THÈSE

pour l'obtention du Grade de

DOCTEUR DE L'UNIVERSITÉ DE POITIERS

Faculté des Sciences Fondamentales et Appliquées
Diplôme National - Arrêté du 25 Mai 2016

Ecole Doctorale : Sciences et Ingénierie des Systèmes, Mathématiques, Informatique - SISMI
Secteur de Recherche : Traitement du Signal et des Images

Présentée par :
Hermine Chatoux

Prise en compte métrologique de la couleur dans un contexte de classification et d'indexation

Directrice de Thèse :
Christine Fernandez-Maloigne

Co-directeur de Thèse :
François Lecellier

Soutenue le 21 mai 2019

Devant la Commission d'Examen

JURY

Alice Caplier, Professeur des universités, INP Grenoble ..... Rapporteur

Eduardo Provenzi, Professeur des universités, Université de Bordeaux ..... Rapporteur

Olivier Lezoray, Professeur des universités, Université de Caen ..... Examinateur

Jean-Yves Tourneret, Professeur des universités, INP Toulouse ..... Examinateur

Noël Richard, Maître de Conférences, Université de Poitiers ..... Examinateur

Christine Fernandez-Maloigne, Professeur des universités, Université de Poitiers .. Directeur

François Lecellier, Maître de Conférences, Université de Poitiers ..... Co-directeur

Remerciements

Pour commencer, je remercie Christine Fernandez-Maloigne et François Lecellier pour m'avoir offert l'opportunité de rentrer dans l'équipe ICONES du laboratoire XLIM. Je voudrais aussi les remercier pour leur encadrements tout au long de ma thèse : François pour la liberté qu'il m'a laissée, Christine pour avoir toujours su se rendre disponible malgré ses nombreuses casquettes. Merci au projet CPER-FEDER ESHALI qui a permis le financement et donc la réalisation de cette thèse.

Je souhaite aussi remercier Jean-yves Tourneret, professeur des universités à l'INP de Toulouse, pour m'avoir permis de découvrir l'équipe ICONES du XLIM et pour avoir présidé mon jury de thèse. Je tiens aussi à remercier Alice Caplier, professeur des universités à l'INP de Grenoble, et Edoardo Provenzi, professeur des universités à l'université de Bordeaux pour avoir accepté de rapporter ce manuscrit de thèse. Merci également à Olivier Lézoray, professeur des universités à l'université de Caen pour sa participation à mon jury de thèse.

Je voudrais aussi remercier le *vilain pas beau qui n'aime personne* qui m'a pris sous son aile colorée, j'ai nommé Noël Richard. Sans lui, je n'aurais pu faire fleurir toutes les idées de cette thèse. Tour à tour pluie ou beau temps, il a été le tuteur aux volubilis de cette aventure. Merci aux arbres de connaissances que sont les enseignants-chercheurs (notamment Thierry, Benoît, Hakim, Philippe M., et les autres) pour les fruits qu'ils m'ont donnés. Un éducatif merci à Yves, Mickaël, et Frédéric pour leur accueil sans préjugés dans l'équipe de mathématiques de l'IUT.

Un chaleureux merci aux BIATS (acronyme dont nous avons si souvent discuté) : Alicia, Bruno, Caroline, Françoise, Lætitia, Sophie, Yoann, … pour leur travail essentiel et aussi pour leur bonne humeur et leur soutien en toutes circonstances.

Un affectueux merci aux différents membres de mon bureau, passé et futur, pour tous ces jours ensembles constitués d'instants de concentration, d'entraide, de détente, de réprimande ou de craquage complet. Merci à Mathieu pour ses achats compulsifs, Alex pour nos dialogues de sourds, étrangement, constructifs, Pascal pour ses blagues (ou pas ?), Lydie pour nos échanges philosophiques et parentaux...

Merci au café social qui réunit tout ce beau monde, et bien d'autre encore (citons notamment les, pas si réservés, membres de l'équipe géométrie discrète), dans des moments qu'il m'est interdit de raconter en dehors de la cafétéria.

Un merci sportif à Lydie et Raphaël pour m'avoir accompagnée dans ces escapades qui m'étaient indispensables durant la rédaction de ce manuscrit. Je souhaite aussi remercier la *Blonde* et le *Barbu* pour tant de chose : du baby-sitting, du shopping, mais surtout des échanges riches, bigarrés et enthousiastes. Un ineffable merci à Julien pour m'avoir soutenu avant et pendant ma thèse (et j'espère encore longtemps après) et à Iris dont un seul sourire suffit à illuminer ma journée. Un merci filial à mes parents pour leur soutien pendant la rédaction et les 35 ans qui ont précédée. Un merci studieux à mon père pour les heures de relecture qu'il m'a accordées. Un merci polychrome à ma mère qui m'a présenté la marmite des couleurs pour que je puisse tomber dedans quand j'étais petite.

Table des matières

Introduction **15**

1 Gradient couleur **19**
1.1 Qu'est-ce qu'un gradient ? . . . 21
1.1.1 En $1D$ . . . 22
1.1.2 Vers la nD . . . 23
1.1.3 Mise en œuvre informatique . . . 24
1.2 En regard de l'état de l'art . . . 26
1.2.1 Domaine physique . . . 27
1.2.2 Espace transformé . . . 27
1.2.3 Domaine perceptuel . . . 29
1.3 Résultats . . . 29
1.3.1 Base de données artificielles . . . 30
1.3.2 Comparaison subjective des performances pour des images réelles . . . 38
1.3.3 Exemples d'applications de calcul de gradient . . . 41
1.4 Synthèse . . . 49

2 Détecteurs de points d'intérêt **51**
2.1 État de l'art . . . 53
2.1.1 Détecteurs de coins . . . 54
2.1.2 Détecteurs basés *blob* . . . 58
2.1.3 Détecteurs basés régions . . . 60
2.2 Détecteurs de points d'intérêt couleur . . . 61
2.2.1 Adaptation aux multi-échelles . . . 62
2.2.2 Adaptation à la couleur . . . 64
2.2.3 Mise en forme . . . 64
2.2.4 Sélection des “bons” points d'intérêt . . . 67
2.3 Preuve de l'approche . . . 69
2.3.1 Base de données artificielles . . . 71
2.3.2 Base de données Oxford . . . 78
2.3.3 Comparaison point d'intérêt/point visuellement saillant . . . 86
2.4 Synthèse . . . 93

3 Critère de dissimilarité **95**
3.1 Distances et mesures de similarité . . . 97
3.1.1 Distances et espaces vectoriels . . . 97
3.1.2 Distances, divergences entre probabilités . . . 97
3.1.3 Autres distances . . . 101
3.1.4 Mesures intégrant une modélisation statistique . . . 101
3.2 Nature des données et mesures utilisées . . . 103
3.2.1 Séquences des bits . . . 103
3.2.2 Les histogrammes comme densités de probabilité . . . 105
3.2.3 Modélisation ou paramétrisation d'une pdf . . . 106
3.2.4 Extension aux contextes couleur et spectral . . . 106
3.3 Éléments de décision d'une mesure de similarité . . . 108
3.3.1 Comportement avec une distribution de synthèse . . . 108
3.3.2 Distributions d'images couleurs . . . 113
3.4 Synthèse . . . 122

4 Descripteurs de points d'intérêt **125**
4.1 Décrire le contenu spatio-chromatique local . . . 126
4.1.1 Valeurs de pixels . . . 130
4.1.2 Décomposition fréquentielle . . . 130
4.1.3 Gradients . . . 131
4.1.4 Différences de valeurs . . . 133
4.1.5 Motifs . . . 134
4.2 Invariance des descripteurs . . . 136
4.2.1 Protocole . . . 136
4.2.2 Résultats . . . 140
4.3 Un descripteur couleur déduit des conjectures de Julesz . . . 146
4.3.1 Première conjecture : cooccurrence . . . 146
4.3.2 Deuxième conjecture : ordres supérieurs à deux . . . 148
4.3.3 Troisième conjecture : texton . . . 149
4.3.4 Relocated colour contrast occurrence . . . 150
4.4 Preuve de l'approche . . . 152
4.4.1 Base de données artificielles . . . 153
4.4.2 Classification de textures . . . 160
4.5 Synthèse . . . 164

Conclusion et Perspectives **167**
Conclusion . . . 167
Perspectives . . . 168

A Teinte *vs* luminosité **171**

B Évaluation des gradients couleur pour des images générées dans $CIELAB$ **173**

C Impact de la forme de distribution pour des images générées en RGB **175**

D Impact de la forme de distribution pour des images générées en $CIELAB$ **179**

E Images de la base Oxford utilisées **183**

F Tests d'hypothèse sur le modèle Gaussien des images couleur **186**

G Invariance de descripteur **188**
G.1 Influence de la taille de la zone d'intérêt par transformation . . . 188
G.2 Influence de la taille de la zone d'intérêt par descripteur . . . 190

H Deuxième conjecture de Julesz **196**

I Construction de la base de données synthétiques fractales **197**

Table des figures

1 L'objectif de la métrologie . . . 16

1.1 Le graal : un résultat indépendant du capteur . . . 20
1.2 Différence d'approche entre les différents gradients . . . 21
1.3 Exemple de gradient de fonction $1D$ . . . 23
1.4 Angle estimé par le gradient de Di Zenzo . . . 26
1.5 Etat de l'art des gradients couleur . . . 28
1.6 Exemples de distributions naturelles . . . 30
1.7 Exemples d'images générées . . . 31
1.8 Exemples d'enveloppes de distributions . . . 33
1.9 Illustration du taux de séparabilité . . . 34
1.10 Exemple de segmentation pratique comparé à la théorique . . . 35
1.11 Exemples d'extraction de bords pour des images générées avec $\Delta E = 3$ . . . 36
1.12 Exemples de détection de bords pour des images générées avec un $\Delta E = 3$ . . . 37
1.13 Comparaison des gradients sur l'image nuage de la base VisTex . . . 38
1.14 Comparaison d'une fleur de la base Stex . . . 39
1.15 Différences en extraction de bords entre le FVG et Di Zenzo . . . 40
1.16 Domaine de perception des différents appareils . . . 41
1.17 Fonctions de sensibilité spectrale de trois appareils d'acquisition . . . 41
1.18 Matrices de Gram pour les capteurs couleur et multi-spectral . . . 42
1.19 Matrice de Gram pour le capteur hyper-spectral . . . 43
1.20 Résultats pour trois types de capteurs restreints au visible . . . 43
1.21 Résultats pour les capteurs multi- et hyper-spectral sur leur spectre complet . . . 45
1.22 Courbes de sensibilité pour les différentes visions . . . 46
1.23 Normes des gradients estimés pour les différentes visions . . . 47
1.24 Zoom des différentes normes de gradients estimés pour les différentes visions couleur 48
1.25 Différences entre les visions couleur normale et déficientes . . . 48

