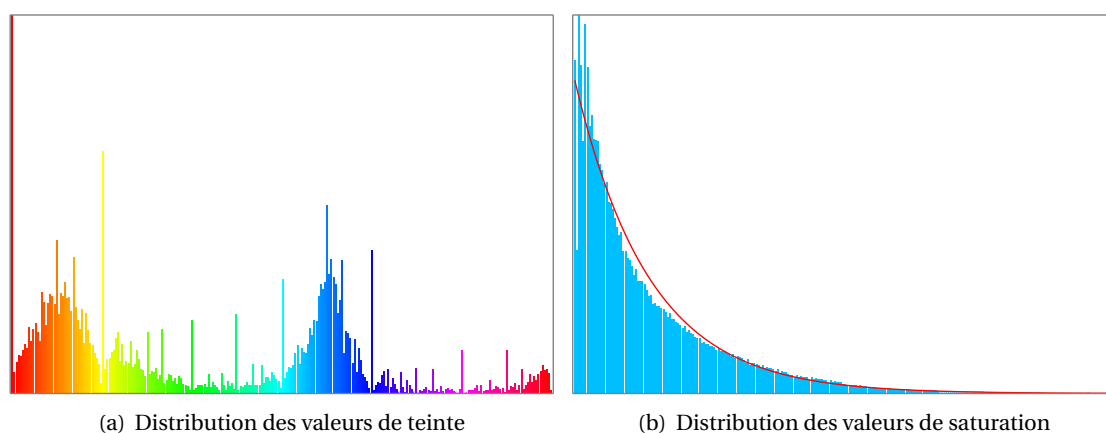


FIG. 7.5 Distributions obtenues à partir de 470,000 images d'une résolution de 96×128 .

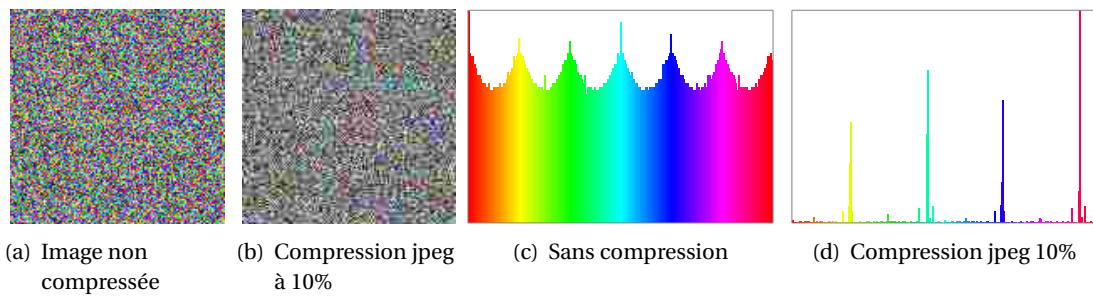


FIG. 7.6 Effets de la quantification JPEG sur la distribution de teinte. Exemple sur une réalisation d'un bruit blanc uniforme.

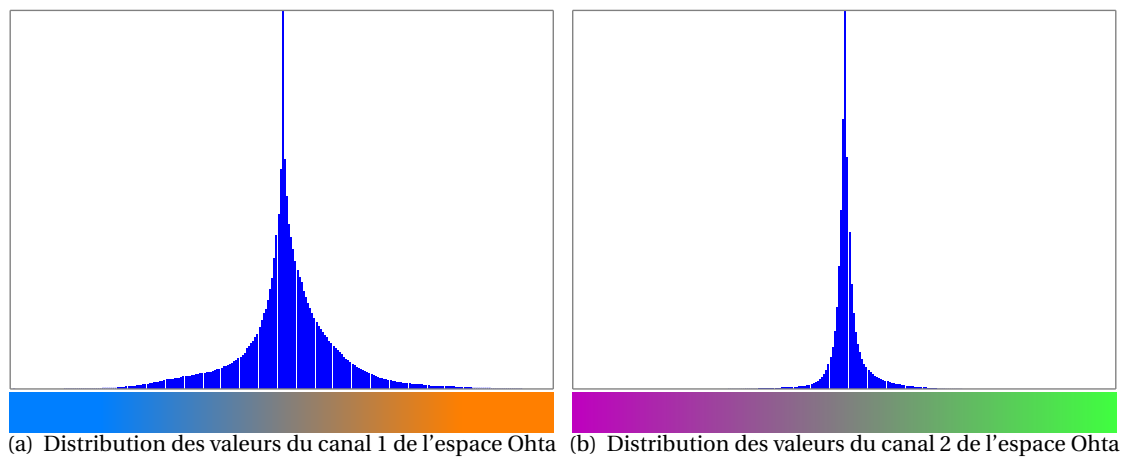


FIG. 7.7 Distributions obtenues à partir de 470,000 images d'une résolution de 96×128 .

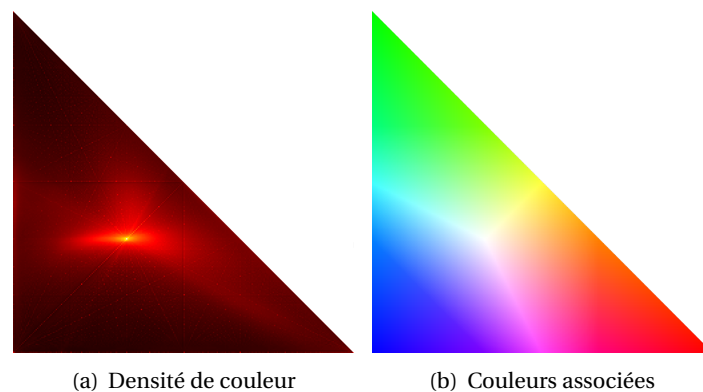


FIG. 7.8 La Figure 7.8(a) donne la projection de la distribution des pixels RGB dans le triangle de Maxwell. Cette distribution est élevée à la puissance $1/10$ pour la visualisation. En réalité, elle est bien plus piquée autour du point gris. Les couleurs correspondantes sont données par la Figure 7.8(b)



FIG. 7.9 Couleur moyenne obtenue suivant l'hypothèse utilisée. La luminance de chaque patch a été ajustée de façon à être identique. La couleur normalisée est obtenue en étirant la dynamique couleur tout en conservant l'intensité inchangée. Il n'a pas été tenu compte des éventuelles saturations pour calculer ces patches.

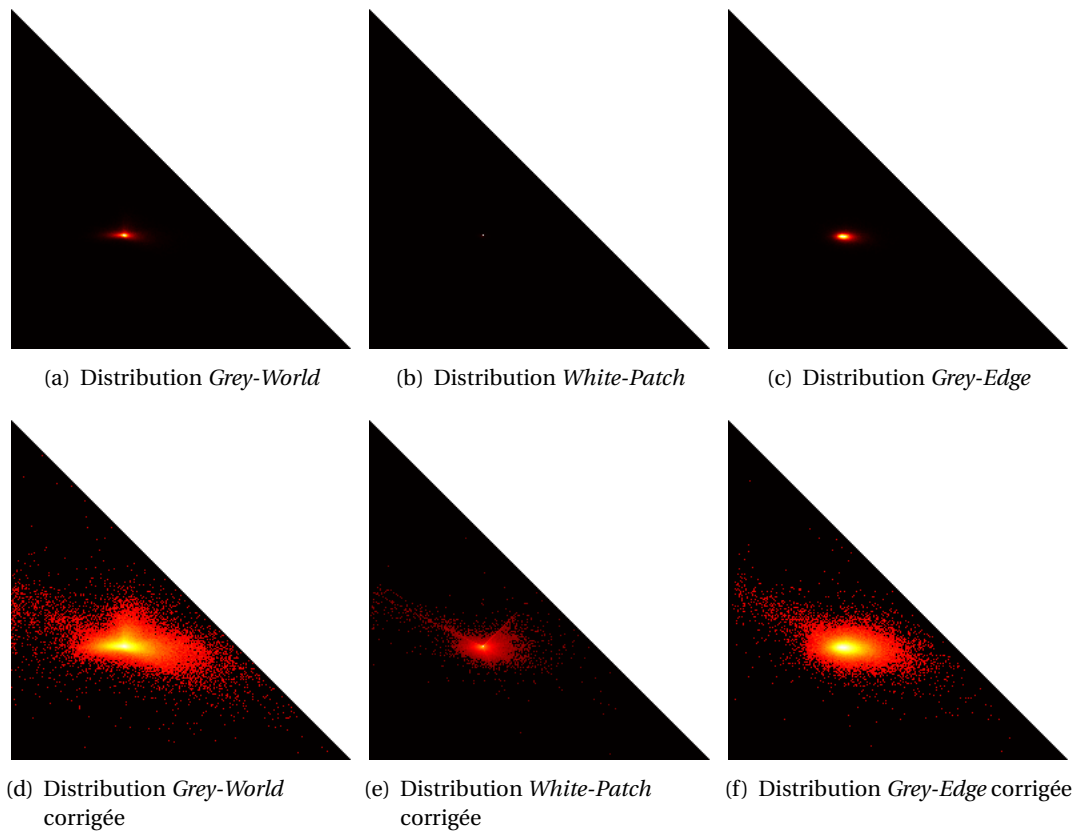


FIG. 7.10 Distributions obtenues suivant l'hypothèse faite sur la scène. Les couleurs associées à la densité sont données par la Figure 7.8(b). La correction est obtenue en appliquant une puissance 0.15 aux distributions. La distribution correspondant à l'hypothèse *White-Patch* est très piquée, sur l'image non normalisée, seul un pixel est allumé.

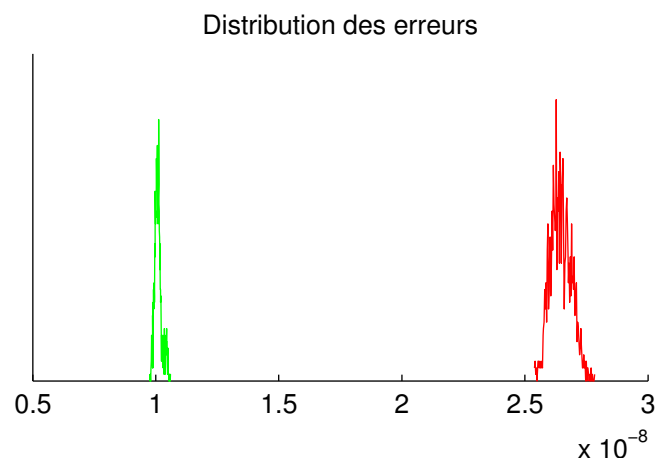


FIG. 7.11 Histogramme des erreurs entre distributions couleur jointes RGB. En vert pour des distributions calculées à partir de 10,000 images et en rouge sur 1,000 images. Les distributions sont normalisées et les erreurs sont en erreurs moyennes par classe.

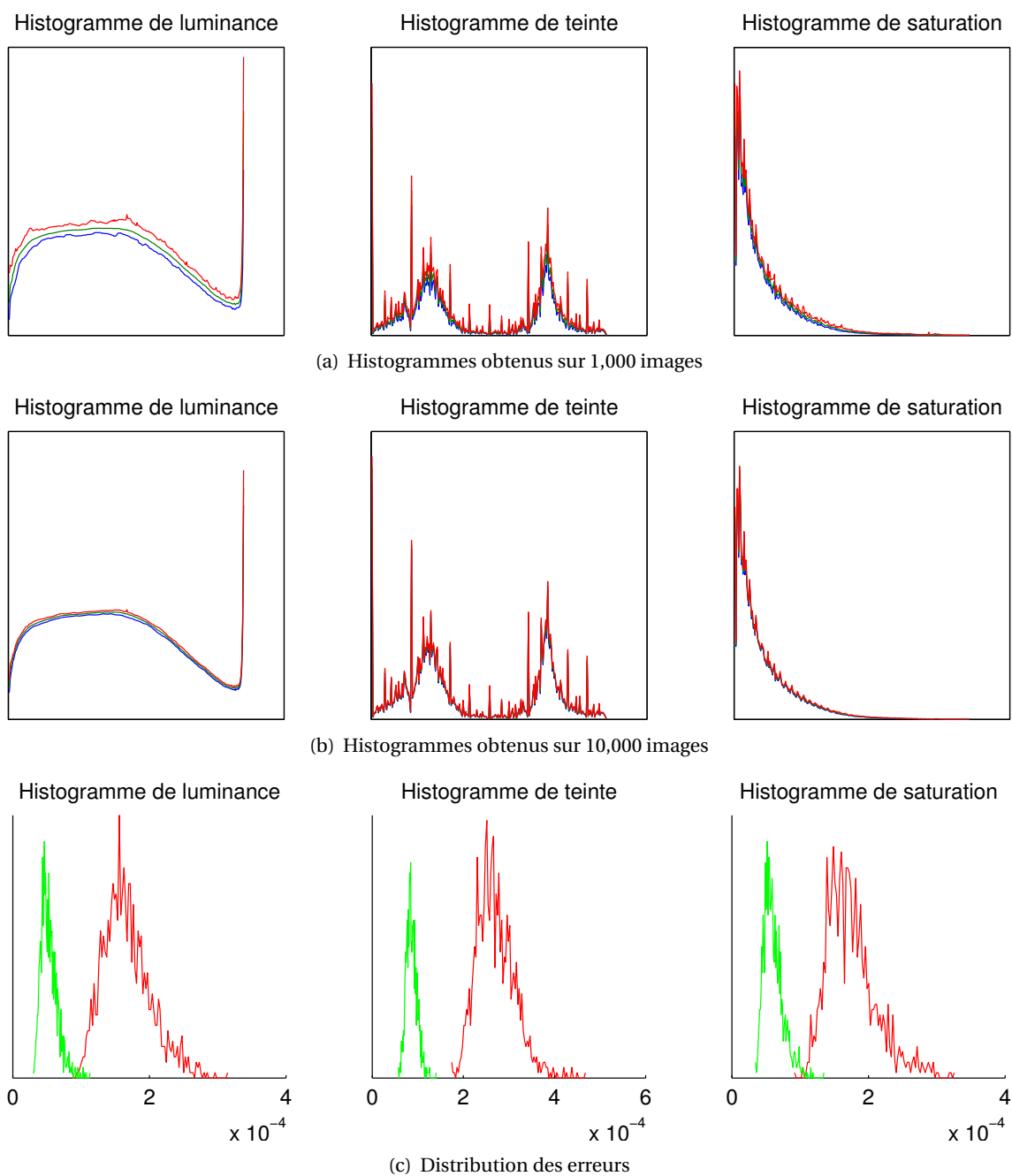


FIG. 7.12 Sur les deux premières lignes sont tracées en vert les histogrammes moyens de luminance, de teinte et de saturation obtenus sur 470,000 images. Également sur ces deux lignes, les courbes rouges et bleues représentent les valeurs extrêmes prises par l'ensemble des histogrammes moyens obtenus à partir de 1,000 images et 10,000 images. Sur la dernière ligne, les distributions des différences mesurées entre histogrammes à partir de paquets de 1,000 images sont représentées en rouge et celles des différences mesurées entre histogrammes provenant de paquets de 10,000 sont représentées en vert. Les distributions sont normalisées et les erreurs sont en erreurs moyennes par classe.