2.1 Décomposition du détecteur de Harris . . . 52
2.2 Etat de l'art des détecteurs couleur . . . 55
2.3 Schéma de fonctionnement du détecteur multi-échelles Harris-Laplace . . . 57
2.4 Illustration du fonctionnement de SUSAN . . . 57
2.5 Différence entre LoG et DoG . . . 59
2.6 Illustration du fonctionnement multi-échelles des DoG . . . 60
2.7 Comparaison des étapes de calcul du détecteur de Harris . . . 62
2.8 Exemples de filtres dérivateurs Gaussiens . . . 63
2.9 Illustration de la nécessité du filtre intégrateur . . . 65
2.10 Limites de l'utilisation du filtre intégrateur . . . 66
2.11 Instabilité de la constante k . . . 67
2.12 Résultat sans sélection de points d'intérêt . . . 67
2.13 Résultat avec sélection de points d'intérêt . . . 70
2.14 Zoom des résultats de sélection pour l'image réelle . . . 70
2.15 Exemples d'images artificielles pour la détection de points d'intérêt . . . 71
2.16 Détection des points d'intérêt en fonction de l'écart couleur . . . 72
2.17 Exemples d'images pour $\Delta E = 30$ . . . 72
2.18 Rappel en fonction de la distance aux points exacts . . . 73
2.19 Images de points d'intérêt détectés pour les détecteurs couleur . . . 74
2.20 Courbes de rappel . . . 77
2.21 Influence de la complexité sur le rappel et la précision . . . 78

2.22 Exemple de points d'intérêt communs entre la scène et une vue 80
2.23 Résultats de répétabilité des détecteurs sur la base Oxford 82
2.24 Résultats de répétabilité des détecteurs sur la base Oxford 83
2.25 Résultats de répatabilité des détecteurs sur la base Oxford 84
2.26 Impact du nombre de points d'intérêt détectés 85
2.27 Protocole de l'expérience 87
2.28 Exemples d'images utilisées pour l'expérience 88
2.29 Exemples de résultats de points extraits 89
2.30 Résultats préliminaires 90
2.31 Adéquation entre points d'intérêt et points saillants 91
2.32 Adéquation entre points d'intérêt et points saillants 92

3.1 Organisation des distances à utiliser en fonction du type des données du descripteur 96
3.2 Schéma de saturation de la norme $\mathcal{L}_2$ 98
3.3 Exemple de répartition de distribution pour l'EMD 99
3.4 Organisation des distances utilisées en fonction du type des données du descripteur dans la pratique 104
3.5 Illustration des variations d'une loi normale $2D$ 109
3.6 Réponse de la divergence de KL à l'éloignement 110
3.7 Réponse de la divergence de KL à un changement de variance 111
3.8 Réponse de la divergence de KL à une variation de covariance 112
3.9 Comparaison d'échelle des réponses pour les différentes mesures 113
3.10 Quatre textures de même nature avec leurs histogrammes 115
3.11 Trois textures neutres et leurs histogrammes 117
3.12 Trois textures colorées et leurs histogrammes 118
3.13 Comparaison de chevauchement de distributions 119

4.1 Illustration du positionnement sémantique des termes utilisés 127
4.2 Organisation des attributs pour la caractérisation de l'aspect non-uniforme 127
4.3 Etat de l'art des descripteurs couleur 128
4.4 Schéma de conception de SIFT 132
4.5 Illustrations des approches par différence de valeurs 134
4.6 Exemple de transformations appliquées à une image 138
4.7 Exemple d'une image de la base *photo CVKODAK* 140
4.8 Invariance à l'échelle 140
4.9 Invariance à la rotation 141
4.10 Invariance aux cisaillements 142
4.11 Invariance à la luminosité 142
4.12 Comparaison des résultats sur l'invariance d'échelle en fonction de la zone d'intérêt 144
4.13 Comparaison des résultats d'invariance à la rotation avec ou sans à priori 145
4.14 Illustration de la première conjecture de Julesz 147
4.15 Exemples de matrices $\mathrm{C_2O}$ 148
4.16 Textures dont les statistiques d'ordre un et deux sont identiques 148
4.17 Illustration de l'approche longueur de plage 149
4.18 Illustration de l'approche LBP 150
4.19 Similarités et différences entre les corrélogrammes et les LBP 150
4.20 Comparaison des approches s'inspirant de la première conjecture de Julesz 151
4.21 Différence entre les distributions C_2O et $\mathrm{RC_2O}$ 152
4.22 Incohérence des moyennes générées 157
4.23 Incohérence des moyennes générées 158
4.24 Variations des précisions avec le coefficient de Hurst 159

4.25 Variations du $\mathrm{C_2O}$ et $\mathrm{RC_2O}$ en fonction de la distance et de l'angle 160
4.26 Quelques images de texture de la base *VisTex*, et leurs distributions $\mathrm{C_2O}$ 162
4.27 Textures feuille de laitue de la base *KTH-TIPS2-b* et leurs distributions $\mathrm{C_2O}$ 163

A.1 Modèle de courbes $\alpha(S)$ 171

B.1 Exemple d'extraction de bords pour des images générées avec $\Delta E = 3$ 174

C.1 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 175
C.2 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 176
C.3 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 177
C.4 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 178

D.1 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 179
D.2 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 180
D.3 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 181
D.4 Résultat de séparabilité du gradient pour différentes formes de distributions . . . 182

E.1 Images de la base Oxford 184
E.2 Images de la base Oxford 185

G.1 Comparaison des résultats d'invariances des descripteurs 188
G.2 Comparaison des résultats d'invariances des descripteurs 189
G.3 Comparaison des résultats sur l'invariance d'échelle 191
G.4 Comparaison des résultats sur l'invariance de rotation 192
G.5 Comparaison des résultats sur l'invariance de cisaillement horizontal 193
G.6 Comparaison des résultats sur l'invariance de cisaillement vertical 194
G.7 Comparaison des résultats sur l'invariance à la luminosité 195

H.1 Solution de segmentation pour les statistiques d'ordre deux identiques 196

Liste des tableaux

1 Notations . . . 13

1.1 Moyennes et écarts-types pour des taux de séparabilité choisis . . . 36
1.2 Matrices de Gram pour les différentes visions couleur . . . 46

2.1 Synthèse de sélection des points . . . 68
2.2 Approches utilisées . . . 69
2.3 Statistiques du nombre de points extraits . . . 75
2.4 Statistique des points détectés sur la base Oxford . . . 81
2.5 Nombre de points d'intérêt détectés par nos propositions . . . 89

3.1 Table des mesures entre textures proches . . . 116
3.2 Table des distances entre images à partir de la différence couleur entre leurs moyennes 116
3.3 Divergence de KL entre les six images . . . 119
3.4 Distance $\mathcal{L}_2$ entre les six images . . . 120
3.5 Distance D_E entre les six images . . . 121
3.6 Tests d'hypothèse de normalité . . . 122

4.1 Résumé de transformations créées . . . 137
4.2 Invariance des descripteurs selon leurs auteurs . . . 139
4.3 Taux d'appariement après transformation par réflexion . . . 143
4.4 Invariances des descripteurs calculés . . . 144
4.5 Récapitulatif des paramètres de l'exprérience . . . 155
4.6 Classification sur la base de données fractales avec H et Σ constants . . . 156
4.7 Classification sur la base de données fractales avec μ et H constants . . . 156
4.8 Vérification des moyennes et écarts-type par μ générés . . . 157
4.9 Classification sur la base de données fractales avec μ et Σ constants . . . 158
4.10 Classification sur la base de données fractales . . . 159
4.11 Bases utilisées pour les expériences . . . 161
4.12 Divergence D_{KL} inter et intra-classe pour les images de la base *Outex* . . . 161
4.13 Divergence D_{KL} pour des images de la base *VisTex* . . . 162
4.14 Divergence D_{KL} pour les variations d'échelle d'une feuille de laitue . . . 162
4.15 Comparaison des taux de classification(%) obtenus sur les différentes bases présentées . . . 163

B.1 Moyennes et écarts-type de la précision . . . 174

F.1 Tests d'hypothèse de normalité pour les images de textures proches . . . 186
F.2 Tests d'hypothèse de normalité pour les images de textures différentes . . . 187

TABLE 1 – **Notations**

Notation	**Signification**
x, y, z	Localisation spatiale dans $\mathbb{R}^n$: $x = (x_1, x_2)$ pour un espace $2D$.
a, b, c	Coordonnées couleur/multivaluée dans $\mathbb{R}^m$, $c = (c_1, c_2, c_3)$ pour un espace trichromatique couleur.
$s_i, S_i()$	Canal $s_i, i \in [1, m]$ définit par sa courbe de sensibilité $S_i()$.
I, F, G	Ensemble ou image, deux définitions possibles : $I : \mathbb{R}^n \to \mathbb{R}^m \;\; I(x) = c$, c est une fonction ; $I \in \mathbb{R}^n \times \mathbb{R}^m \;\; I(x, c)$ fonction en x et c.
$\mathcal{X}, \mathcal{C}, \mathcal{F}$	Support de la variable spécifiée x, c, f : $x \in \mathcal{X} = \{x_i \in [xmin_i; xmax_i]; i \in [1; n]\}$.
$\mathcal{S}(x)$	Voisinage spatial de x.
$\#\mathcal{S}$	Cardinal de l'ensemble.
$\langle ., . \rangle_p$	Produit scalaire : $\langle x, y \rangle_p = \sum_{i=1}^{n} (x_i y_i)^p$.
$\|.\|_p$	L_p norme (Minkowski d'ordre p) : $\|x\|_p = \sqrt[\frac{1}{p}]{\langle x, x \rangle_p}$.
$d_\Sigma(., .)$	Distance de Mahalanobis associée à la matrice de covariance Σ.
∇	Gradient.

Espace continu	
$\frac{\partial I(a)}{\partial b}$	Dérivée partielle de $I(a)$ selon b : $\frac{\partial I(c_2)}{\partial x_1}$: partie x_1 du gradient en c_2.
∇I	Gradient de l'image I : for $I : \mathbb{R}^2 \to \mathbb{R}^3$ $\nabla I(x, c) = \begin{pmatrix} \frac{\partial I(c_1)}{\partial x_1} & \frac{\partial I(c_2)}{\partial x_1} & \frac{\partial I(c_3)}{\partial x_1} \\ \frac{\partial I(c_1)}{\partial x_2} & \frac{\partial I(c_2)}{\partial x_2} & \frac{\partial I(c_3)}{\partial x_2} \end{pmatrix}$.

Espace discret	
$\Delta_b I(a)$	Dérivée partielle de $I(a)$ selon b : $\Delta_{x_1} I(c_2)$ est la différence de deux c_2 suivant la direction x_1.
∇I	Gradient de l'image I : for $I : \mathbb{R}^2 \to \mathbb{R}^3$ $\nabla I(x, c) = \begin{pmatrix} \Delta_{x_1} I(c_1) & \Delta_{x_1} I(c_2) & \Delta_{x_1} I(c_3) \\ \Delta_{x_2} I(c_1) & \Delta_{x_2} I(c_2) & \Delta_{x_2} I(c_3) \end{pmatrix}$.

$[a]_b$	a Modulo b, *i.e.* reste de la division Euclidienne de a par b.
$*$	Produit de convolution $(f * g)(x) = \int_\Omega f(t) g(x - t) dt$.
(A, B)	Droite passant par les points A et B.
$[A, B]$	Segment borné par les points A et B.
$p(A\|B)$	Probabilité de A sachant B.
ROI	Région d'intérêt.
POI	Point d'intérêt.

Introduction

La couleur fait partie intégrante de la vision humaine et de la nature physique du monde naturel. Elle apporte un peu de pep's à la vie et aux images niveaux de gris. La couleur est aussi source d'informations. McIlhagga and Mullen (2018) ont prouvé à l'aide d'une expérience psychovisuelle que l'œil était sensible aux "bords" couleur. Cependant, sur l'ensemble des outils de traitement d'images, plus de 80% des outils sont développés en niveaux de gris. Certains auteurs justifient leurs choix en citant Novak and Shafer (1987) qui énoncent que 90% des informations d'une scène sont contenues dans les images niveaux de gris. Nous pourrions bien évidement nous interroger sur la nature du contenu considéré : manufacturé ou naturel, ou simplement choisir de nous intéresser aux 10% restants. De plus, parmi les outils s'intéressant à la couleur, nombreuses sont les propositions qui se contentent d'une approche marginale dans un espace RGB, considérant les canaux comme orthogonaux. Mais cela revient à nier la complexité du monde vivant pour le lier à une décomposition théorique sommaire.

Plus qu'une incorporation basique de la couleur dans les outils de traitement, nous voulons rigoureusement utiliser celle-ci. En effet, la couleur prend essence dans la physique des objets et est perçue via un capteur pour le traitement d'images ou via l'œil pour l'humain. Deux cas se présentent, soit le traitement à effectuer doit employer l'humain à une étape, soit le traitement ne nécessite aucune analyse humaine. Dans le premier cas, il est important d'incorporer la perception dans l'analyse. Sinon, rester au niveau physique du capteur diminue la perte d'information. Nous proposons une approche répondant aux deux cas avec une unique formulation.

La question de la perception des couleurs n'est pas nouvelle, puisqu'elle a conduit la Commission Internationale de l'Éclairage (CIE) à mener une expérience psychovisuelle dès 1931 visant à établir le spectre des couleurs visibles par l'humain ainsi que sa concordance avec un affichage RGB. Ce type d'expérimentations montre que la métrologie est accessible à tous les niveaux de la chaîne d'analyse et de traitement de l'image couleur. Dans ce travail, nous souhaitons respecter ces contraintes métrologiques, pour produire des outils robustes, génériques (quel que soit le capteur) et dont les incertitudes peuvent être maîtrisées.

La métrologie consiste à définir une unité de mesure fixe et immuable pour ensuite pouvoir comparer les mesures effectuées avec différents outils basés sur cette même unité. La figure 1 illustre cette idée. La taille de l'enfant à l'instant où la photo a été prise est une constante. Quel que soit l'appareil de mesure utilisé (mètre de couturière, règle, mètre pliant, mètre laser, etc.), nous devons retrouver la même mesure de longueur, en revanche la "précision" de la mesure dépendra de l'appareil.

Le souhait d'immuabilité de la mesure, entraîne, une recherche depuis plusieurs siècles, notamment au cours des Conférences Générale des Poids et des Mesures. La 17e, qui a eu lieu en 1983, a proposé de supprimer le mètre étalon pour baser l'unité de longueur sur une constante physique fixe afin d'éviter toute dérive. Notre travail s'insère dans cette veine. Sans avoir la prétention d'établir une mesure de traitement couleur immuable, nous souhaitons pouvoir comparer des mesures effectuées avec différents capteurs.

Notre analyse se place dans un contexte d'indexation ou de classification. La classification cherche à attribuer une dénomination unique, ou un ensemble de dénominations à une image donnée. Ces dénominations sont extraites d'un nombre fini de classes, généralement identifiées au préalable. L'indexation vise à extraire des images similaires à celle donnée en entrée. Ces deux tâches consistent à analyser une image pour extraire des caractéristiques afin de les comparer à celles de la base de données. Cela permet d'extraire ainsi les images similaires ou de choisir la classe d'appartenance selon des critères préalablement déterminés par les caractéristiques choisies. Deux outils sont indispensables pour réaliser ces tâches : le détecteur et le descripteur. Une analyse plus approfondie, nous montrera que la mesure de similarité choisie est très importante.

Le détecteur de points d'intérêt, comme son nom l'indique, permet de situer les points d'intérêt dans l'image. La définition de points d'intérêt mérite d'être détaillée de par sa nature mathématique

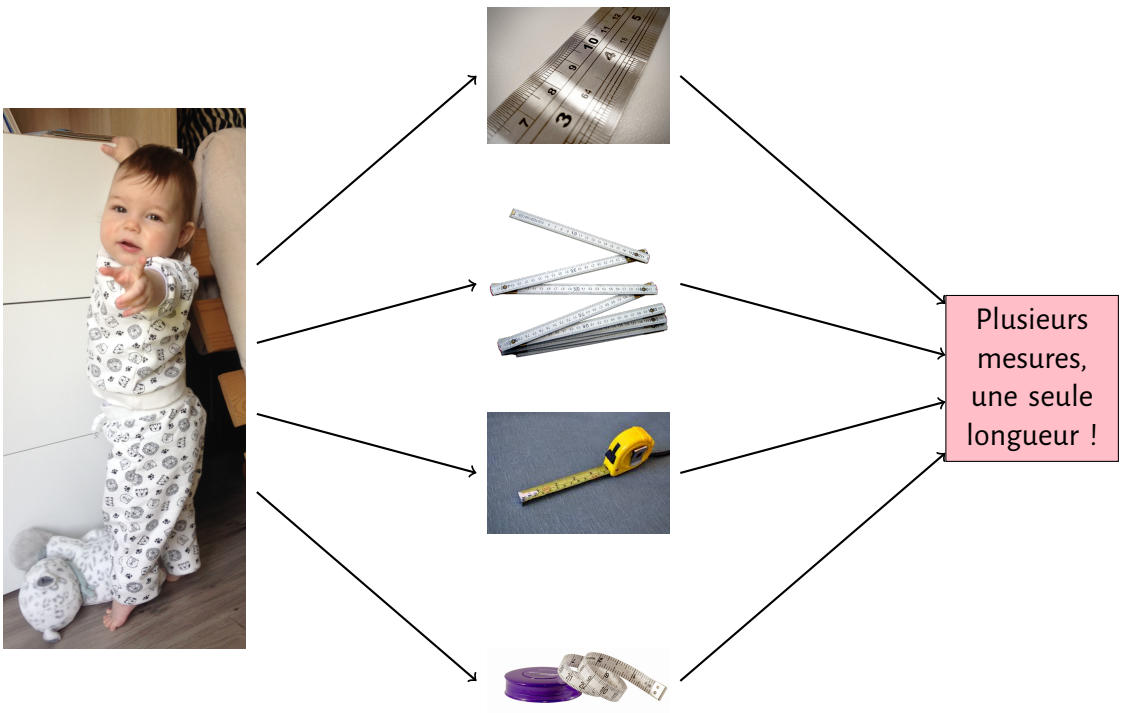


FIGURE 1 – **L'objectif de la métrologie**. Ces différents appareils donneront la même mesure de longueur $l \pm \epsilon$ à l'exactitude prés.

liée au contenu couleur. Cette nature de point saillant ou d'intérêt le destine directement à une étude sous le crible métrologique.

Identifier des points fixes ou des points d'intérêt n'a de sens que pour les retrouver dans d'autres images. Il est donc naturel de les associer à des descripteurs locaux. Le rôle du descripteur est de décrire le point d'intérêt de manière à ce qu'il puisse être reconnu dans une autre image. Comme pour le détecteur, les descripteurs sont construits avec des invariances qui les rendent robustes à certaines transformations. Peut-on s'absoudre ici aussi du regard métrologique ?