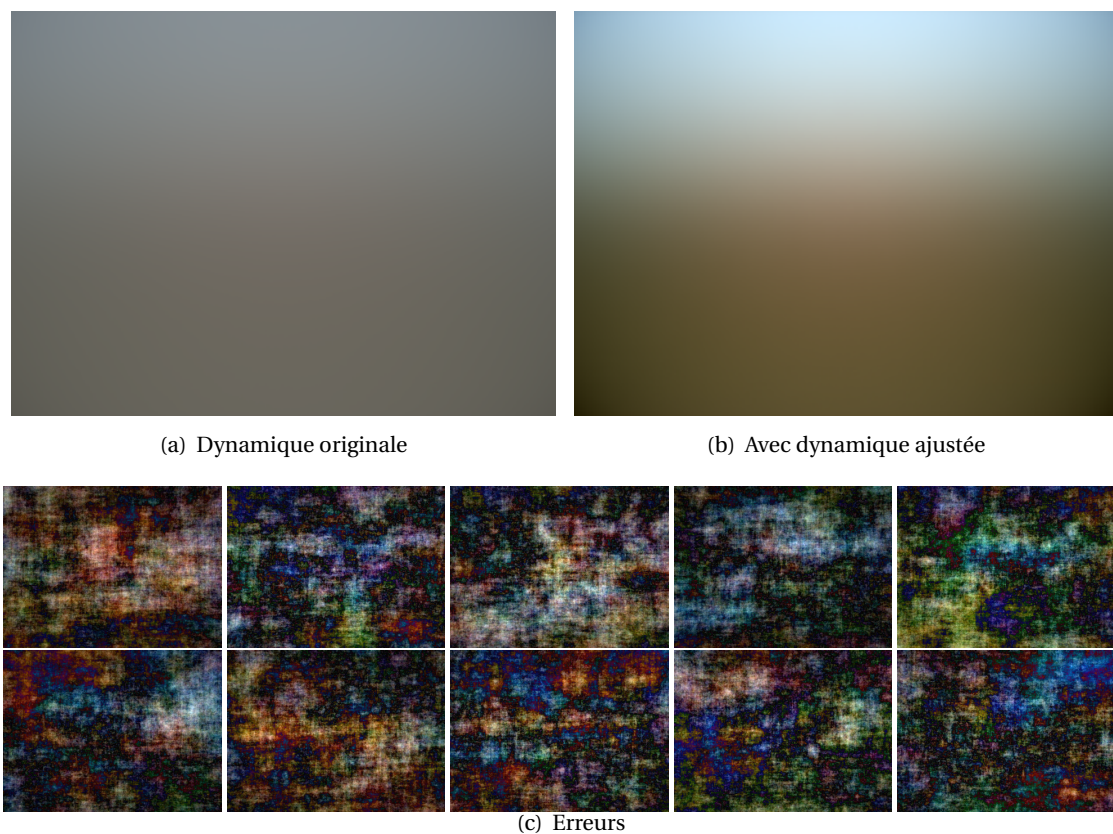


FIG. 7.13 Image moyenne obtenue à partir de 470,000 images RGB. Chaque image d'erreur de la Figure 7.13(c) correspond à la soustraction entre une image moyenne calculée sur 10,000 images et l'image moyenne de la Figure 7.13(b). La dynamique de l'image résultante est ajustée pour la visualisation. Cet ajustement est réalisé de façon à ce que la valeur RGB maximale corresponde à 255 et la minimale à 0.

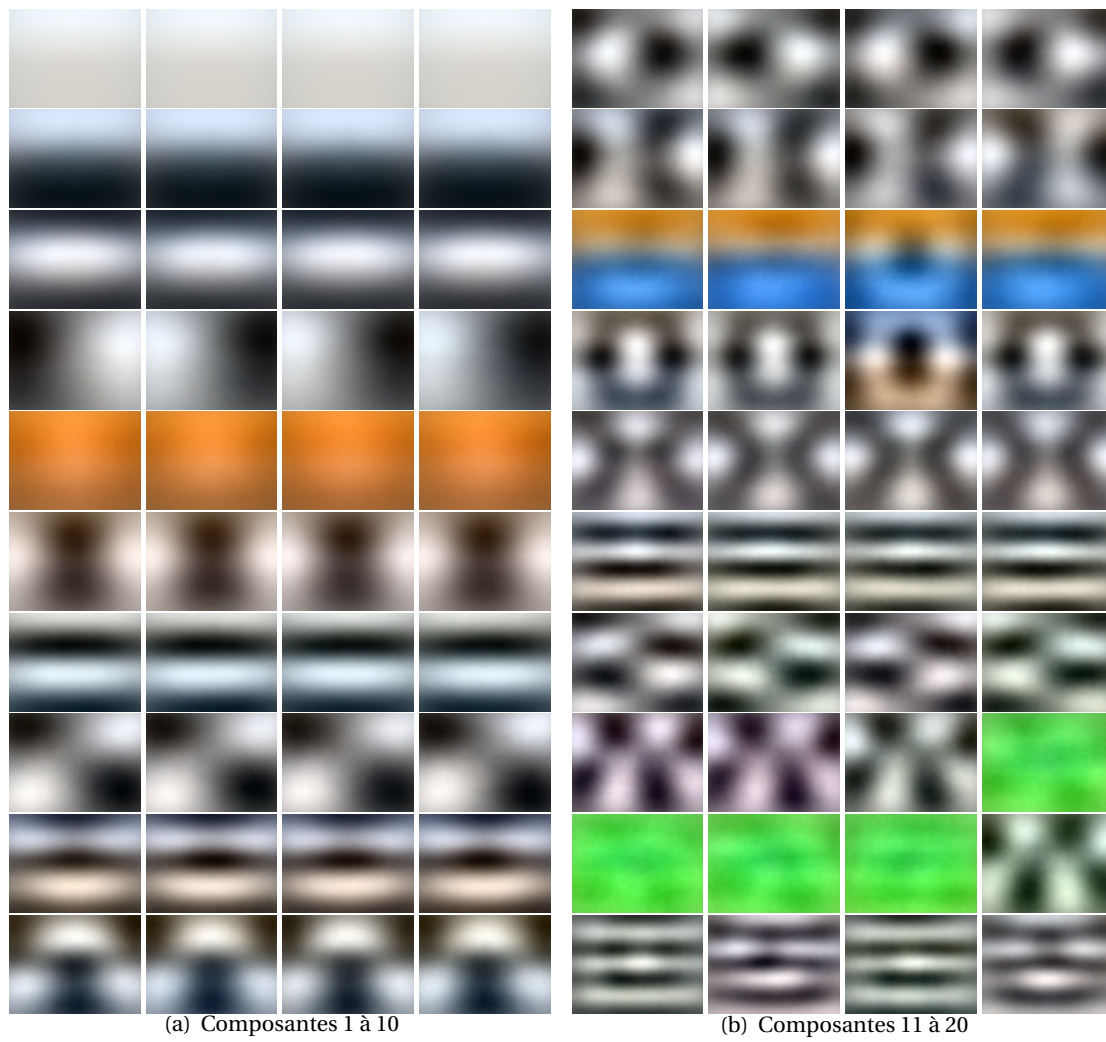


FIG. 7.14 Les 20 premières composantes principales obtenues sur 4 paquets de 100,000 images.

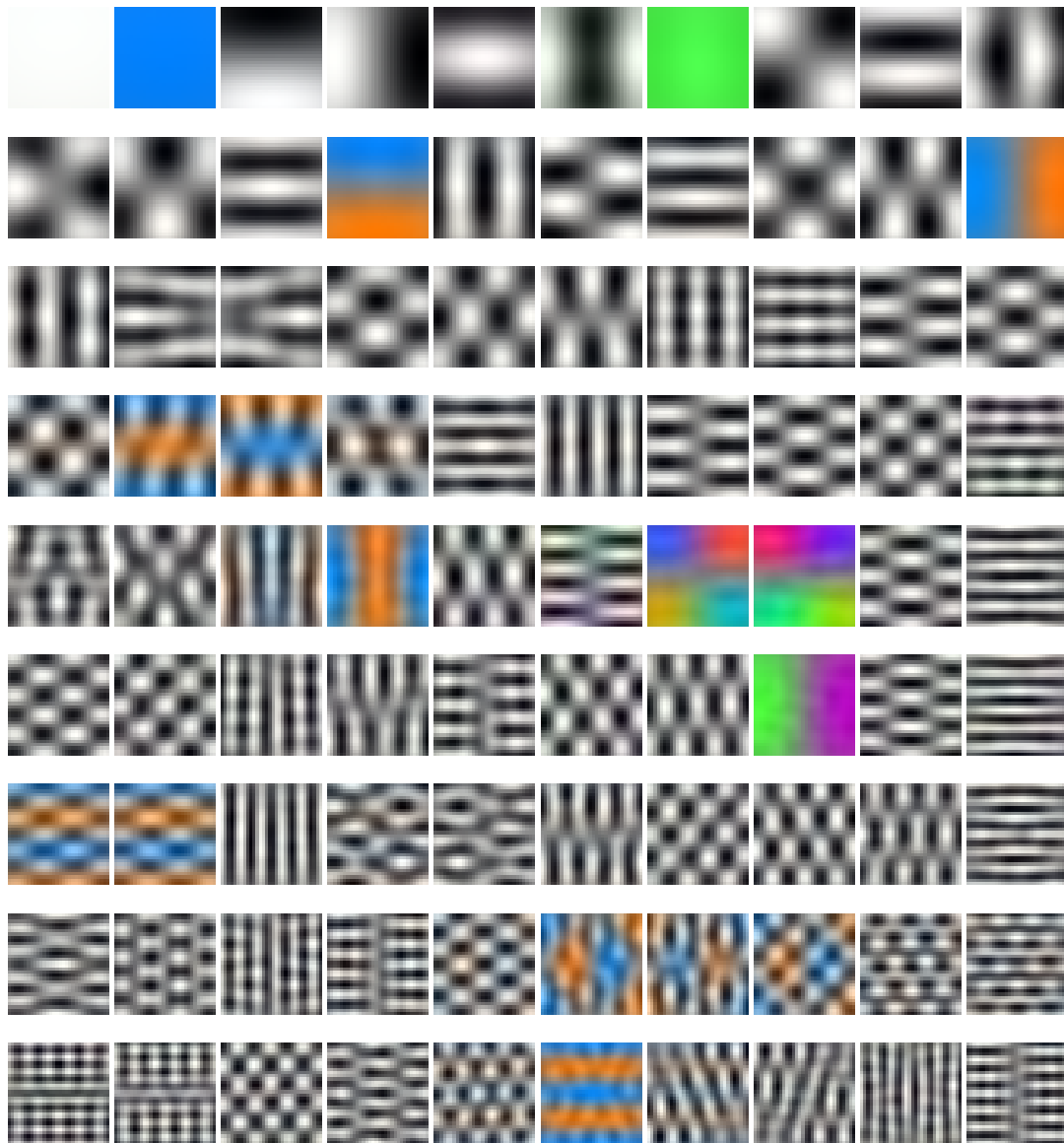


FIG. 7.15 Composantes principales obtenues à partir de 1,000,000 de patches d'un dimension 32×32 .