Comme nous l'avons déjà dit ces outils sont principalement développés en niveaux de gris. Nous voulons étendre ces outils à la couleur, au sens large du terme : trois canaux ou plus. L'objectif de cette thèse est de développer des outils qui prennent en compte **la couleur en respectant la physique et/ou la perception**. Nous souhaitons aussi démontrer **en quoi une prise en compte de la couleur améliore l'analyse et le traitement des images couleur**.

Nous avons choisi pour l'écriture de cette thèse de découper les chapitres par outil de traitement d'images. Ainsi, chaque chapitre commencera par un état de l'art de l'outil étudié. Cette bibliographie ne sera pas exhaustive. Comment pourrait-elle l'être après plusieurs décennies sur le sujet ? Cependant, nous cherchons à mettre en avant les points importants à la compréhension afin de révéler les manques de l'état de l'art. Nous proposons une taxonomie des outils existants dans le domaine couleur.

Nous souhaitons prouver l'apport des approches couleur que nous proposons. Pour cela, chaque chapitre présentera, dans la section résultat, une étude sur une base de données de synthèse mettant en avant l'apport de notre contribution dans le domaine couleur. Nous éprouvons ensuite la proposition sur des bases de données réelles régulièrement utilisées dans l'état de l'art.

Les outils que nous souhaitons revisiter se situent à différents niveaux d'analyse de traitement d'images. Nous commencerons au niveau élémentaire avec l'analyse du gradient couleur. Ce sera l'objet du chapitre 1. Puis, nous regarderons un niveau intermédiaire et étudierons les détecteurs de points d'intérêt couleur dans le chapitre 2. Le plus haut niveau d'analyse sera étudié sur deux chapitres. Nous nous intéresserons aux descripteurs de points d'intérêt et à la mesure de similarité qui

leur est associée. Or, les performances de ceux-ci sont directement liées à la distance associée. Nous montrerons l'indissociabilité de la paire {descripteur, mesure de similarité} dans le chapitre 3. Nous pourrons ensuite étudier les descripteurs couleur à proprement parler dans le chapitre 4.

Gradient couleur

Contenu

1.1 Qu'est-ce qu'un gradient ? 21
1.1.1 En $1D$ 22
1.1.2 Vers la nD 23
1.1.2.1 Formulation de Di Zenzo 23
1.1.2.2 Formulation du *Full-vector gradient* 24
1.1.3 Mise en œuvre informatique 24
1.2 En regard de l'état de l'art 26
1.2.1 Domaine physique 27
1.2.2 Espace transformé 27
1.2.3 Domaine perceptuel 29
1.3 Résultats 29
1.3.1 Base de données artificielles 30
1.3.1.1 Construction de la base de données 30
1.3.1.2 Impact de la forme de la distribution 34
1.3.1.3 Combinaison de formes de distributions couleur 35
1.3.2 Comparaison subjective des performances pour des images réelles . . . 38
1.3.3 Exemples d'applications de calcul de gradient 41
1.3.3.1 Vers le domaine physique, ... 41
1.3.3.2 Et au delà ! 45
1.4 Synthèse 49

Le calcul du gradient est une étape essentielle pour de nombreuses tâches en traitement d'images comme la détection de bords Belmamoun et al. (2014), la segmentation Luccheseyz and Mitray (2001), l'extraction de coins van de Weijer and Gevers (2004) ou d'attributs Abdel-Hakim and Farag (2006); Martínez et al. (2015). Les usages de ces gradients ne sont pas restreints au visible. Des combinaisons des plages du visible, proche infra-rouge et infra-rouge sont maintenant exploitées Xing et al. (2018). Face à une bibliographie prolifique sur le sujet dans le domaine couleur, mais beaucoup plus restreinte dans les domaines multi- et hyper-spectral, nous nous sommes intéressés à une approche générique sans restriction de plages spectrales, du nombre de canaux ou de sensibilité de ces canaux (fig. 1.1).

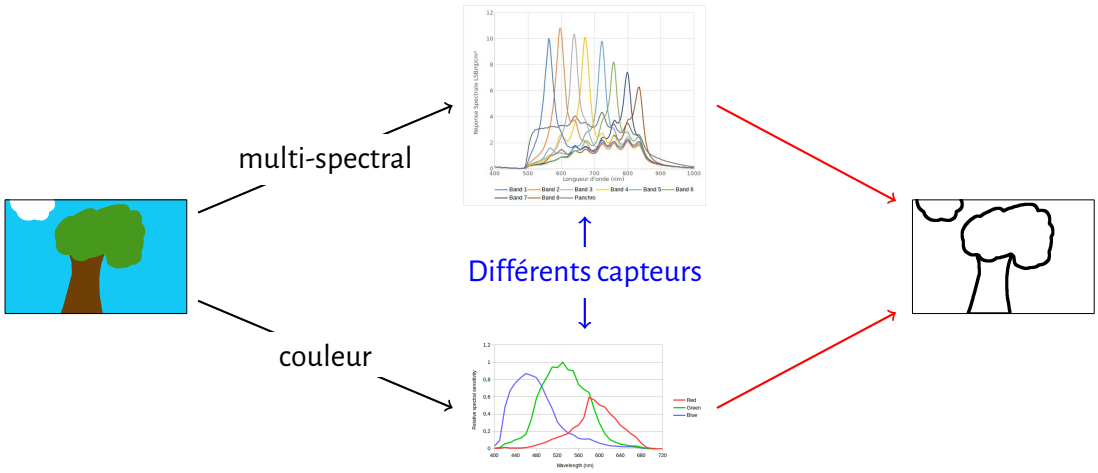


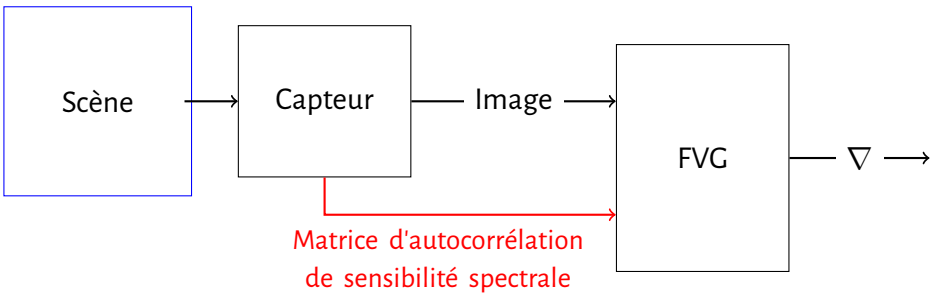
FIGURE 1.1 – **Le graal : un résultat indépendant du capteur**. L'objectif de la métrologie est de s'affranchir de l'outil de mesure.

Notre objectif est l'évaluation d'un gradient dans un contexte métrologique. Pour cela nous nous plaçons dans l'espace d'acquisition du capteur défini par ses courbes de sensibilité spectrale. D'un point de vue strictement métrologique, la mesure du gradient doit être indépendante du système de capture de l'image (fig. 1). Cependant, le capteur induit des incertitudes/limites liées à ses résolutions spatiale et spectrale. Nous choisissons donc de prendre en compte les caractéristiques du capteur (fig 1.2a) pour pouvoir s'en extraire. En effet, sans cette prise en compte, le gradient extrait est la combinaison de la scène et de l'appareil d'acquisition (fig. 1.2b). Sans dissociation des deux, la comparaison puis la combinaison naturelle des gradients calculés sur des capteurs différents doit être interdite dans un objectif métrologique.

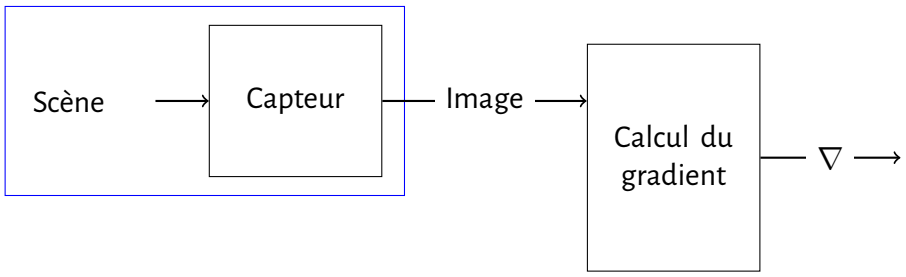
Dans ce chapitre, nous allons souvent faire référence à la proposition de Di Zenzo (1986), au cœur de nombreuses approches utilisées. Nous allons montrer les limites implicites de la formulation de Di Zenzo qui suppose l'espace de représentation comme orthogonal. Malheureusement, cette hypothèse n'est que rarement satisfaite dans le cas de capteurs couleur ou spectraux. L'inter-corrélation entre les canaux d'acquisition a été considérée par certains auteurs comme Ehrhardt and Arridge (2014) dans le domaine couleur ou par Qin and Clausi (2010) dans un contexte multivarié. Cependant, leurs approches ne distinguent pas la corrélation due à la scène de celle due au capteur comme dans la figure 1.2b.

L'adaptation à un plus grand nombre de canaux est une question récurrente depuis Cumani (1991). Ce dernier a adapté la solution de Di Zenzo au cas multi- et hyper-spectral. Cependant, il souffre des mêmes limites que l'expression de Di Zenzo. Plus récemment, Chen et al. (2018) proposent une solution pour le recalage d'images multi-spectrales. Ils ne répondent cependant pas à la question de la combinaison des différents canaux.

Dans la section suivante, nous définissons la notion de gradient et l'étendons au cas multivarié pour introduire le *full-vector gradient* (FVG) qui est générique. La section 1.2 présente l'état de l'art en



(a) Approche FVG



(b) Approches supposant l'orthogonalité

FIGURE 1.2 – **Différence d'approche entre les différents gradients**. Le *full-vector gradient* (FVG) s'affranchit des caractéristiques du capteur contrairement aux approches supposant que l'espace couleur utilisé est orthogonal.

regard de la méthode proposée dans la section 1.1. Ensuite, nous étudions les résultats associés à ce nouveau gradient. L'impact des caractéristiques du capteur est montré dans la section 1.3.3, comparant des gradients calculés sur un capteur couleur, multi- et hyper-spectral simulant la même scène. Puis nous ouvrons la proposition en l'appliquant au système visuel humain. Nous étudions ensuite le gain objectif de cette proposition sur une base de données artificielles. Nous terminons l'analyse du FVG avec quelques résultats subjectifs sur des images réelles. La dernière section synthétise et conclut ce chapitre sur les gradients.

1.1 Qu'est-ce qu'un gradient ?

Qu'est ce qu'un gradient : comment le définir autrement que par sa définition mathématique ? Il semble que cette question pose un problème aux auteurs. Jähne (2005) ne répond pas à cette question, mais traite de l'usage du gradient dans le contexte de l'extraction des discontinuités (*edges*) entre parties d'une image. Dans ce contexte, il développe la question d'abord sous l'angle des différences avant d'arriver à la formulation du gradient. Il est intéressant de noter que, pour lui, plusieurs définitions sont possibles dans le cas multivalué. Il en va de même pour Pratt (2001), qui développe de façon approfondie les étapes de calcul numérique du gradient à partir d'une image en niveaux de gris. Pour Horaud and Monga (1995), le gradient n'est évoqué que pour son apport à la détection des contours. Dans cette introduction, Horaud et Monga ne considèrent que l'aspect spatial du gradient, dont "*le vecteur se trouve dans le plan image et est perpendiculaire à la direction du contour*". Même s'ils précisent que le calcul peut s'établir pour des images de dimension quelconque. Gonzalez et al. (2001), quant à eux, n'abordent le gradient que pour l'amélioration de l'image, sans définir ce qu'il représente.

Soille (2002), suivant le travail effectué avec Rivest et al. (1993) dans le contexte de la morphologie mathématique, ramène la question à l'extraction des limites entre objets définis par des bords plus ou moins nets. Les objets composant l'image et le fond se caractérisant par des valeurs plus ou moins homogènes. Soille indique typiquement l'analogie qui peut être faite entre la norme du gra-

dient et le gradient morphologique dans un contexte continu.

Enfin, pour enrichir cette liste non-exhaustive des livres de référence du traitement d'images par une étude couleur, Trémeau et al. (2004) développent les gradients pour l'extraction de contours dans les images couleur. Ceux-ci ne sont pas définis autrement que par leur expression mathématique. Plusieurs solutions y sont proposées pour prendre en compte la couleur dans le calcul du gradient, en particulier l'approche proposée par Di Zenzo.

Nonobstant, aucun des ouvrages cités ne répond à la question. Hasardons nous à une définition générique : le gradient est un vecteur établi en chaque point de l'image mesurant les variations locales et pouvant indiquer l'amplitude ou les directions de changement amenant la notion de tangence.

Dans un contexte couleur, le gradient permet d'extraire une direction spatiale d'amplitude maximale, mais aussi l'orientation couleur de cette variation d'un rouge vers le bleu par exemple. Lorsque le point considéré est dans une zone plus ou moins uniforme, l'amplitude considérée est faible et les directions quasi-aléatoires. Mais lorsque le point appartient à une zone de rupture, une cohérence apparaît sur ces trois mesures.

Le gradient est trop souvent confondu avec la notion de bords d'un objet (*edge*). Ce qui est considéré comme un bord dans une image correspond à des lignes continues ayant un gradient de haute amplitude, à la différence du bruit blanc, qui peut aussi être défini comme des pixels à gradient de haute amplitude mais dont la distribution spatiale n'a pas de cohérence. De par sa liaison directe avec le contenu spatio-fréquentiel de l'image, le gradient ne se limite pas à l'analyse ou l'extraction des contours. Il est également au cœur de l'analyse de l'aspect non-uniforme (texturé) des régions de l'image (moyennes et hautes fréquences de faible intensité). Disposer d'une formulation générique et complète du gradient pour des images multivaluées (couleur, multi-, hyper-spectrales et $3D\ldots$) a donc des répercussions allant bien au-delà de la détection de contours ou de points d'intérêt. Le *full-vector gradient* répond à cette exigence.

Nous nous intéressons ici, au calcul du gradient, introduit en une dimension avant de l'étendre aux dimensions multiples. Comme nous travaillons en numérique, une adaptation est nécessaire, celle-ci sera présentée par la suite.

1.1.1 En $1D$

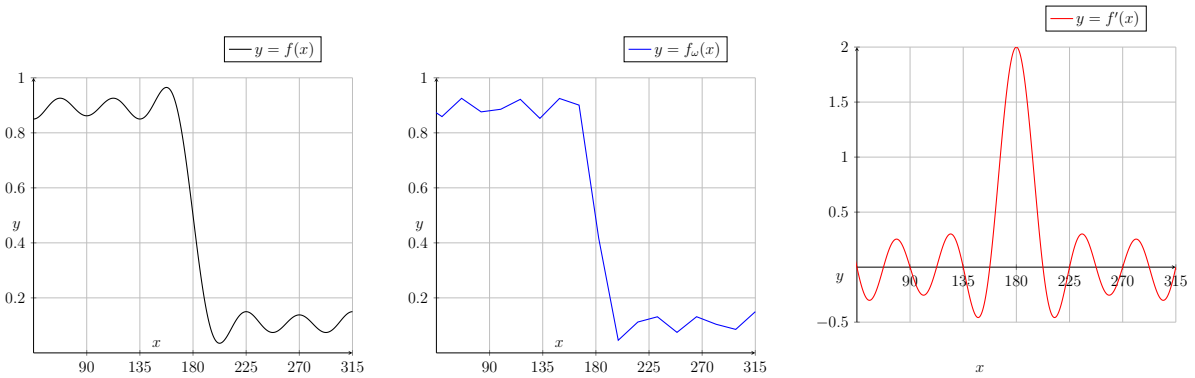
En $1D$, le gradient est la dérivée d'une fonction ou d'un signal selon l'axe de propagation de celle ou celui-ci. La figure 1.3 met en valeur un des problèmes du traitement d'image pour l'extraction de gradient. Nous avons au premier abord, une scène ou un signal continu (fig. 1.3a). Celui-ci est mesuré, ce qui implique une discrétisation (échantillonnage + quantification) de l'information (fig. 1.3b). Lors du calcul (fig. 1.3c), notre objectif est d'obtenir le gradient du signal initial et non celui du signal discrétisé.

La définition mathématique de la dérivée selon l'axe x est :

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}, \tag{1.1}$$

où $f(x)$ est le signal et $f'(x)$ sa dérivée. La définition de la limite implique que l'élément h soit petit.

Cette définition nécessite une fonction de classe C_1. Cette hypothèse est vérifiée dans le cas du signal de la figure 1.3a et plus généralement pour les scènes du monde physique qui sont continues. Mais elle ne l'est pas pour la figure 1.3b à cause de l'échantillonnage et de la quantification des images signaux initiaux. Une méthode de résolution sera discutée dans la sous-section 1.1.3.



(a) Signal réel (b) Signal mesuré (c) Signal gradient désiré

FIGURE 1.3 – **Exemple de gradient de fonction** $1D$. Nous trouvons respectivement, le signal, sa mesure numérique et la dérivée que l'on souhaite obtenir. Cela illustre la notion de gradient et un des problèmes liés au numérique.

1.1.2 Vers la nD

Quand il y a plus d'une direction, une dérivée devient une différentielle et dépend de la direction choisie :

$$\lim_{\|h\|_2 \to 0} \frac{\|I(p+h) - I(p) - dI(p) \cdot h\|_2}{\|h\|_2} = 0, \tag{1.2}$$

où $I : E \to F$ est une fonction (E et F deux espaces vectoriels) et dI la différentielle avec $p, h \in E$. La limite est toujours présente, mais h peut prendre une infinité de directions.

Quand h est orienté sur un vecteur de la base, la différentielle obtenue est appelée dérivée partielle. Un vecteur constitué de toutes les dérivées partielles est appelé gradient. Ainsi, appliqué à une image couleur $2D$ ($I(x, c)$: carte en $x \in \mathbb{R}^2$ et $c \in \mathbb{R}^2$), nous devrions obtenir un gradient comme celui-ci :

$$\nabla I(x, c) = \begin{pmatrix} \frac{\partial I(x,c)}{\partial x_1} \\ \frac{\partial I(x,c)}{\partial x_2} \\ \frac{\partial I(x,c)}{\partial c_1} \\ \frac{\partial I(x,c)}{\partial c_2} \\ \frac{\partial I(x,c)}{\partial c_3} \end{pmatrix}. \tag{1.3}$$

Le vecteur $\nabla I(x, c)$ est composé de deux parties. La première porte sur l'information spatiale $\left(\frac{\partial I(x,c)}{\partial x_i}\right)$ et la seconde pour la couleur $\left(\frac{\partial I(x,c)}{\partial c_i}\right)$. Les deux premières dérivées partielles contiennent l'information spatiale du gradient. Les trois dernières analysent l'impact d'une petite variation de couleur sur la localisation des pixels. Dans la suite de ce chapitre, nous nous concentrerons sur l'analyse du “gradient spatial”.

1.1.2.1 Formulation de Di Zenzo

Di Zenzo (1986) propose d'analyser les caractéristiques spatiales du gradient en exploitant la ma-

trice d'auto-corrélation spatiale :

$$ACor = \begin{pmatrix} \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle_2 & \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \\ \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 & \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \end{pmatrix}. \tag{1.4}$$

Appliquée à une image niveau de gris, nous pouvons analyser les valeurs et vecteurs propres directement pour obtenir la direction principale et l'intensité du gradient. Le produit scalaire complet en nD est donné par :

$$\left\langle \frac{\partial I(x,c)}{\partial x_i}, \frac{\partial I(x,c)}{\partial x_j} \right\rangle_2 = \sum_{j=1}^{n} \sum_{k=1}^{n} \left\langle \frac{\partial I(x,c_j)}{\partial x_i}, \frac{\partial I(x,c_k)}{\partial x_j} \right\rangle_2. \tag{1.5}$$

Avec une base orthogonale, tous les produits croisés sont nuls, ce qui simplifie le calcul.