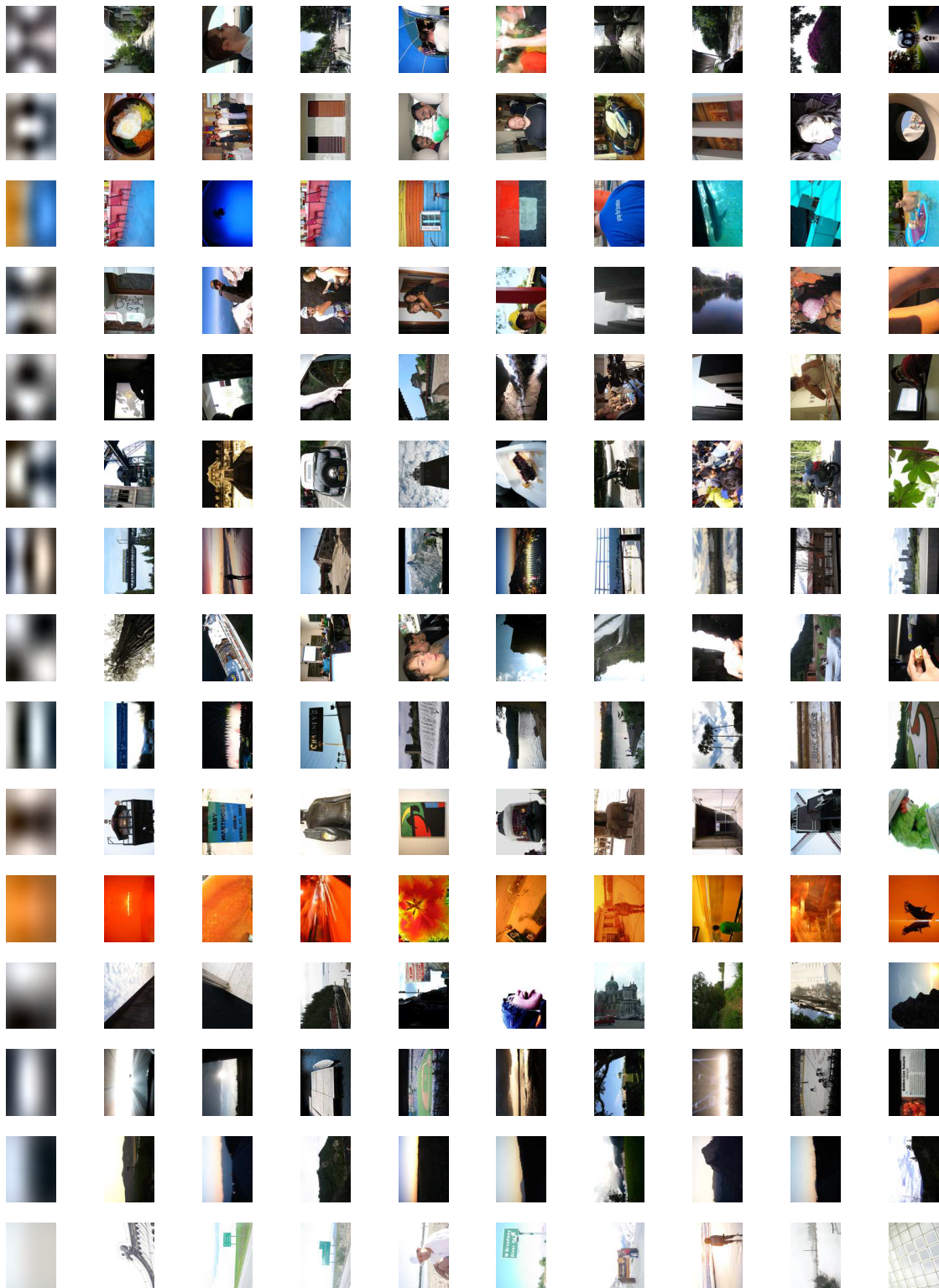


FIG. 7.16 Chaque ligne correspond à une composante principale et fournit des exemples d'images dont la décomposition possède un fort coefficient positif associé à cette composante.

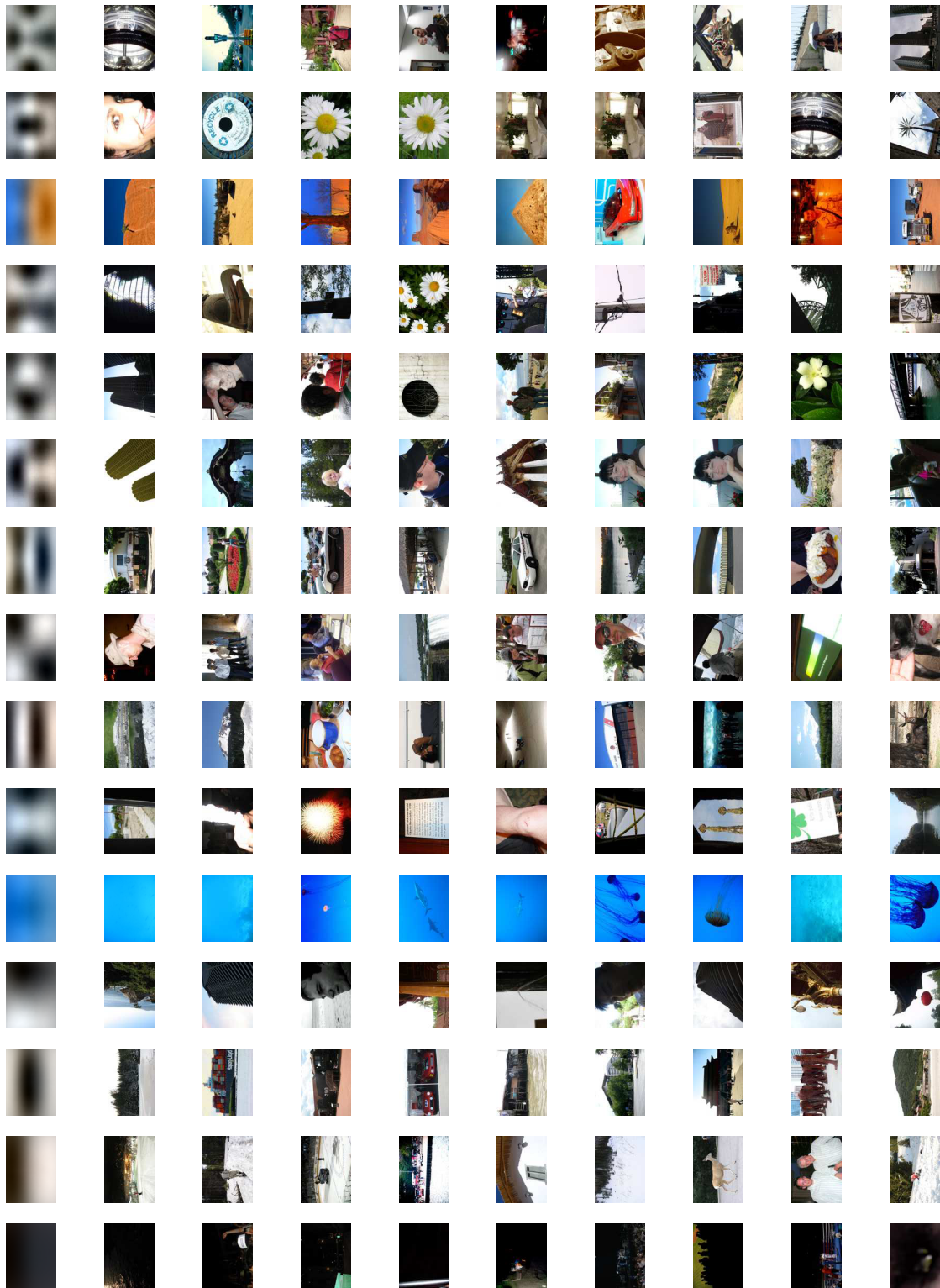


FIG. 7.17 Chaque ligne correspond à une composante principale et fournit des exemples d'images dont la décomposition possède un fort coefficient négatif associé à cette composante. La composante représentée par la première image de chaque ligne est représentée par son opposée.

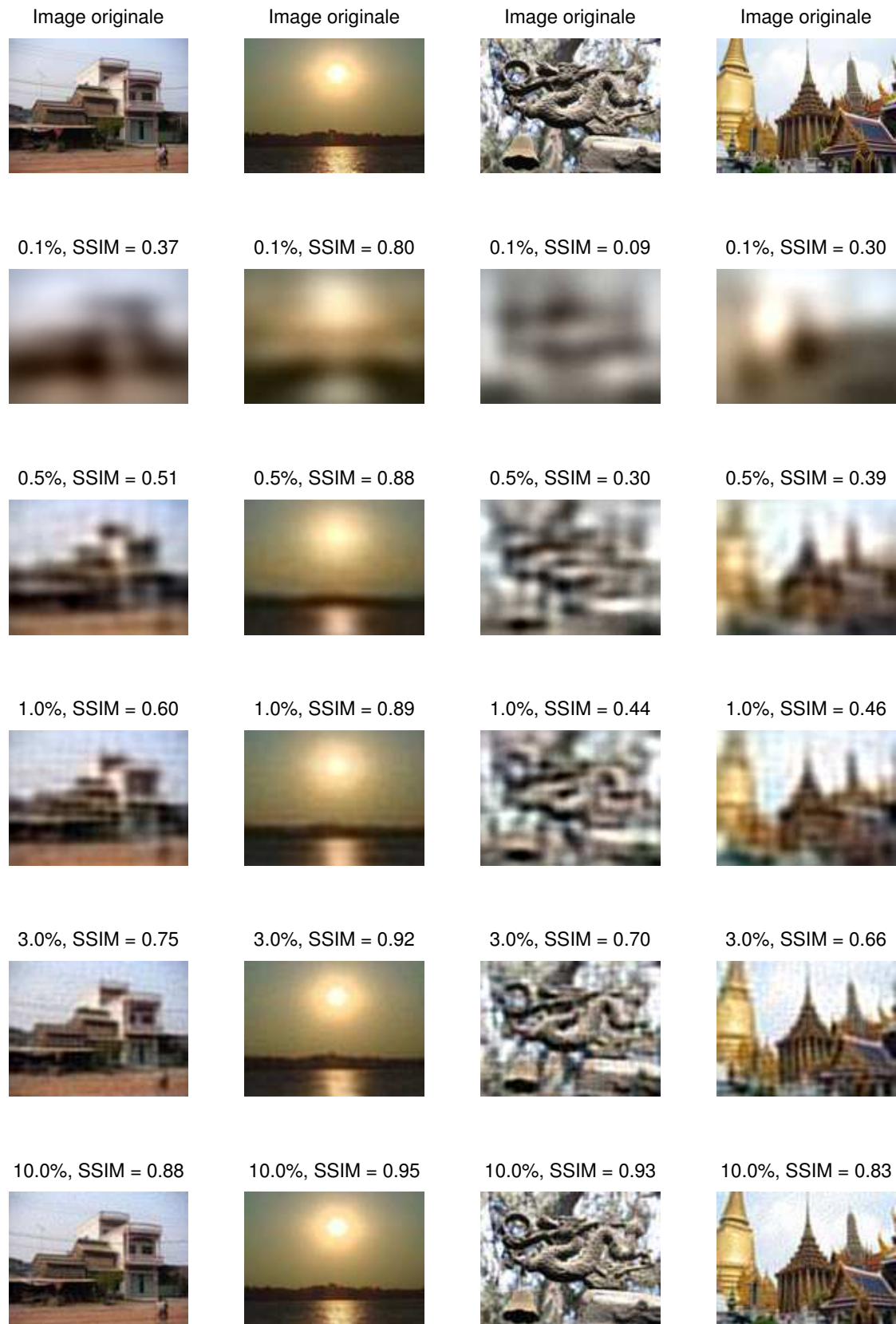


FIG. 7.18 Images reconstruites à partir d'un pourcentage des composantes principales. On peut noter que les images les mieux reconstruites sont celles avec le moins de hautes fréquences.

Conclusion

8.1 Conclusion

Cette thèse a abordé différents problèmes liés à l'information couleur dans les images numériques.

Nous avons d'abord présenté une méthode simple pour la détection automatique d'illuminants multiples dans les images numériques. Cette approche repose sur des hypothèses physiques généralement réalistes (approximation des illuminants par des corps noirs, présence de pixels achromatiques et lumineux dans les images numériques) et sur une procédure de vote après une projection dans un espace de chromaticité bien choisi. Son intérêt par rapport à l'état de l'art réside dans sa simplicité, sa capacité à détecter plusieurs illuminants simultanément présents et dans le fait qu'elle ne nécessite pas d'apprentissage. Sa capacité à bien estimer les illuminants a été démontrée sur plusieurs bases d'images.

Nous nous sommes également intéressés dans cette thèse à la pertinence de l'information colorée pour la comparaison d'images. Constatant que la couleur est souvent sous-utilisée pour l'appariement de descripteurs locaux, nous avons proposé des descripteurs couleurs à la fois robustes aux changements de balance des blancs et discriminants. Nous avons montré que l'information couleur n'était pleinement représentée qu'en tenant compte à la fois de distributions couleur locales et des distributions de gradients couleur. De nombreux tests sur des bases d'images réelles ou synthétiques ont montré un gain très significatif dans la qualité des appariements en utilisant ces descripteurs. Une méthode a contrario a été proposée pour l'appariement automatique de ces descripteurs hybrides, différents scénarios sur l'indépendance entre secteurs et/ou entre information de couleur et de luminance ayant été envisagés et testés.

La dernière partie de cette thèse correspond à un travail sur les statistiques couleur des images naturelles. Ce travail expérimental, toujours en cours, porte notamment sur une base de plusieurs millions d'images et consiste à explorer les distributions de différents attributs couleur et la stabilité de ces distributions en fonction du nombre d'images utilisées pour les estimer.

8.2 Perspectives

Les travaux développés dans ce manuscrit soulèvent de nombreuses questions.

Tout d'abord, nous envisageons de poursuivre le travail sur l'estimation des illuminants en l'étendant aux séquences d'images. Le cas des vidéos est particulièrement intéressant puisqu'il permet d'apporter plus d'information, ce qui est nécessaire à la résolution du problème de l'estimation qui est très mal posé pour les images fixes.

Concernant les descripteurs couleur, nous avons commencé à explorer la possibilité de construire des descripteurs compacts, par exemple en fusionnant l'information portée par les différents canaux Ohta sur chaque secteur du descripteur, ou en réduisant le nombre de classes utilisées dans les histogrammes. Nos premiers tests montrent des résultats étonnamment bons pour des descripteurs significativement réduits en taille. Une autre direction que nous souhaitons explorer est celle de descripteurs reposant sur des modes d'histogrammes, en s'inspirant de la méthode décrite dans l'Annexe A. Nous sommes également persuadés de l'intérêt de détecter des points clés directement dans les images couleur plutôt que sur les images en niveau de gris.

Enfin, nous comptons poursuivre nos travaux ébauchés sur les statistiques couleur dans les photographies numériques. Nous envisageons de nombreuses pistes, suggérées dans ce qui suit :

- vitesse de convergence des distributions couleurs ;
- étude du biais introduit par la compression JPEG dans les statistiques colorées ;
- représentations parcimonieuses de la couleur ;
- étude de l'influence des fonctions de sensibilité des capteurs numériques sur la définition des espaces antagonistes (obtenus par ACP sur de grosses bases d'images) ;
- statistiques sur les images multispectrales.