1.1.2.2 Formulation du *Full-vector gradient*

En réalité les canaux couleur ou spectraux sont corrélés. En effet, leurs courbes de sensibilité spectrale (SSF pour *spectral sensitivity functions*) se chevauchent. Nous proposons d'intégrer une matrice de Gram dans le produit scalaire qui permet de prendre en compte cette non-orthogonalité :

$$\langle I(x), I(y) \rangle = I(x)^T \cdot Gm \cdot I(y), \tag{1.6}$$

$$\text{avec } Gm = \begin{pmatrix} \|S_0\|_2^2 & \langle S_0, S_1 \rangle_2 & \dots & \langle S_0, S_m \rangle_2 \\ \langle S_1, S_0 \rangle_2 & \|S_1\|_2^2 & \dots & \langle S_1, S_m \rangle_2 \\ \vdots & \vdots & \ddots & \vdots \\ \langle S_m, S_0 \rangle_2 & \dots & \langle S_m, S_{m-1} \rangle_2 & \|S_m\|_2^2 \end{pmatrix}. \tag{1.7}$$

Habituellement, la matrice de Gram Gm utilise un produit scalaire défini dans une base orthogonale. Il n'existe pas de base orthogonale connue dans l'espace d'acquisition des capteurs. Les courbes de sensibilité spectrale qui caractérisent les capteurs sont des fonctions semi-définies positives sur un support compact. Ce sont donc des mesures. Le produit scalaire associé aux fonctions intégrables définies sur $\mathbb{R}$ est :

$$\langle S_i, S_j \rangle_2 = \int_{\mathbb{R}} S_i(\lambda) S_j(\lambda) d\lambda. \tag{1.8}$$

C'est le produit scalaire que nous utiliserons dans la matrice Gm. Cela implique une matrice de Gram associée et fixée pour chaque capteur.

Si le produit scalaire entre deux SSF est nul, la matrice Gm devient diagonale. Dans ce cas, l'équation (1.6) correspond à l'expression présentée par Di Zenzo (1986) et Koschan and Abidi (2005).

1.1.3 Mise en œuvre informatique

Dans la sous-section précédente, la formulation est développée pour un monde continu (fig. 1.3), où la condition de classe C_1 est respectée. Dans une approche numérique, nous devons nous adapter à un contexte discret.

Dans ce cas, le pas h de l'équation (1.2) doit être entier. La condition limite étant toujours présente, il doit être petit. Pour $h = 1$, Roberts (1963) propose :

$$\begin{aligned} \Delta_{x_1} I(x) &= I(x_1 + 1, x_2) - I(x_1, x_2), \\ \Delta_{x_2} I(x) &= I(x_1, x_2 + 1) - I(x_1, x_2). \end{aligned} \tag{1.9}$$

Pour éviter l'asymétrie induite par l'expression précédente, des valeurs paires de h sont préférées. Par exemple, le cas où $h = 2$:

$$\begin{aligned} \Delta_{x_1} I(x) &= I(x_1 + 1, x_2) - I(x_1 - 1, x_2), \\ \Delta_{x_2} I(x) &= I(x_1, x_2 + 1) - I(x_1, x_2 - 1). \end{aligned} \tag{1.10}$$

Comme mentionné précédemment, une image discrète n'a pas de dérivée continue. Ce problème est classiquement résolu en appliquant un filtre passe-bas G avant le calcul du gradient. C'est le cas de Canny (1986) qui filtre (eq. (1.11)) les images avant de calculer les dérivées.

$$G(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-\|x\|^2}{2\sigma^2}}. \tag{1.11}$$

Le filtre peut être appliqué en même temps que le calcul du gradient, comme le font Prewitt (1970); Sobel and Feldman (1968). Ces méthodes introduisent un filtrage passe-bas orthogonal à la dérivée partielle calculée. Prewitt (1970) effectue une moyenne, quand Sobel and Feldman (1968) introduisent un filtre binomial :

$$\begin{aligned} \frac{\partial I(x)}{\partial x_1} &= \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} * I(x), \\ \text{et } \frac{\partial I(x)}{\partial x_2} &= \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} * I(x). \end{aligned} \tag{1.12}$$

D'une façon plus générale, l'utilisation des fonctions Gaussiennes et dérivées de Gaussiennes permet l'approximation, à une taille quelconque, des filtres moyenneur et dérivateur pour débruiter (fig. (1.13)). Les filtres de Sobel and Feldman (1968) représentent alors une approximation entière et compacte de ce principe par le biais des filtres binomiaux.

$$\frac{\partial Filter}{\partial x_i} = \begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix} \times \begin{pmatrix} \\ \\ \\ \\ \\ \end{pmatrix}^T . \tag{1.13}$$

La Gaussienne est centrée sur le pixel dont nous estimons la dérivée. L'estimation de son support peut être faite en fonction de l'écart-type σ sélectionné : $6\sigma + 1 \times 6\sigma + 1$.

Une fois les dérivées partielles obtenues, elles sont combinées grâce à la matrice de Gram introduite précédemment (eq. (1.7)) et nous obtenons la matrice d'auto-corrélation spatiale présentée dans l'équation (1.4). Cette matrice d'auto-corrélation nous permet d'extraire deux informations : la norme et l'angle du gradient, à partir des valeurs propres de la matrice. Celles-ci sont données par λ^+ et λ^- :

$$\begin{aligned} \lambda^+ = \frac{1}{2}\Bigg(&\left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle_2 + \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \\ &+ \sqrt{\left(\left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle_2 - \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \right)^2 + 4 \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2^2} \Bigg), \end{aligned} \tag{1.14}$$

$$\begin{aligned} \lambda^- = \frac{1}{2}\Bigg(&\left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle_2 + \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \\ &- \sqrt{\left(\left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle_2 - \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \right)^2 + 4 \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2^2} \Bigg). \end{aligned} \tag{1.15}$$

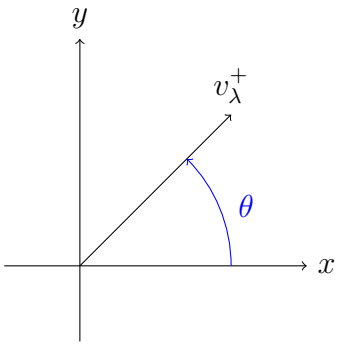


FIGURE 1.4 – **Angle estimé par le gradient de Di Zenzo** par rapport aux axes de l'image.

La mesure de l'angle a été initialement proposée par Di Zenzo (1986) à $\pm\frac{\pi}{2}$ près. Jin et al. (2012) ont levé cette imprécision :

$$\theta = \begin{cases} \text{Indéfini si } \left(\left\| \frac{\partial I(x,c)}{\partial x_1} \right\|_2^2 - \left\| \frac{\partial I(x,c)}{\partial x_2} \right\|_2^2 \right)^2 + (\tau)^2 = 0 \\ \text{sign}\,(\tau) \arcsin \left(\dfrac{\lambda_+ - \left\| \frac{\partial I(x,c)}{\partial x_1} \right\|_2^2}{2\lambda_+ - \left\| \frac{\partial I(x,c)}{\partial x_1} \right\|_2^2 - \left\| \frac{\partial I(x,c)}{\partial x_2} \right\|_2^2} \right) \\ \text{avec } \tau = \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle_2 \text{ sinon.} \end{cases} \tag{1.16}$$

Elle correspond à l'angle spatial entre l'axe initial et la direction du vecteur propre associé à la valeur propre maximale. La figure 1.4 représente l'angle estimé.

Le calcul de la norme a pris différentes formes au cours des articles. Koschan and Abidi (2005) propose de retenir la valeur propre maximale comme norme du gradient λ^+ donné dans l'équation (1.15).

Au contraire Shi and Tomasi (1993) utilisent la valeur propre minimale. Cependant, leur problématique est définie pour la détection de points d'intérêt. Sapiro (1997) va plus loin en impliquant les deux valeurs propres. Une de ces propositions est une fonction de λ^+ et λ^- :

$$\begin{aligned} Mag &= (\lambda^+ - \lambda^-)^{\frac{1}{p}} \\ &= \left(\left(\left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle - \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle \right)^2 + 4 \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle^2 \right)^{\frac{1}{p}}, \end{aligned} \tag{1.17}$$

pour $p > 0$. Dans la suite du chapitre, nous garderons cette expression avec $p = 2$.

1.2 En regard de l'état de l'art

Beaucoup d'articles ayant une vue d'ensemble sur le calcul du gradient couleur ou multi-spectral ont été publiés, y compris ces dernières années James (2016); Subhasini and Singh (2017). Être exhaustif est impossible. Nous allons présenter ici un état de l'art avec un point de vue orienté vers la métrologie. L'objectif est de progresser vers la validation physique de la proposition. Avec cet objectif, nous avons organisé les entrées bibliographiques en fonction du domaine où est calculé le gradient (fig. 1.5). L'extraction du gradient est beaucoup moins développée dans les domaines multi- et hyperspectral, c'est pour cela que cet état de l'art explore principalement le domaine couleur.

Si la notion de gradient est générique dans des espaces n-dimensionnels, elle est cependant dépendante du domaine dans lequel elle est exprimé. Deux grands domaines sont utilisés pour l'expression du gradient.

Le domaine perceptuel est celui qui a le plus attiré les chercheurs et la production scientifique. Ce domaine impose l'estimation de gradient en concordance avec le système visuel humain (HVS pour *Human Visual System*). Nous supposons, dans ce domaine, que les données de calibrage du processus sont bien connues pour éviter tout biais de calcul.

De façon complémentaire, le domaine physique ou capteur cherche une expression au plus prés du contenu physique, sans les non-linéarités induites par de HVS. Dans ce domaine, une image est considérée acquise en RGB ou à l'aide d'un capteur multi- ou hyper-spectral.

D'un point de vue métrologique, la première approche s'intéresse à la perception de l'objet par un observateur, alors que la seconde s'intéresse uniquement à la surface de l'objet d'intérêt. Les deux se combinent dans l'analyse de l'apparence des matériaux comme le font Huang et al. (2012).