Détection de modes dans les histogrammes pondérés

Les histogrammes sont un outil très largement utilisé dans l'analyse de données et en traitement des images. Ils permettent une description synthétique des statistiques globales des données observées. Les régions dans lesquelles la masse de l'histogramme est abondante (modes), comme les régions dans lesquelles elle est rare (creux), sont particulièrement informatives. Détecter ces régions constitue le but de la segmentation d'histogrammes. Les algorithmes permettant de résoudre ce problème introduisent le plus souvent des *a priori* indésirables. Les algorithmes de classification non supervisée tels que l'algorithme des *k*-moyennes ou encore l'algorithme *Mean-Shift* [71] supposent la connaissance *a priori* du nombre de modes. L'estimation d'un mélange de gaussiennes avec un critère de longueur minimum de description [126] permet de surmonter ce problème mais introduit un *a priori* sur la forme du mode.

Un algorithme alternatif a été proposé par Desolneux *et coll.* [33]. Cet algorithme repose sur l'idée qu'un mode est une région en forte contradiction avec l'hypothèse que l'histogramme suit une distribution uniforme. Cette méthode a l'avantage de ne pas introduire d'*a priori* sur le nombre de modes ni sur leur forme. Néanmoins, elle utilise l'hypothèse, souvent vérifiée, que tous les éléments ayant servi à la construction de l'histogramme contribuent de manière identique. Cependant, il existe de nombreux cas dans lesquels la contribution de chaque élément est différente. Citons par exemple les histogrammes de valeurs de teinte pondérées par la saturation ou encore les histogrammes d'orientation du gradient pondérés par la magnitude. Dans cette thèse, l'histogramme de vote utilisé au Chapitre 3, et dont la construction est détaillée à la Section 3.3, est un histogramme pondéré.

Nous décrivons dans la suite une version généralisée aux histogrammes pondérés de l'algorithme proposé par Desolneux *et coll.* [33]. Nous illustrons cet algorithme en l'appliquant à la segmentation d'histogrammes de valeurs de teinte pondérées par la saturation.

A.1 Modèle de fond pour les histogrammes non pondérés

Les méthodes *a contrario* reposent sur la détection d'évènements dont la probabilité d'apparition est petite d'après un modèle de fond. Le modèle de fond définit les conditions dans lesquelles aucun événement ne doit être détecté. Dans [33], ce modèle est défini par un histogramme uniforme, c'est-à-dire obtenu à partir de points uniformément distribués, ajoutant chacun une masse unitaire à la classe à laquelle ils appartiennent. Dans ce cas,

« une masse surprenante » est une masse qui contredit notre modèle de fond. Définissons précisément ce modèle de fond.

Soit H un histogramme discret réparti en N classes (*bins*) constitué à partir de M points m_1, \dots, m_M . Pour tout m_i , on note $c_i \in \{1, \dots, N\}$ la classe qui lui correspond. L'histogramme non pondéré est alors pour tout $k \in \{1, \dots, N\}$

$$H_k = \sum_{i=1}^M \mathbb{1}_{\{c_i=k\}} \quad (\text{A.1})$$

Nous cherchons ici à détecter les modes $[a, b]$ pour lesquels la masse observée $H_{[a,b]} = \sum_{k=a}^b H_k$ est significativement supérieure à la masse attendue sous un modèle de fond prédéfini. Ces intervalles sont alors appelés modes significatifs.

Hypothèse nulle \mathcal{H}_0 Pour un intervalle $[a, b]$ associons à chaque point m_i une variable aléatoire C_i correspondant à l'évènement « la classe c_i appartient à l'intervalle $[a, b]$ ». On définit ainsi l'hypothèse \mathcal{H}_0 : les variables C_i sont *iid* et suivent une loi de Bernoulli $\mathcal{B}(p)$ avec $p = \frac{b-a+1}{N}$. Sous l'hypothèse \mathcal{H}_0 , la probabilité de mesurer dans un intervalle $[a, b]$ une masse $D_{[a,b]}$ supérieure à la masse $H_{[a,b]}$ observée est alors donnée par

$$P(D_{[a,b]} > H_{[a,b]}) = \sum_{i=k}^M \binom{M}{k} p^i (1-p)^{M-i}. \quad (\text{A.2})$$

Creux significatifs Symétriquement, il est également possible de chercher à détecter des creux significatifs. Ces creux sont des intervalles pour lesquels la masse observée est très en deçà de la masse attendue. La probabilité d'observer une masse $D_{[a,b]}$ inférieure à la masse $H_{[a,b]}$ observée est ici

$$P(D_{[a,b]} < H_{[a,b]}) = \sum_{i=k}^M \binom{M}{k} p^i (1-p)^{M-j}. \quad (\text{A.3})$$

A.2 Modèle de fond pour les histogrammes pondérés

Le modèle de fond décrit ci-dessus ne correspond pas aux histogrammes pondérés, c'est-à-dire ceux pour lesquels la contribution de chaque point est pondérée par une quantité annexe. Dans ce cas, les points ne contribuent plus nécessairement à l'histogramme avec une masse unitaire et le modèle de fond doit être redéfini.

Nous considérons maintenant que chaque point m_i vote pour une classe c_i avec un poids w_i . La valeur retenue par l'histogramme pour une classe a est alors

$$H_k = \sum_{i=1}^M w_i \mathbb{1}_{\{c_i=k\}}. \quad (\text{A.4})$$

Afin de détecter les modes significatifs dans ce type d'histogramme, il faut adapter l'hypothèse \mathcal{H}_0 . Pour cela, nous associons à chaque pixel m_i une nouvelle variable aléatoire W_i correspondant à la masse w_i . L'hypothèse \mathcal{H}_0 devient alors :

- Les W_i sont *iid* de loi μ_W ,
- les C_i sont *iid* de loi $\mathcal{B}(p)$ avec $p = \frac{b-a+1}{N}$.

— les W_i et les C_i sont indépendantes entre elles.

Considérons les variables $X_i = W_i C_i$, la distribution associée à chacune d'entre elles est alors

$$\mu_{X_i} = (1-p)\delta_0 + p\mu_W. \quad (\text{A.5})$$

Ces variables étant indépendantes par hypothèse, la distribution de leur somme est donnée par le produit de convolution suivant

$$\mu_{(\sum_{i=1}^M X_i)} = \star_{i=1}^M \mu_X = \sum_{k=0}^M \binom{M}{k} (1-p)^k p^{M-k} (\star_{i=1}^{M-k} \mu_W) \star \delta_0. \quad (\text{A.6})$$

Le but est de calculer $P[\sum_{i=1}^M X_i > H_{[a,b]}]$. Nous considérons dans ce qui suit le cas où les variables aléatoires W_i suivent une loi exponentielle.

Loi exponentielle En supposant que μ_W est une distribution exponentielle de paramètre λ , alors la somme de ℓ variables aléatoires indépendantes de ce type suit une distribution d'Erlang μ_E de paramètres (ℓ, λ) dont la densité est

$$\mu_E(x; \ell, \lambda) = \lambda^\ell x^{\ell-1} \frac{e^{-\lambda x}}{(\ell-1)!}.$$

La fonction de répartition de la loi d'Erlang nous permet d'écrire que

$$\int_d^{+\infty} d\mu_E(x; \ell, \lambda) = e^{-\lambda d} \sum_{j=0}^{\ell-1} \frac{(\lambda d)^j}{j!}.$$

Dans notre cas, en considérant $\ell = M - k$ et $d = H_{[a,b]}$, On obtient que

$$P \left[\sum_{i=1}^{M-k} W_i > H_{[a,b]} \right] = e^{-\lambda H_{[a,b]}} \sum_{j=0}^{M-k-1} \frac{(\lambda H_{[a,b]})^j}{j!}. \quad (\text{A.7})$$

Ainsi

$$P \left[\sum_{i=1}^M X_i > H_{[a,b]} \right] = \sum_{k=0}^{M-1} \binom{M-1}{k} (1-p)^k p^{M-k} P \left[\sum_{i=1}^M W_i > H_{[a,b]} \right] \quad (\text{A.8})$$

$$= \sum_{k=0}^{M-1} \binom{M-1}{k} (1-p)^k p^{M-k} e^{-\lambda H_{[a,b]}} \sum_{j=0}^{M-k-1} \frac{(\lambda H_{[a,b]})^j}{j!}. \quad (\text{A.9})$$

$$(\text{A.10})$$

Remarquons que la somme va de 0 à $M - 1$ et non jusqu'à M puisque la distribution δ_0 n'a pas de masse entre $H_{[a,b]}$ et $+\infty$.

$$P \left[\sum_{i=1}^M X_i > H_{[a,b]} \right] = e^{-\lambda H_{[a,b]}} \sum_{j=0}^{M-k-1} \frac{(\lambda H_{[a,b]})^j}{j!} (1 - B(M, M - j, 1 - p)),$$

où B est la queue de la binomiale.

Application du théorème central limite Lorsque le nombre de points est suffisamment grand (typiquement quelques centaines), le théorème central limite peut être appliqué pour simplifier le calcul de cette probabilité. La moyenne et la variance de chacune des variables X_i est alors

$$E[X_i] = p\lambda^{-1}, \quad \text{Var}[X_i] = p\lambda^{-2}(2-p). \quad (\text{A.11})$$

Le théorème central limite prédit que la somme des X_i suit une loi normale de paramètres $\mathcal{N}(Mp\lambda^{-1}, Mp\lambda^{-2}(2-p))$. En conséquence, la probabilité que la somme des X_i soit supérieure à la masse $H_{[a,b]}$ est

$$P \left[\sum_{i=1}^M X_i > H_{[a,b]} \right] = P \left[\frac{\sum_{i=1}^M X_i - p\lambda^{-1}}{p\lambda^{-2}(2-p)} \sqrt{M} > H_{[a,b]} \right] \simeq \int_{H_{[a,b]}}^{+\infty} \frac{e^{-\frac{t^2}{2}}}{\sqrt{2\pi}} dt. \quad (\text{A.12})$$

A.3 Significativité d'un intervalle et intervalles maximum significatifs

Le nombre de fausses alarmes est défini comme l'espérance du nombre d'occurrences de l'évènement. Dans notre cas, il correspond à la réponse à la question : « Combien d'intervalles ayant une masse supérieure à celle observée dans l'intervalle $[a, b]$ vais-je observer en moyenne si je teste N_{test} intervalles contre \mathcal{H}_0 ? ». Ce nombre est défini par

$$\text{NFA}_{[a,b]} = N_{\text{test}} P \left[\sum_{i=0}^M X_i > H_{[a,b]} \right]. \quad (\text{A.13})$$

Un intervalle $H_{[a,b]}$ est dit ε -significatif si $\text{NFA}_{[a,b]} < \varepsilon$ c'est-à-dire si

$$P \left[\sum_{i=0}^M X_i > H_{[a,b]} \right] < \frac{\varepsilon}{N_{\text{test}}}. \quad (\text{A.14})$$

Le nombre de tests dans un histogramme est $N_{\text{test}} = \frac{L(L+1)}{2}$. Dans le cadre d'un histogramme circulaire, comme pour un histogramme d'orientation du gradient ou de teinte, le nombre de test est alors $N_{\text{test}} = L(L+1)$.

Intervalles maximum significatifs Afin de limiter les détections d'intervalles significatifs multiples se recouvrant, il a été proposé dans [33] d'appliquer un principe de maximalité aux détections. Un intervalle maximal significatif est défini comme étant un intervalle qui :

- ne contient pas d'intervalle plus significatif que lui,
- n'est pas contenu dans un intervalle plus significatif.

De plus un mode maximal significatif est un intervalle maximal significatif qui ne contient pas de creux significatif.

A.4 Illustrations de la détection de modes dans les histogrammes pondérés

L'algorithme présenté précédemment, est utilisé au Chapitre 3 afin de segmenter un histogramme de vote. Il est également utilisé au paragraphe 4.3.2 pour détecter l'orientation principale des descripteurs locaux. Nous appliquons ici à titre d'exemple la détection de modes à des histogrammes de valeurs de teinte pondérées par la saturation. Une application éventuelle est le choix automatique d'une palette de couleur pour une image [29].

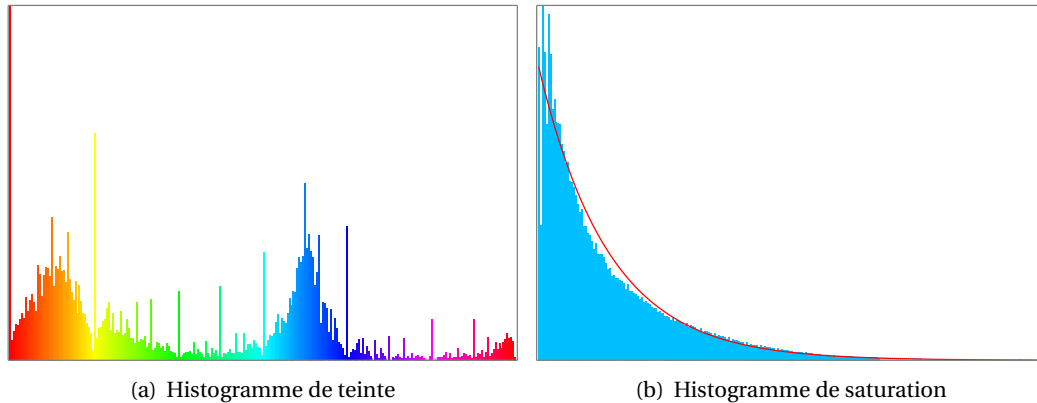


FIG. A.1 Histogrammes calculés sur 10000 images réduites à une résolution de 256×192 pixels (cf. Chapitre 7). Sur l’histogramme de saturation, une distribution exponentielle de paramètre $\lambda = 0.14$ est tracée en rouge.