Nous avons défini une troisième catégorie d'espaces, appelée espaces transformés. Ils sont apparus au cours des années, et sont généralement liés à l'espace RGB d'entrée. Ils ne nécessitent aucune donnée de calibrage. Leur objectif est d'imiter la perception humaine.

Nous présenterons tout d'abord les approches du domaine physique puisqu'elles sont les plus directes à mettre en œuvre. Étant construite depuis l'espace RGB avec un but perceptuel, la deuxième catégorie présentée sera les espaces transformés. Enfin, les propositions du domaine perceptuel seront introduites. Nous conclurons sur l'insertion de notre proposition dans ce schéma.

1.2.1 Domaine physique

Pour utiliser l'information couleur, certains auteurs Chen and Chen (2010); Shivakumara et al. (2013) combinent de manière empirique les gradients marginaux pour estimer une norme de gradient couleur. Par exemple, Tarabalka et al. (2010) proposent de sommer les gradients marginaux ou d'approximer les gradients par une distance entre différences de vecteurs. Di Zenzo (1986) propose de prendre en compte la couleur avant d'utiliser une matrice de corrélation spatiale pour mesurer le gradient. Comme indiqué dans la section précédente, il suppose implicitement que l'espace dans lequel les gradients sont calculés est orthogonal. Cumani (1991) présente une extension au multi-spectral de l'approche de Di Zenzo. Cependant, il ne prend pas en compte la corrélation entre les canaux comme présentée dans les sections précédentes. L'hypothèse de validité est donc elle aussi basée sur l'orthogonalité des canaux. Kimmel et al. (1997) proposent de projeter les données sur une surface de Riemann connaissant la métrique associée à ces données. Il obtient une matrice de Gram et s'interroge sur le choix des coefficients à appliquer dans la matrice.

Rivest et al. (1993) ont introduit la morphologie mathématique en niveaux de gris et défini le gradient de Beucher comme étant la différence de deux résultats morphologiques. Dès sa création, la morphologie mathématique s'inscrit dans un contexte métrologique et pourrait répondre à nos besoins. Mais si la morphologie a été définie dans un but métrologique, son extension aux dimensions multiples n'est pas aisée. Cela demande de définir un ordre dans un espace nD. Une approche basée sur la médiane vectorielle dans le domaine couleur a été proposée par Astola et al. (1990). Pour résoudre la difficulté d'ordonnancement en nD, de nombreux auteurs (Lezoray et al. (2008); Ledoux et al. (2012); Shih and Liu (2016); Deborah et al. (2018)) présentent la nécessité de choisir des références. Aujourd'hui la sélection automatique de références est toujours une question ouverte malgré certaines propositions Shih and Liu (2016).

1.2.2 Espace transformé

La première proposition Carron and Lambert (1994) estime les gradients dans l'espace couleur HLS. La construction empirique de Carron and Lambert (1994) favorise l'information luminosité ou teinte en fonction de la saturation. Leur proposition est présentée plus en détail en annexe A.

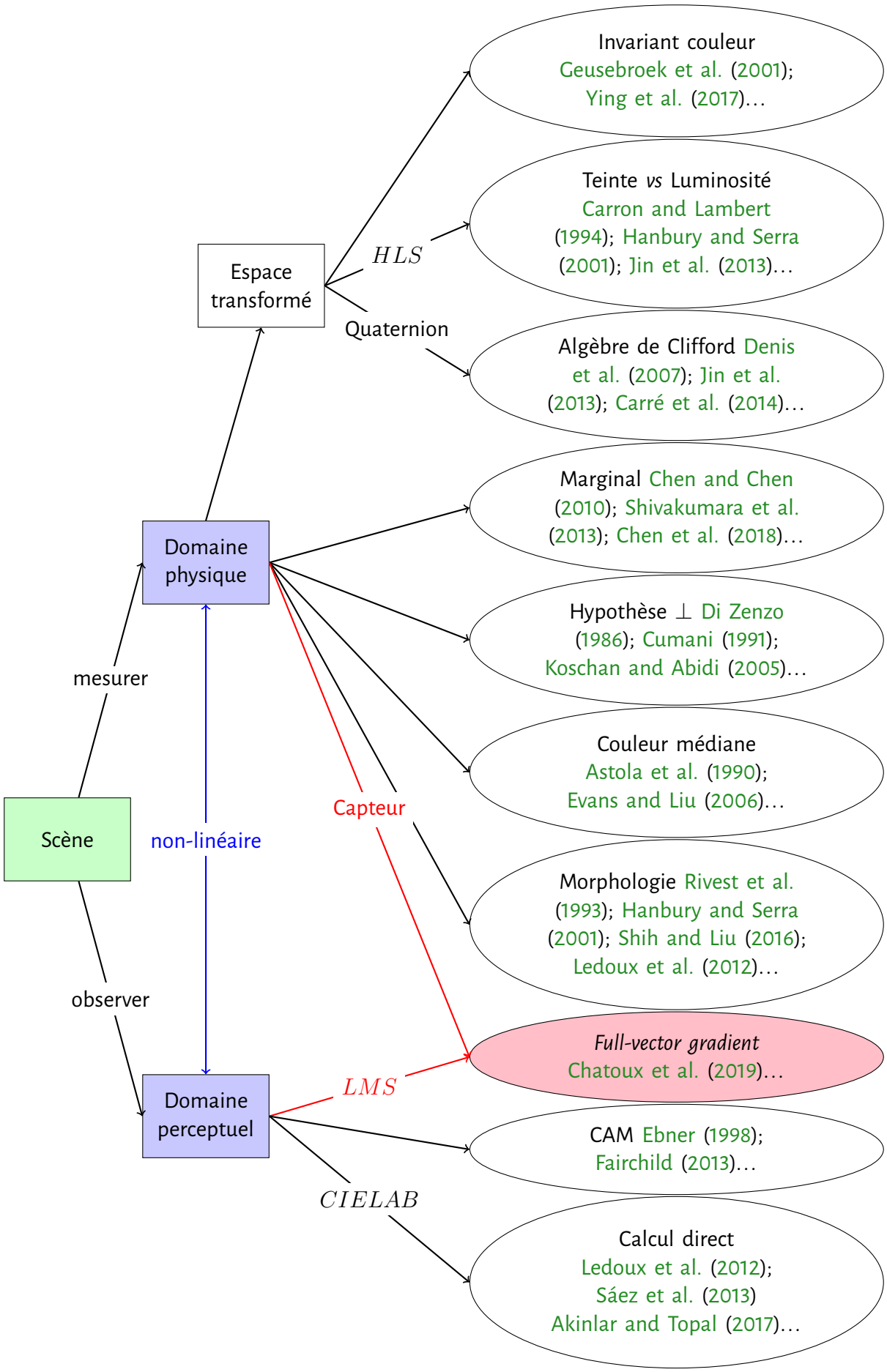


FIGURE 1.5 – **État de l'art des gradients couleur** organisés en fonction de l'espace dans lequel ils sont calculés.

Le système visuel humain s'adapte en présence d'ombres ou de spécularités. C'est-à-dire que même si la couleur varie en terme de mesure, l'œil la traduit comme invariée. Cela a mené à une grande famille calculant des invariants couleur visant à reproduire ce phénomène. Geusebroek et al. (2001) proposent d'estimer un gradient à l'aide de ces invariants couleur. Plus récemment, Ying et al. (2017) utilisent directement ces invariants dans l'espace RGB.

Travailler dans le domaine de Fourier est très attractif pour les images en intensité. En suivant cette idée, l'algèbre de Clifford, utilisée par Denis et al. (2007); Jin et al. (2013); Carré et al. (2014), offre la possibilité d'utiliser les quaternions pour l'estimation du gradient. Ces espaces couleur sont généralement des rotations depuis l'espace RGB initial. La non-orthogonalité est donc préservée.

1.2.3 Domaine perceptuel

Depuis 1913, la Commission Internationale de l'Éclairage (CIE) a conduit des expériences pour concevoir des espaces perceptuels. $CIELAB$ est un espace créé pour estimer les différences couleur selon la vision humaine. Depuis 1976, plusieurs différences couleur ΔE on été proposées. Sáez et al. (2013) comparent ces différentes fonctions de ΔE pour une tâche de calcul de gradient. Akinlar and Topal (2017) utilisent l'expression de Di Zenzo dans l'espace $CIELAB$ qui est orthogonal par construction. Cependant, les adaptations récentes des estimations de différences couleur s'éloignent du propos original. En effet, le ΔE initial était une simple mesure Euclidienne, les plus récentes (ΔE_{94}, ΔE_{2000}, etc.) en sont loin et ne sont précises que pour des faibles différences couleur. Cette contrainte est rarement vérifiée dans le cas d'un calcul de gradient couleur.

Pour aller plus loin, les modèles d'apparence couleur (CAM pour *Color Appearance Models*) sont conçus pour embarquer des modèles plus complets du HVS en prenant en compte l'arrière-plan, et l'impact de l'environnement sur la perception locale des éléments visuels. Ces travaux sont essentiels pour mieux comprendre le système visuel humain mais ils ne sont pas adaptés à un objectif métrologique. Pour être métrologiquement valides, ces approches demandent identification, caractérisation et calibrage des éléments de la chaîne d'acquisition, ce sont des connaissances impératives pour un traitement métrologique. Le livre de Fairchild (2013) présente en détail de nombreux modèles.

Les gradients sont calculés depuis l'image afin de prendre des décisions à propos de la scène physique acquise. Pour respecter les contraintes métrologiques, le gradient ne doit être lié qu'au contenu de la scène et non au produit scène-capteur. Dans cette optique, nous avons proposé le *full-vector gradient* dans la sous-section 1.1.2.2. C'est un lien entre l'expression de Di Zenzo et une construction générique établie dans l'espace capteur (couleur ou spectral). En embarquant les courbes de sensibilité spectrale du capteur, l'expression permet de calculer un gradient quel que soit son nombre de canaux ou ses SSF. Ainsi, les gradients estimés par différents capteurs peuvent être comparés (sec. 1.3.3.1). Dans la sous-section 1.3.3.2, nous nous interrogeons sur l'intérêt de calculer ce gradient dans un espace perceptuel : LMS.