A.4.1 Segmentation d’histogrammes de valeurs de teinte pondérées

Vérification du modèle de fond Au Chapitre 7, afin de vérifier expérimentalement la validité des hypothèses faites, nous avons calculé les histogrammes de teinte et de saturation à partir de 10000 images d’une résolution de 256×192 pixels. La Figure A.1 présente ces histogrammes. Les exemples de détection sont donnés par les figure A.2 et A.3.

Histogramme de teinte De forts effets de quantification sont visibles sur l’histogramme de teinte. Le pic en zéro est en partie dû aux pixels ayant une saturation nulle, car dans ce cas, la teinte qui leur est attribuée est également nulle. L’histogramme présente deux modes : un dans le rouge/jaune ainsi qu’un dans le bleu. Il n’est donc pas uniforme. Néanmoins utiliser cette distribution comme *a priori* pour le modèle de fond conduirait à rendre plus difficile la détection de modes de couleur bleu ou rouge.

Histogramme de saturation L’histogramme de saturation permet de juger la justesse de l’approximation par une loi exponentielle. Cette dernière n’est pas parfaite mais elle est néanmoins satisfaisante pour l’application suivante. Le paramètre λ de la loi exponentielle est appris empiriquement comme l’inverse de la moyenne des saturations mesurées.

A.4.2 Application à la détection d’orientations principales

La Figure A.4 illustre la détection d’orientations principales par détection de modes dans les histogrammes de valeurs d’orientations du gradient pondérées par la norme du gradient.

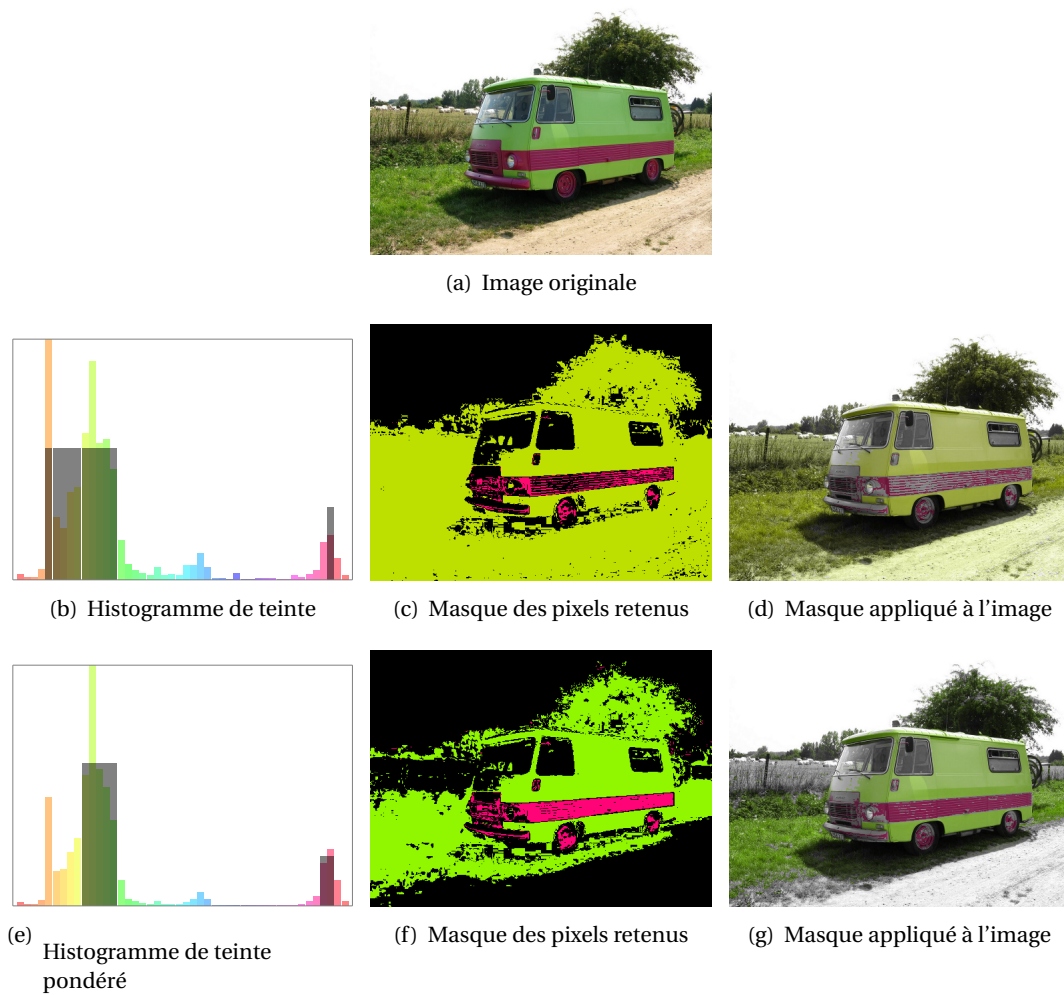


FIG. A.2 Exemples de détection de mode de teinte sur une image. La Figure A.2(a) donne l'image de départ, la Figure A.2(b) l'histogramme de teinte ainsi que les modes détectés par l'algorithme de [33]. L'histogramme de teinte pondéré et les modes associés sont donnés par la Figure A.2(e).

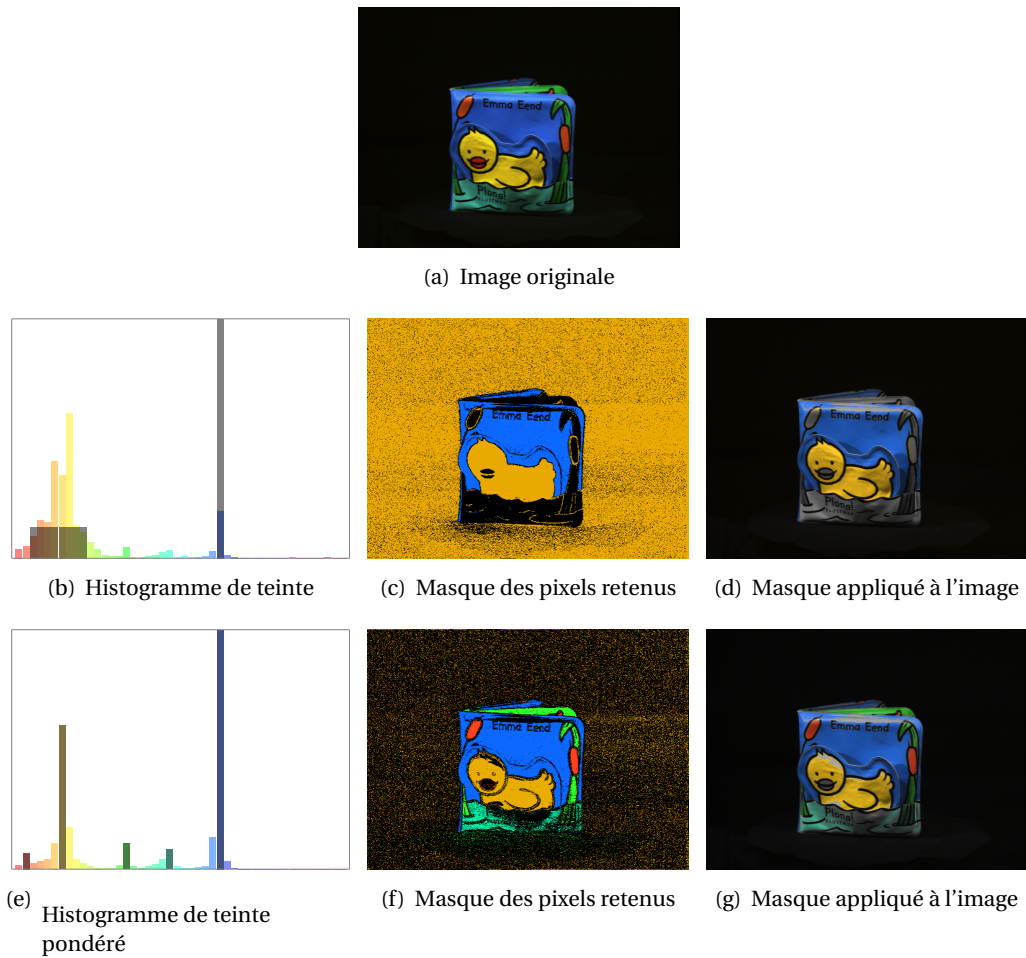


FIG. A.3 Exemples de détection de modes de teinte sur une image de la base ALOI [56]. La Figure A.3(a) donne l'image de départ, la Figure A.3(b) l'histogramme de teinte ainsi que les modes détectés par l'algorithme de [33]. L'histogramme de teinte pondéré et les modes associés sont donnés par la Figure A.3(e).

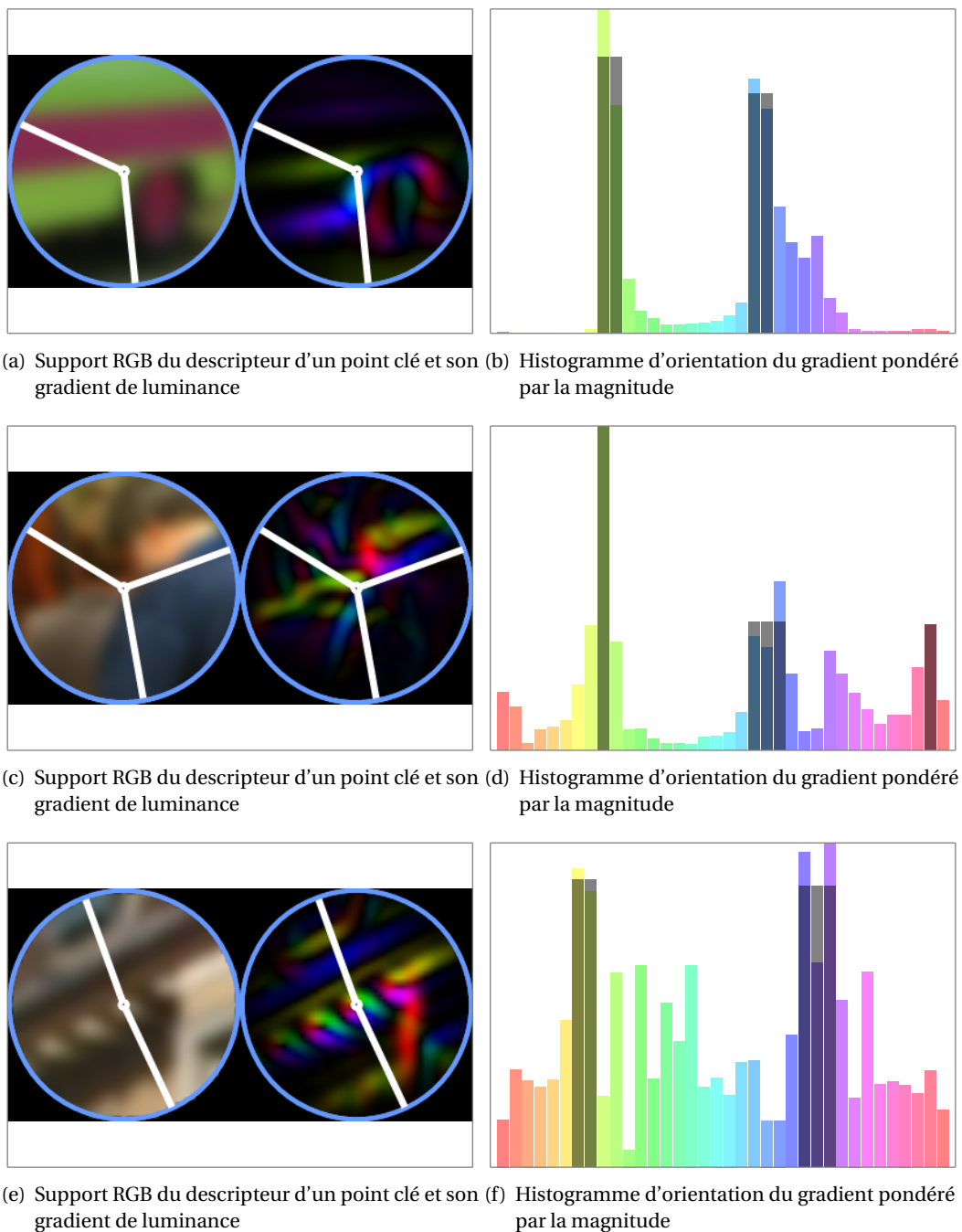


FIG. A.4 Exemples de sélection d'orientations principales pour des points clés par segmentation d'un histogramme d'orientation du gradient. Sur le patch du gradient et sur l'histogramme, la couleur correspond à la phase du gradient. Les orientations sont obtenues en calculant le barycentre de l'intervalle associé à chaque mode.

Modèle à deux paramètres pour le changement d'illuminant

Dans cette annexe, nous présentons un modèle de changement d'illuminant. Ce modèle peut être considéré comme un cas particulier du modèle diagonal de von Kries [151]. Celui-ci (cf. chapitre 2.1) décrit un changement d'illuminant par un facteur multiplicatif appliqué à chacun des canaux RGB. Cette approche physique a l'inconvénient de mélanger changement d'intensité et changement dans la distribution spectrale de l'illuminant. Le modèle que nous proposons n'utilise que deux paramètres. Le premier décrit un changement d'intensité et le second décrit une modification du contenu spectral de l'illuminant, telle une balance rouge/bleu.

Un modèle très proche est utilisé par Gevers et Smeulders [60], sans précision sur son origine. Nous montrons le lien entre ce modèle et l'équation de Planck et évaluons sa crédibilité sur la base d'images ALOI [56].

B.1 Simplification du modèle de Von Kries pour l'adaptation chromatique

Rappelons que pour une même surface \mathbf{s} éclairée sous deux illuminants différents \mathbf{l}^1 et \mathbf{l}^2 , le modèle de von Kries établit la relation suivante entre les mesures \mathbf{p}^1 et \mathbf{p}^2 effectuées dans l'espace RGB de l'appareil photo

$$\begin{bmatrix} p_R^1 \\ p_G^1 \\ p_B^1 \end{bmatrix} = \begin{bmatrix} l_R^1/l_R^2 & 0 & 0 \\ 0 & l_G^1/l_G^2 & 0 \\ 0 & 0 & l_B^1/l_B^2 \end{bmatrix} \begin{bmatrix} p_R^2 \\ p_G^2 \\ p_B^2 \end{bmatrix}. \quad (\text{B.1})$$

Ce modèle est général et n'utilise pas *a priori* sur la distribution des illuminants. Si l'on suppose que les illuminants ont généralement une couleur proche de celles prédites par l'équation de Planck 2.20, il est possible de simplifier ce modèle. En effet, comme cela est visible sur la Figure B.1, le lieu de Planck dans l'espace de chromaticité rg peut être grossièrement approché par une droite obtenue par une régression linéaire sur une sous partie du lieu de Planck ([2300K, 22000K]). Cette approximation du lieu de Planck peut elle-même

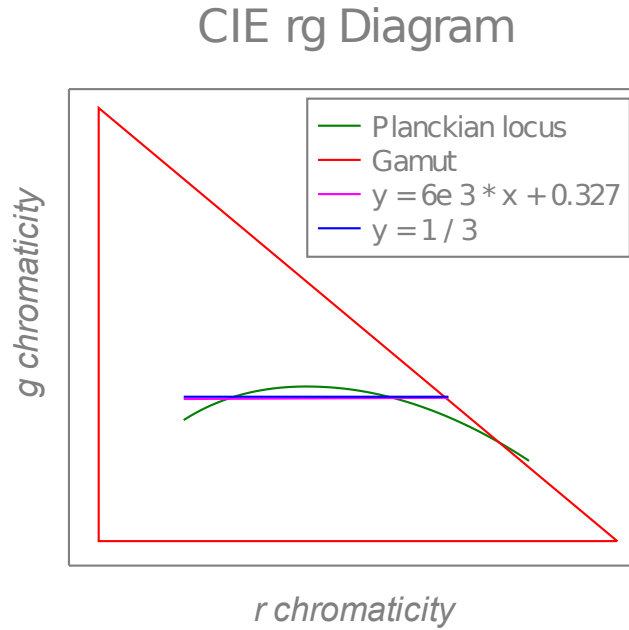


FIG. B.1 Régression linéaire sur le lieu de Planck pour des températures comprises entre 2300K et 22000K (droite Fuchsia). La droite bleue correspond au modèle à deux paramètres ainsi qu'à l'équation $y = \frac{1}{3}$.

être simplifiée par la droite d'équation $g = \frac{1}{3}$. En paramétrant $r = \frac{1}{3}(1 - d)$ et sachant que $r + g + b = 1$, il est alors possible d'écrire :

$$\begin{cases} r = \frac{1}{3}(1 - d) \\ g = \frac{1}{3} \\ b = 1 - r - g = \frac{1}{3}(1 + d) \end{cases}, \quad (\text{B.2})$$

avec d un paramètre. En posant que $R + G + B = 3k$, et sachant que

$$R = r(R + G + B), \quad G = g(R + G + B), \quad B = b(R + G + B) \quad (\text{B.3})$$

On obtient le modèle à deux paramètres suivant :

$$\begin{bmatrix} p_R^1 \\ p_G^1 \\ p_B^1 \end{bmatrix} = \begin{bmatrix} k(1 - d) & 0 & 0 \\ 0 & k & 0 \\ 0 & 0 & k(1 + d) \end{bmatrix} \begin{bmatrix} p_R^2 \\ p_G^2 \\ p_B^2 \end{bmatrix}. \quad (\text{B.4})$$

Ce modèle simplifié a l'avantage de ne comporter que 2 paramètres : le paramètre d permet de régler la balance rouge/bleu tandis que le paramètre k règle le changement d'intensité de l'illuminant.

Contrairement à l'équation de Planck, le changement d'illuminant se fait le long d'une droite et non plus suivant une courbe dans le diagramme de chromaticité. En particulier, et ceci peut être très utile en pratique, il n'est plus nécessaire de connaître l'illuminant de départ pour déterminer l'illuminant d'arrivée. Il est donc possible en utilisant ce modèle de parcourir en aveugle le lieu de planck.

B.2 Observation des changements d'illumination de la base ALOI

Nous proposons de tester la qualité du modèle sur la base ALOI [56]. Cette base contient 1000 objets photographiés dans différentes conditions : changement de point de vue, changement de la disposition de l'éclairage, et changement de la température de couleur de l'illuminant. C'est ce dernier type de changement qui nous intéresse. Dans cette base, les changements d'illuminant sont simulés à l'aide d'un ensemble de 5 lampes à incandescence tungstène Osram (type 64637, 12 V, 100 W), possédant une température de couleur de 3100 K à 12 V. Lors des prises de vue, la balance des blancs de l'appareil photographique est calibrée à une température de couleur de 3075 K. On observe donc une variation de l'illumination de la lumière blanche vers une teinte rouge. Les images correspondant à l'identifiant $i250$ sont considérées comme images de référence. Dans le but d'obtenir des changements dans le spectre d'illumination, une tension variable est appliquée aux 5 lampes. Les différentes conditions expérimentales testées sont résumées par la Table B.1 et un exemple d'images capturées pour un objet est donné sur la Figure B.2.

id.	i110	i120	i130	i140	i150	i160
V	5.17	5.64	6.11	6.58	7.05	7.52
T(K)	2175	2250	2325	2400	2475	2550
id.	i170	i180	i190	i210	i230	i250
V	8.00	8.47	8.94	9.88	10.82	11.76
T(K)	2625	2675	2750	2850	2975	3075

TAB. B.1 Équivalences identifiant/tension/température



FIG. B.2 Exemple d'un objet de la base ALOI et des modifications colorimétriques appliquées (de $i110$ à $i250$).

Afin de vérifier la qualité du modèle décrit précédemment pour la modélisation du changement de température de couleur, nous proposons de le comparer au modèle de von Kries. Pour cela, une estimation des paramètres des deux modèles et des erreurs qu'ils induisent a été réalisée sur la base ALOI. Les paramètres de la transformation sont estimés par moindres carrés.

Estimation des paramètres du modèle de Von Kries Dans la base ALOI Pour vérifier la validité du modèle de Von Kries sur les changements d'illuminant de la base ALOI, nous avons estimé pour chaque illuminant les paramètres α, β, γ moyens de la matrice diagonale de

l'équation 2.26 et observé l'erreur qu'ils engendrent lorsqu'on les utilise pour retrouver l'image de référence.

On note \mathbf{p}^{250} l'ensemble des n pixels de toutes les images éclairées sous l'illuminant de référence i_{250} et \mathbf{p}^v l'ensemble des n pixels de toutes les images éclairées sous l'illuminant $v \in \{110, \dots, 230\}$. On estime alors par moindres carrés les paramètres α , β et γ de la matrice M tels que :

$$\mathbf{p}_v = M\mathbf{p}_{250} \quad (\text{B.5})$$

$$\begin{bmatrix} R_1^{250} & \dots & R_n^{250} \\ G_1^{250} & \dots & G_n^{250} \\ B_1^{250} & \dots & G_n^{250} \end{bmatrix} = \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{bmatrix} \begin{bmatrix} R_1^v & \dots & R_n^v \\ G_1^v & \dots & G_n^v \\ B_1^v & \dots & G_n^v \end{bmatrix}.$$

la Figure B.3 donne les résultats de cette estimation ainsi que les erreurs causées par le modèle diagonal. On peut noter que le modèle fournit des prévisions étranges pour l'illuminant i_{140} .

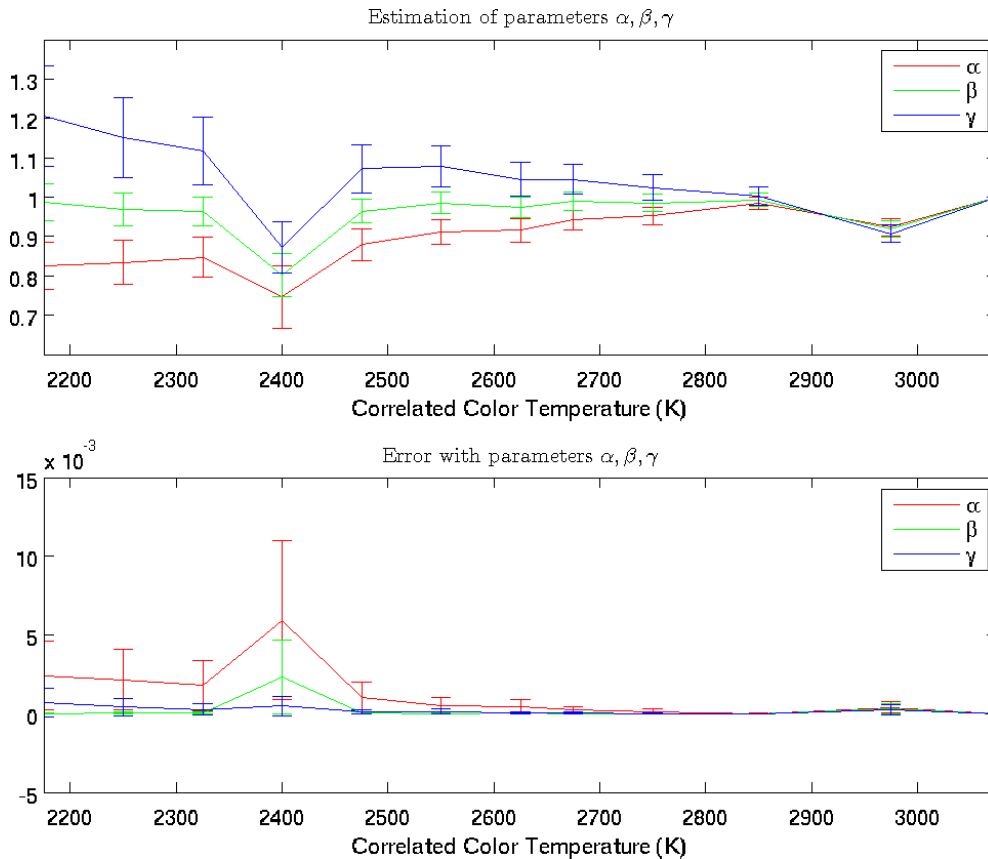


FIG. B.3 Estimations des paramètres du modèle de Von Kries sur la base ALOI et erreurs liées.

Cela semble être une erreur dans la base car un saut d'intensité est observable sur toutes les images pour cette configuration d'illumination.

Conclusion Comme le montre les erreurs obtenues (cf. Figure B.4), le modèle proposé constitue une approximation raisonnable des changements colorimétriques observés sur la base ALOI. Nous avons montré que ce modèle constitue une approximation simple du lieu de Planck. En outre, ce modèle possède deux avantages. Il permet d'appliquer, directement dans l'espace RGB, des changements colorimétriques similaires à ceux provoqués par un changement d'illuminant. De plus, ces changements ne sont contrôlés que par deux paramètres : l'un contrôlant l'intensité et l'autre les variations couleur. Ainsi, le modèle proposé, contrairement à un changement de température de couleur, ne nécessite pas la connaissance *a priori* de la température de couleur de l'illuminant d'entrée.

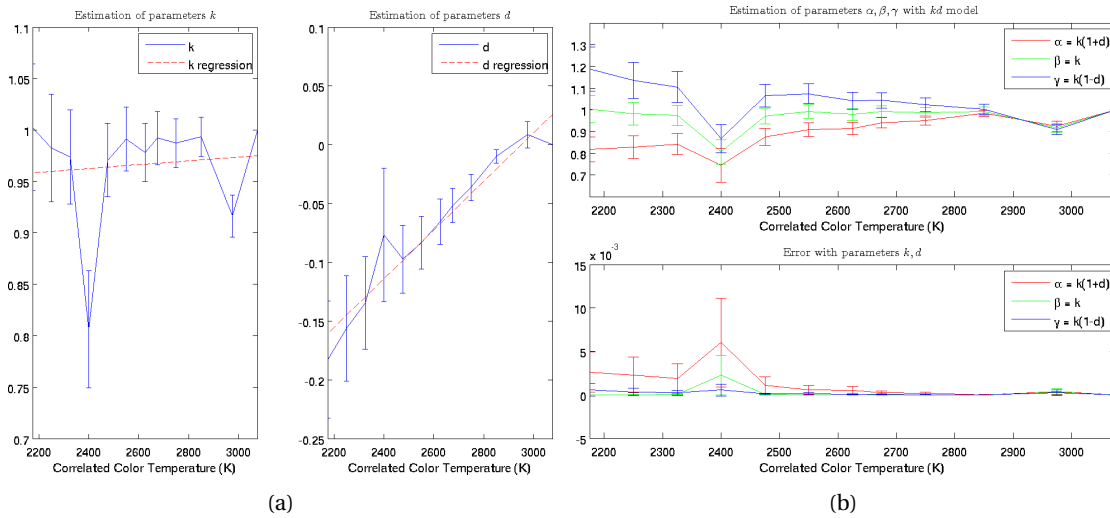


FIG. B.4 Estimations des paramètres et erreurs sur la base ALOI avec le modèle proposé.



Bibliographie

- [1] Alaa E. Abdel-Hakim and Aly A. Farag. CSIFT : A SIFT descriptor with color invariant characteristics. In *Conference on Computer Vision and Pattern Recognition*, pages 1978–1983, 2006.
- [2] Fred Attneave. Some informational aspects of visual perception. *Psychological review*, 61(3) :183, 1954.
- [3] HB Barlow. The coding of sensory messages. *Current problems in animal behaviour*, pages 331–360, 1961.
- [4] Kobus Barnard, Vlad Cardei, and Brian Funt. A comparison of computational color constancy algorithms. i : Methodology and experiments with synthesized data. *Image Processing, IEEE Transactions on*, 11(9) :972–984, 2002.
- [5] Kobus Barnard, Graham Finlayson, and Brian Funt. Color constancy for scenes with varying illumination. *Computer vision and image understanding*, 65(2) :311–321, 1997.
- [6] Kobus Barnard, Lindsay Martin, Adam Coath, and Brian Funt. A comparison of computational color constancy algorithms. *IEEE Trans. Image Process.*, 11 :2002, 2002.
- [7] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf : Speeded up robust features. In *ECCV*, pages 404–417, 2006.
- [8] Anthony J Bell and Terrence J Sejnowski. The “independent components” of natural scenes are edge filters. *Vision research*, 37(23) :3327–3338, 1997.
- [9] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4) :509–522, 2002.
- [10] Simone Bianco, Gianluigi Ciocca, Claudio Cusano, and Raimondo Schettini. Improving color constancy using indoor–outdoor image classification. *Image Processing, IEEE Transactions on*, 17(12) :2381–2392, 2008.
- [11] Michael Bleier, Christian Riess, Shida Beigpour, Eva Eibenberger, Elli Angelopoulou, Tobias Tröger, and André Kaup. Color constancy and non-uniform illumination : Can existing algorithms work? In *ICCV Workshops*, 2011.
- [12] Michael Bleyer and Sylvie Chambon. Does color really help in dense stereo matching? In *International Symposium 3D Data Processing, Visualization and Transmission (3DPVT)*

- 2010, pages 1–8, 2010. Vortrag : International Symposium 3D Data Processing, Visualization and Transmission (3DPVT) 2010, Paris, France ; 2010-05-17 – 2010-05-20.
- [13] Anna Bosch, Andrew Zisserman, and Xavier Muoz. Scene classification using a hybrid generative/discriminative approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(4) :712–727, 2008.
- [14] Ivaylo Boyadzhiev, Kavita Bala, Sylvain Paris, and Frédo Durand. User-guided white balance for mixed lighting conditions. *ACM Trans. Graph.*, 31(6) :200 :1–200 :10, November 2012.
- [15] Matthew Brown and S Susstrunk. Multi-spectral sift for scene category recognition. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 177–184. IEEE, 2011.
- [16] G Buchsbaum. A spatial processor model for object colour perception. *J. Franklin Inst.*, 310(1) :1–26, 1980.
- [17] Gershon Buchsbaum and A Gottschalk. Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the Royal society of London. Series B. Biological sciences*, 220(1218) :89–113, 1983.
- [18] G. J. Burghouts and J. M. Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113 :48–62, 2009.
- [19] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief : Binary robust independent elementary features. In *Computer Vision–ECCV 2010*, pages 778–792. Springer, 2010.
- [20] Ayan Chakrabarti, Keigo Hirakawa, and Todd Zickler. Color constancy with spatio-spectral statistics. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 34(8) :1509–1519, 2012.
- [21] Sylvie Chambon and Alain Crouzil. Colour correlation-based matching. *International Journal of Robotics and Automation*, 20(2) :78–85, 2005.
- [22] Peng Chang and John Krumm. Object recognition with color cooccurrence histograms. In *CVPR*, pages 2498–2504, 1999.
- [23] Yizong Cheng. Mean shift, mode seeking, and clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(8) :790–799, 1995.
- [24] Dung Manh Chu and Arnold WM Smeulders. Color invariant surf in discriminative object tracking. In *Trends and Topics in Computer Vision*, pages 62–75. Springer, 2012.
- [25] Florian Ciurea and Brian Funt. A large image database for color constancy research, 2003.
- [26] Dave Coffin. Decoding raw digital photos in linux, 2013.
- [27] Robert T Collins, Yanxi Liu, and Marius Leordeanu. Online selection of discriminative tracking features. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(10) :1631–1643, 2005.
- [28] A.B. Dahl and H. Aanaes. Effective image database search via dimensionality reduction. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on*, pages 1–6, 2008.

- [29] Julie Delon, Agnes Desolneux, Jose Luis Lisani, and Ana Belen Petro. Automatic color palette. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–706. IEEE, 2005.
- [30] NG Deriugin. The power spectrum and the correlation function of the television signal. *Telecommunications*, 1(7) :1–12, 1956.
- [31] Joel B Derrico and Gershon Buchsbaum. A computational model of spatiochromatic image coding in early vision. *Journal of Visual Communication and Image Representation*, 2(1) :31–38, 1991.
- [32] A. Desolneux, L. Moisan, and J.-M. Morel. A grouping principle and four applications. *PAMI*, 25(4) :508–513, 2003.
- [33] A. Desolneux, L. Moisan, and J. M. Morel. *From Gestalt Theory to Image Analysis*, volume 34. Springer-Verlag, 2008.
- [34] Agnès Desolneux, Lionel Moisan, and Jean-Michel Morel. Edge detection by helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3) :271–284, 2001.
- [35] Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer Vision, Graphics, and Image Processing*, 33(1) :116–125, 1986.
- [36] Marc Ebner. A parallel algorithm for color constancy. *J. Parallel Distrib. Comput.*, 64(1) :79–88, January 2004.
- [37] Marc Ebner. *Color constancy*, volume 6. Wiley, 2007.
- [38] Marc Ebner. Estimating the color of the illuminant using anisotropic diffusion. In Walter G. Kropatsch, Martin Kampel, and Allan Hanbury, editors, *Computer Analysis of Images and Patterns, 12th International Conference, CAIP 2007, Vienna, Austria, August 27-29, 2007, Proceedings*, volume 4673 of *Lecture Notes in Computer Science*, pages 441–449. Springer, 2007.
- [39] David J Field. Relations between the statistics of natural images and the response properties of cortical cells. *JOSA A*, 4(12) :2379–2394, 1987.
- [40] G. D. Finlayson and S. D. Hordley. Gamut constrained illuminant estimation. *Int. J. Comput. Vision*, 67 :2006, 2006.
- [41] Graham Finlayson, Steven Hordley, Gerald Schaefer, and Gui Yun Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern recognition*, 38(2) :179–190, 2005.
- [42] Graham D. Finlayson, Mark S. Drew, and Brian V. Funt. Color constancy : Generalized diagonal transforms suffice. *J. Opt. Soc. Am. A*, 11 :3011–3020, 1994.
- [43] Graham D Finlayson, Mark S Drew, and Brian V Funt. Spectral sharpening : sensor transformations for improved color constancy. *JOSA A*, 11(5) :1553–1563, 1994.
- [44] Graham D. Finlayson, Brian V. Funt, and Kobus Barnard. Color constancy under varying illumination. In *ICCV*, pages 720–725, 1995.
- [45] Graham D. Finlayson and Steven D. Hordley. Color constancy at a pixel. *J. Opt. Soc. Am. A*, 18(2) :253–264, Feb 2001.

- [46] Graham D Finlayson and Steven D Hordley. Color constancy at a pixel. *JOSA A*, 18(2) :253–264, 2001.
- [47] Graham D. Finlayson, Steven D. Hordley, and Paul M. Hubel. Color by correlation : A simple, unifying framework for color constancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11) :1209–1221, 2001.
- [48] Graham D. Finlayson and Gerald Schaefer. Solving for colour constancy using a constrained dichromatic reflection model. *Int. J. Comput. Vision*, 42(3) :127–144, May 2001.
- [49] Graham D Finlayson, Bernt Schiele, and James L Crowley. Comprehensive colour image normalization. In *Computer Vision—ECCV’98*, pages 475–490. Springer, 1998.
- [50] Graham D. Finlayson and Elisabetta Trezzi. Shades of gray and colour constancy. In *Color Imaging Conference*, pages 37–41, 2004.
- [51] D. A. Forsyth. A novel algorithm for color constancy. *Int. J. Comput. Vision*, 5(1) :5–36, 1990.
- [52] Brian V. Funt and Graham D. Finlayson. Color constant color indexing. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 17(5) :522–529, 1995.
- [53] P. V. Gehler, C. Rother, A. Blake, T. Minka, and T. Sharp. Bayesian color constancy revisited. In *Conference on Computer Vision and Pattern Recognition*, pages 1–8, 06 2008.
- [54] T Geodeme, Tinne Tuytelaars, G Vanacker, M Nuttin, and Luc Van Gool. Omnidirectional sparse visual path following with occlusion-robust feature tracking. In *OMNIVIS Workshop, ICCV*, 2005.
- [55] Ron Gershon, Allan D Jepson, and John K Tsotsos. From [r, g, b] to surface reflectance : Computing color constant descriptors in images. In *IJCAI*, pages 755–758, 1987.
- [56] J. M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1) :103–112, 2005.
- [57] J. M. Geusebroek, T. Gevers, and A. W. M. Smeulders. The Kubelka-Munk theory for color image invariant properties. In *European Conference on Colour in Graphics*, pages 463–467, 2002.
- [58] J. M. Geusebroek, R. van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(12) :1338–1350, 2001.
- [59] Jan-Mark Geusebroek, Rein Van Den Boomgaard, Arnold WM Smeulders, and Anuj Dev. Color and scale : The spatial structure of color images. In *Computer Vision-ECCV 2000*, volume 1, pages 331–341. Springer, 2000.
- [60] T. Gevers and WM Smeulders. Color based object recognition. *Pattern recognition*, 32(3) :453–464, 1999.
- [61] Theo Gevers and Arnold WM Smeulders. Content-based image retrieval by viewpoint-invariant color indexing. *Image and vision computing*, 17(7) :475–488, 1999.
- [62] Theo Gevers and Harro Stokman. Robust histogram construction from color invariants for object recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(10) :113–118, 2003.
- [63] A. Gijsenij, T. Gevers, and J. van de Weijer. Generalized gamut mapping using image derivative structures for color constancy. *Int. J. Comput. Vision*, 86(2-3) :127–139, 2010.

- [64] A. Gijsenij, T. Gevers, and J. van de Weijer. Computational color constancy : Survey and experiments. *IEEE Trans. Image Process.*, 20(9) :2475–2489, 2011.
- [65] A. Gijsenij, Th. Gevers, and M.P Lucassen. A perceptual analysis of distance measures for color constancy. *Journal of the Optical Society of America A*, 26(10) :2243–2256, 2009.
- [66] A. Gijsenij, Th. Gevers, and J. van de Weijer. Physics-based edge evaluation for improved color constancy. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2009 (CVPR'09)*, pages 1–8, Miami Beach, Florida, USA, June 2009.
- [67] A. Gijsenij, Rui Lu, and T. Gevers. Color constancy for multiple light sources. *Image Processing, IEEE Transactions on*, 21(2) :697–707, feb. 2012.
- [68] Arjan Gijsenij and Theo Gevers. Color constancy using natural image statistics and scene semantics. *IEEE Trans. Pattern Anal. Mach. Intell.*, 33(4) :687–698, 2011.
- [69] Allan Hanbury. Constructing cylindrical coordinate colour spaces. *Pattern Recogn. Lett.*, 29 :494–500, March 2008.
- [70] Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK, 1988.
- [71] Trevor Hastie, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.
- [72] James Hays and Alexei A Efros. Scene completion using millions of photographs. *Communications of the ACM*, 51(10) :87–94, 2008.
- [73] Glenn Healey and David Slater. Global color constancy : recognition of objects by use of illumination-invariant properties of color distributions. *JOSA A*, 11(11) :3003–3010, 1994.
- [74] Javier Hernández-Andrés, Raymond L. Lee, and Javier Romero. Calculating correlated color temperatures across the entire gamut of daylight and skylight chromaticities. *Appl. Opt.*, 38(27) :5703–5709, Sep 1999.
- [75] Heiko Hirschmuller and Daniel Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(9) :1582–1599, 2009.
- [76] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédo Durand. Light mixture estimation for spatially varying white balance. *ACM Trans. Graph.*, 27(3) :70 :1–70 :7, August 2008.
- [77] Jinggang Huang and David Mumford. Statistics of natural images and models. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1. IEEE, 1999.
- [78] Aapo Hyvärinen, Jarmo Hurri, and Patrick O. Hoyer. *Natural Image Statistics*, volume Vol. 39 of *Computational Imaging and Vision*. Springer, 2010.
- [79] Yoshie Imai, Yu Kato, Hideki Kadoi, Takahiko Horiuchi, and Shoji Tominaga. Estimation of multiple illuminants based on specular highlight detection. In *Proceedings of the Third international conference on Computational color imaging, CCIW'11*, pages 85–98, Berlin, Heidelberg, 2011. Springer-Verlag.

- [80] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. Aggregating local descriptors into a compact image representation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3304–3311. IEEE, 2010.
- [81] Jun Jiang, Dengyu Liu, Jinwei Gu, and Sabine Süsstrunk. What is the space of spectral sensitivity functions for digital color cameras? In *WACV*, pages 168–179, 2013.
- [82] Deane B. Judd, David L. Macadam, Günter Wyszecki, H. W. Budde, H. R. Condit, S. T. Henderson, and J. L. Simonds. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.*, 54(8) :1031–1036, Aug 1964.
- [83] Deane Brewster Judd. *Color in business, science, and industry*. Wiley, N.Y. :, 1952.
- [84] Bongsoon Kang, Ohak Moon, Changhee Hong, Honam Lee, Bonghwan Cho, and Youngsun Kim. Design of advanced color - temperature control system for hdtv applications. *Journal of the Korean Physical Society*, Vol. 41(No. 6) :pp. 865–871, December 2002.
- [85] Rei Kawakami, Robby T Tan, and Katsushi Ikeuchi. A robust framework to estimate surface color from changing illumination. In *Asian Conference on Computer Vision (ACCV2004)*, 2004.
- [86] Yan Ke and Rahul Sukthankar. Pca-sift : a more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition, CVPR'04*, pages 506–513, 2004.
- [87] James M Kraft and David H Brainard. Mechanisms of color constancy under nearly natural viewing. *Proceedings of the National Academy of Sciences*, 96(1) :307–312, 1999.
- [88] Hyuk-Ju Kwon, Sung-Hak Lee, Tae-Wuk Bae, and Kyu-Ik Sohng. Compensation of desaturation effect in {HDR} imaging using a real scene adaptation model. *Journal of Visual Communication and Image Representation*, 24(6) :678 – 685, 2013.
- [89] E.H. Land and John J. McCann. Lightness and retinex theory. *J. Opt. Soc. Am.*, 61(1) :1–11, Jan 1971.
- [90] Ann B Lee, David Mumford, and Jिंगgang Huang. Occlusion models for natural images : A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision*, 41(1-2) :35–59, 2001.
- [91] Reiner Lenz, Linh Viet Tran, and Peter Meer. Moment based normalization of color images. In *Multimedia Signal Processing, 1999 IEEE 3rd Workshop on*, pages 103–108. IEEE, 1999.
- [92] Bing Li, De Xu, Weihua Xiong, and Songhe Feng. Color constancy using achromatic surface. *Color Research & Application*, 35(4) :304–312, 2010.
- [93] Tony Lindeberg. Scale-space theory : A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2) :225–270, 1994.
- [94] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60 :91–110, 2004.
- [95] David L MacAdam. Visual sensitivities to color differences in daylight. *JOSA*, 32(5) :247–273, 1942.
- [96] Donald M MacKay. Towards an information-flow model of human behaviour. *British Journal of Psychology*, 47(1) :30–43, 1956.

- [97] Stéphane G Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 11(7) :674–693, 1989.
- [98] Laurence T Maloney. Evaluation of linear models of surface spectral reflectance with small numbers of parameters. *JOSA A*, 3(10) :1673–1683, 1986.
- [99] Jiri Matas, Ondrej Chum, Martin Urban, and Tomas Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10) :761–767, 2004.
- [100] Jiri Matas, Dimitri Koubaroulis, and Josef Kittler. Colour image retrieval and object recognition using the multimodal neighbourhood signature. In *Proceedings of the 6th European Conference on Computer Vision-Part I, ECCV '00*, pages 48–64, London, UK, UK, 2000. Springer-Verlag.
- [101] Baptiste Mazin, Julie Delon, and Yann Gousseau. Illuminant estimation from projections on the planckian locus. In *ECCV Workshops (2)*, pages 370–379, 2012.
- [102] Baptiste Mazin, Julie Delon, and Yann Gousseau. Estimation automatique d’illuminant multiples. In *proceedings of GRETSI 2013*, 2013.
- [103] Krystian Mikolajczyk and Cordelia Schmid. Indexing based on scale invariant interest points. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 525–531. IEEE, 2001.
- [104] Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *Computer Vision—ECCV 2002*, pages 128–142. Springer, 2002.
- [105] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell.*, 27 :1615–1630, 2005.
- [106] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2) :43–72, 2005.
- [107] Florica Mindru, Tinne Tuytelaars, Luc Van Gool, and Theo Moons. Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94(1) :3–27, 2004.
- [108] Lionel Moisan, Pierre Moulon, and Pascal Monasse. Automatic Homographic Registration of a Pair of Images, with A Contrario Elimination of Outliers. *Image Processing On Line*, 2012, 2012.
- [109] Pierre Moreels and Pietro Perona. Evaluation of features detectors and descriptors based on 3d objects. *International Journal of Computer Vision*, 73(3) :263–284, 2007.
- [110] Jean-Michel Morel and Guoshen Yu. Asift : A new framework for fully affine invariant image comparison. *SIAM J. Img. Sci.*, 2(2) :438–469, 2009.
- [111] Nathan Moroney, Mark D Fairchild, Robert WG Hunt, Changjun Li, M Ronnier Luo, and Todd Newman. The ciecam02 color appearance model. In *Color and Imaging Conference*, volume 2002, pages 23–27. Society for Imaging Science and Technology, 2002.

- [112] Karsten Muhlmann, Dennis Maier, Jurgen Hesser, and Reinhard Manner. Calculating dense disparity maps from color stereo images, an efficient implementation. *International Journal of Computer Vision*, 47(1-3) :79–88, 2002.
- [113] Sarif Kumar Naik and CA Murthy. Distinct multicolored region descriptors for object recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7) :1291–1296, 2007.
- [114] Shree K Nayar and Ruud M Bolle. Reflectance based object recognition. *International Journal of Computer Vision*, 17(3) :219–240, 1996.
- [115] Remakant Nevatia. A color edge detector and its use in scene segmentation. *IEEE Transactions on systems, Man, and Cybernetics*, 7(11) :820–826, 1977.
- [116] Ron Ohlaxnder, Keith Price, and D Raj Reddy. Picture segmentation using a recursive region splitting method. *Computer Graphics and Image Processing*, 8(3) :313–333, 1978.
- [117] Yu-Ichi Ohta, Takeo Kanade, and Toshiyuki Sakai. Color information for region segmentation. *Computer graphics and image processing*, 13(3) :222–241, 1980.
- [118] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7) :971–987, 2002.
- [119] B. Olshausen and D. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381 :607–609, 1996.
- [120] Bruno A Olshausen et al. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583) :607–609, 1996.
- [121] Patrick Perez, Carine Hue, Jaco Vermaak, and Michel Gangnet. Color-based probabilistic tracking. In *Computer vision—ECCV 2002*, pages 661–675. Springer, 2002.
- [122] Irwin G. Priest. A proposed scale for use in specifying the chromaticity of incandescent illuminants and various phases of daylight. *J. Opt. Soc. Am.*, 23(2) :41–45, Feb 1933.
- [123] Pedro Quelhas and Jean-Marc Odobez. A Color and Gradient Local Descriptor Fusion Scheme For Object Recognition. Technical report, EPFL, 2003. Published in WIAMIS04.
- [124] J. Rabin, J. Delon, and Y. Gousseau. Circular earth mover’s distance for the comparison of local features. In *Proceedings of ICPR 2008*, December 2008.
- [125] Julien Rabin, Julie Delon, and Yann Gousseau. A statistical approach to the matching of local features. *SIAM J. Imaging Sciences*, 2(3) :931–958, 2009.
- [126] Jorma Rissanen. A universal prior for integers and estimation by minimum description length. *The Annals of statistics*, pages 416–431, 1983.
- [127] Charles Rosenberg, Martial Hebert, and Sebastian Thrun. Color constancy using kl-divergence. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 1, pages 239–246. IEEE, 2001.
- [128] Y. Rubner, C. Tomasi, and L. J. Guibas. A metric for distributions with applications to image databases. In *Proceedings of the 1998 IEEE International Conference on Computer Vision*, pages 59–66, January 1998.

- [129] Y. Rubner, C. Tomasi, and L.J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2) :99–121, 2000.
- [130] Daniel L Ruderman. The statistics of natural images. *Network : computation in neural systems*, 5(4) :517–548, 1994.
- [131] Daniel L Ruderman, Thomas W Cronin, and Chuan-Chin Chiao. Statistics of cone responses to natural images : Implications for visual coding. *JOSA A*, 15(8) :2036–2045, 1998.
- [132] Guillermo Sapiro. Color and illuminant voting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 21 :1210–1215, 1999.
- [133] R. Sève. *Physiques de la couleur*. Masson, 2009.
- [134] Steven A Shafer. Using color to separate reflection components. *Color Research & Application*, 10(4) :210–218, 1985.
- [135] F. Shahbaz Khan, J. van de Weijer, and M. Vanrell. Top-down color attention for object recognition. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 979–986, 2009.
- [136] Lilong Shi and Brian. Re-processed version of the gehler color constancy dataset of 568 images, 2010.
- [137] Eero P Simoncelli and Edward H Adelson. Noise removal via bayesian wavelet coring. In *Image Processing, 1996. Proceedings., International Conference on*, volume 1, pages 379–382. IEEE, 1996.
- [138] Alvy Ray Smith. Color gamut transform pairs. *SIGGRAPH Comput. Graph.*, 12 :12–19, August 1978.
- [139] Helman Stern and Boris Efros. Adaptive color space switching for tracking under varying illumination. *Image and Vision Computing*, 23(3) :353–364, 2005.
- [140] Markus A Stricker. Color and geometry as cues for indexing. Technical report, The University of Chicago - Department of Computer Science, 1992.
- [141] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7 :11–32, 1991.
- [142] E. Tola, V. Lepetit, and P. Fua. Daisy : An efficient dense descriptor applied to wide-baseline stereo. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(5) :815–830, may 2010.
- [143] Antonio Torralba, Robert Fergus, and William T Freeman. 80 million tiny images : A large data set for nonparametric object and scene recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(11) :1958–1970, 2008.
- [144] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9) :1582–1596, 2010.
- [145] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Trans. Image Process.*, 16(9) :2207–2214, 2007.

- [146] Joost Van De Weijer, Theo Gevers, and J-M Geusebroek. Edge and corner detection by photometric quasi-invariants. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4) :625–630, 2005.
- [147] Joost Van De Weijer, Theo Gevers, and Arnold WM Smeulders. Robust photometric invariant features from the color tensor. *Image Processing, IEEE Transactions on*, 15(1) :118–127, 2006.
- [148] Joost Van De Weijer and Cordelia Schmid. Coloring local feature extraction. In *European Conference on Computer Vision*, pages 334–348, 2006.
- [149] Joost Van De Weijer, Cordelia Schmid, and Jakob Verbeek. Using high-level visual information for color constancy. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE, 2007.
- [150] Javier Vazquez-Corral, Maria Vanrell, Ramon Baldrich, and Francesc Tous. Color constancy by category correlation. *IEEE Transactions on Image Processing*, 21(4) :1997–2007, April 2012.
- [151] J. von Kries. *Influence of Adaptation on the Effects Produced by Luminous Stimuli*. D.L. MacAdam, ed. MIT Press, 1970.
- [152] Thomas Wachtler, Te-Won Lee, and Terrence J Sejnowski. Chromatic structure of natural scenes. *JOSA A*, 18(1) :65–77, 2001.
- [153] Junqiu Wang and Yasushi Yagi. Integrating color and shape-texture features for adaptive real-time object tracking. *Image Processing, IEEE Transactions on*, 17(2) :235–240, 2008.
- [154] Philippe Weinzaepfel, Hervé Jégou, and Patrick Pérez. Reconstructing an image from its local descriptors. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 337–344. IEEE, 2011.
- [155] Günther Wyszecki and W. S. Stiles. *Color Science : Concepts and Methods, Quantitative Data and Formulae*. Wiley-Interscience, 2 edition, August 2000.
- [156] Dong Zhang, Weiqiang Wang, Wen Gao, and Shuqiang Jiang. An effective local invariant descriptor combining luminance and color information. In *ICME*, pages 1507–1510, 2007.
- [157] Chao Zhu, C-E Bichot, and Liming Chen. Multi-scale color local binary patterns for visual object classes recognition. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 3065–3068. IEEE, 2010.
- [158] Song Chun Zhu and David Mumford. Prior learning and gibbs reaction-diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(11) :1236–1250, 1997.

Méthodes robustes pour l'estimation d'illuminants et la prise en compte de la couleur en comparaison d'images

Baptiste Mazin

Résumé : Cette thèse traite de l'utilisation de la couleur en traitement d'images et en vision par ordinateur. Deux problèmes sont étudiés :

- l'estimation d'illuminants,
- la mise en correspondance de descripteurs locaux pour la comparaison d'images couleur.

Par définition, les surfaces achromatiques renvoient un spectre lumineux ayant la même distribution fréquentielle que le spectre de l'illuminant. Les détecter permet donc de recouvrer l'illuminant. En supposant que l'ensemble des couleurs que peut prendre un illuminant est limité (équation de Planck), il est possible de sélectionner les pixels appartenant à une surface potentiellement grise. Une méthode de vote est alors appliquée, permettant de sélectionner un ou plusieurs illuminants. L'algorithme final possède de nombreux avantages : il est efficace, intuitif, ne nécessite pas de phase d'apprentissage et requiert peu de paramètres, qui s'avèrent stables. De plus, la méthode de vote permet de s'adapter aux cas où plusieurs sources lumineuses de couleurs différentes éclairent la scène photographiée.

Les descripteurs locaux sont des outils très puissants pour comparer finement des images. Cependant, le rôle de la couleur dans l'étape d'appariement a fait l'objet de relativement peu d'études. Le problème principal que nous considérons ici est précisément celui de l'apport de la couleur pour l'appariement de descripteurs locaux. Un usage local de la couleur permet-il de désambiguïser les situations où la luminance seule est insuffisante ? Et si oui, dans quelles proportions ? Nous proposons quatre descripteurs permettant de décrire de manière détaillée le contexte local de points clés dans l'optique de l'appariement de ces derniers. L'idée maîtresse développée ici est qu'une information précise ne peut être obtenue qu'en décrivant à la fois les distributions de couleurs et les transitions entre couleurs. Les nombreuses expériences proposés montrent clairement l'apport positif de la couleur pour la fiabilité de la mise en correspondance locale.

Mots-clefs : couleur, descripteurs locaux, sift, invariance photométrique, estimation d'illuminant.

Abstract : This thesis addresses the use of color in image processing and computer vision. Two problems are studied :

- illuminant estimation,
- local descriptors matching for color images comparison.

Achromatic surfaces are defined as surfaces reflecting a spectrum with the same frequency distribution than the illuminant. Consequently, recovering these surfaces allows to estimate the illuminant. Assuming that the range of colors taken by an illuminant is limited (Planck equation), it is possible to select the pixels belonging to a potentially gray surface. A voting procedure is then applied to select one or more illuminants. The proposed algorithm has many advantages : it is effective, intuitive, does not rely on a learning phase and requires only few parameters. In addition, the voting procedure can be adapted to handle cases where multiple light sources of different colors illuminate the scene.

Local descriptors are powerful tools to compare images. However, few studies concern the influence of color in the matching step. The main problem that we consider here is the contribution of the color matching of local descriptors. Does the local use of color allow to disambiguate situations where the luminance alone is insufficient ? And if so, how much ? We propose four descriptors to precisely describe the local context of key points in the matching step. The main idea developed here is that accurate information can only be obtained by describing both the color distributions and transitions between colors. The many experiences presented clearly show the positive contribution of color to the reliability of the local matching.

Keywords : color, local descriptors, image matching, sift, photometric invariance, illuminant estimation.