1.3 Résultats

Au-delà de ces références, une question récurrente est toujours présente sans solution complète : comment évaluer la qualité, la précision, la robustesse d'une estimation de gradient ? Cette question induit deux sous-questions : Quelles images pour l'évaluation et quelle mesure objective pour comparer les gradients ? Généralement cette question est associée à l'évaluation d'une segmentation. Panetta et al. (2016) proposent différents éléments de mesure : localisation des bords et coins, largeur du bord. Ils s'appuient sur une image de bord de référence. Magnier et al. (2018) évaluent la qualité de détection de bords avec une vérité-terrain créée par un observateur humain. La variabilité entre

observateurs peut être forte comme le montrent Menze et al. (2015) dans le domaine médical. Pour pallier cette variabilité, Fernandez et al. (2015) évaluent la qualité de la segmentation à l'aide d'une tâche de plus haut niveau (classification). Dans tous ces cas, les résultats se limitent à l'analyse des images et sont donc dépendants de leur contenus et de la tâche à accomplir. Sous l'hypothèse que le gradient couleur est lié à la différence entre des distributions couleur ou spectrales, nous proposons de construire une base de données artificielles. Les images sont générées mathématiquement, pour évaluer objectivement les performances du calcul d'un gradient en fonction de la difficulté de séparer deux distributions.

1.3.1 Base de données artificielles

Dans cette section, nous comparons la précision d'extraction d'un bord sur des images synthétiques. Nous formulons l'hypothèse que plus les distributions des régions à discriminer se chevauchent plus la tâche est difficile. Notre objectif est donc d'évaluer la capacité des gradients à extraire les bords lorsque les distributions de l'objet et du fond se chevauchent. La figure 1.6 présente des exemples d'images dont les distributions des deux textures naturelles s'intersectent ou non. Sur la figure 1.6c, nous remarquons que la séparation des deux textures est plus difficile en bas à gauche du cercle où les distributions se chevauchent plus qu'en haut à droite.

Pour maîtriser ce chevauchement, nous créons des distributions Gaussiennes $3D$ (aussi nommées formes ellipsoïdales). Ensuite, nous déterminons où les positionner dans l'espace pour établir le taux de séparabilité. Il représente le taux de chevauchement des ellipsoïdes et sera le critère utilisé pour mesurer la difficulté d'extraction des bords.

Nous étudierons les résultats de deux expériences sur cette base de données après l'avoir présentée. La première expérience évalue l'impact de la forme des distributions sur les performances de détection de bords. La seconde évalue les performances globales.



(a) Sans chevauchement (b) Avec chevauchement (c) Avec chevauchement

FIGURE 1.6 – **Exemples de distributions naturelles** avec chevauchement des distributions. Il est visuellement plus facile de discerner les deux textures de la figure 1.6a que celles des figures 1.6b et 1.6c.

1.3.1.1 Construction de la base de données

Afin de mesurer l'apport de nos contributions, nous avons mis en place une base de données synthétique. Elle permet d'évaluer la qualité de notre gradient et du détecteur associé. Bien que notre proposition soit générique, le manque de résultat de référence en dehors du visible dans le domaine multi- ou hyper-spectral nous a conduit à effectuer cette évaluation dans le domaine couleur. De plus,

la comparaison de gradient couleur est assez prolifique dans la littérature pour pouvoir nous y comparer.

Nous présenterons comment les distributions couleurs sont créés, puis la mesure du taux de séparabilité. Ce dernier évalue la difficulté d'extraire un bord. Grâce à celui-ci et aux formes de distributions couleurs nous avons une base de données synthétique. Enfin, nous présenterons le critère de comparaison basé sur la classification.

Magnier et al. (2018) proposent une méthode d'évaluation des gradients niveaux de gris sur des images réelles. La vérité-terrain dépendra toujours de l'humain qui a crée la vérité-terrain. Dans le domaine médical, de nombreux articles présentent les désaccords qu'il peut y avoir parmi les experts. Dans le domaine du patrimoine culturel, ces désaccords sont aussi fréquents. C'est pour ces raisons que nous avons choisi de construire une base de données synthétiques.

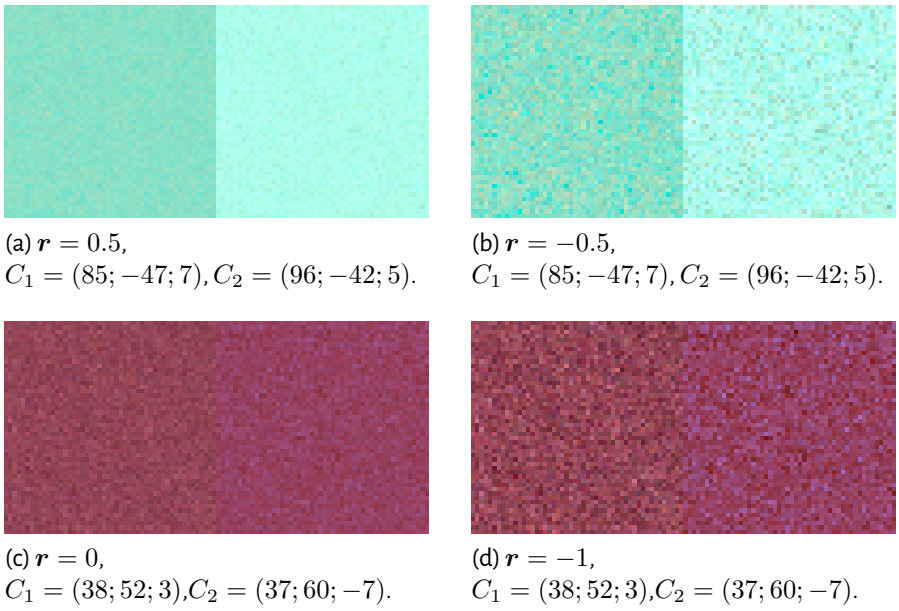
Nous partons du principe que la détection de régions est facile si les deux distributions sont uniformes même si les couleurs sont proches. Cependant, la tâche se complique si les régions ne sont pas parfaitement uniformes. Dans ce cas, nous supposons que chaque région est définie par une distribution couleur. La complexité de l'extraction du gradient dépend de l'intersection entre les distributions couleur. Plus les distributions couleur se chevauchent, plus la segmentation devient difficile.

Le protocole que nous proposons est basé sur deux distributions couleur générées avec un taux de séparabilité, des formes de distributions et des couleurs moyennes donnés (figure 1.7).

Formes des distributions

La distribution couleur est générée avec un modèle Gaussien multivarié $\mathcal{N}_3(\mu, \Sigma)$ dans un espace couleur donné ($CIERGB$ ou $CIELAB$ par exemple). Pour éviter le biais induit par la quantification, tous nos calculs sont faits avec une précision de type *double*.

La forme de la Gaussienne multivariée est contrôlée par la matrice de variance covariance Σ.



(a) $\boldsymbol{r} = 0.5$,
$C_1 = (85; -47; 7), C_2 = (96; -42; 5)$.

(b) $\boldsymbol{r} = -0.5$,
$C_1 = (85; -47; 7), C_2 = (96; -42; 5)$.

(c) $\boldsymbol{r} = 0$,
$C_1 = (38; 52; 3), C_2 = (37; 60; -7)$.

(d) $\boldsymbol{r} = -1$,
$C_1 = (38; 52; 3), C_2 = (37; 60; -7)$.

FIGURE 1.7 – **Exemples d'images générées**. Les valeurs du taux de séparabilité (r) et la localisation des paires de couleur sont données par images. Pour percevoir les différences de couleur, la distance entre les centres des distributions est fixé à 12 dans l'espace $CIELAB$.

Celle-ci dépend d'une matrice de rotation R et d'une matrice de forme S (eq. 1.18).

$$\Sigma = R^T S R, \tag{1.18}$$

$$\text{avec } S = \begin{pmatrix} a & 0 & 0 \\ 0 & \frac{1}{a} & 0 \\ 0 & 0 & \frac{1}{a^2} \end{pmatrix}. \tag{1.19}$$

Le paramètre $a \in [1;5]$ détermine la forme de la distribution, sphérique lorsque a est proche de 1, plus oblongue lorsque a s'approche de 5 (eq. 1.19). Ce paramètre est choisi aléatoirement. La matrice R est une matrice de rotation classique définie par le produit de trois matrices définissant chacune une rotation d'angle θ_i autour d'un axe principal (eq. 1.20). Les angles sont aussi choisis aléatoirement.

$$R = R_1 R_2 R_3, \tag{1.20}$$

$$\text{avec } R_1 = \begin{pmatrix} \cos(\theta_1) & -\sin(\theta_1) & 0 \\ \sin(\theta_1) & \cos(\theta_1) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{1.21}$$

$$\text{et } R_2 = \begin{pmatrix} \cos(\theta_2) & 0 & \sin(\theta_2) \\ 0 & 1 & 0 \\ -\sin(\theta_2) & 0 & \cos(\theta_2) \end{pmatrix}, \tag{1.22}$$

$$\text{et } R_3 = \begin{pmatrix} 1 & 0 & 0 \\ & cos(\theta_3) & -\sin(\theta_3) \\ 0 & \sin(\theta_3) & \cos(\theta_3) \end{pmatrix}. \tag{1.23}$$

La figure 1.8 montre quelques exemples de formes obtenues. A cette étape, toutes les distributions sont centrées ($\mathcal{N}_3(0, \Sigma)$). L'ellipsoïde représente l'enveloppe de la distribution à une distance de Mahalanobis de 3 du centre.

Ensuite, nous modifions la moyenne et ajoutons un paramètre α pour contrôler la localisation et le volume de la distribution. La matrice de variance covariance finale est définie par $\Sigma'_i = \alpha_i^2 \Sigma_i$. Cette étape permet de contrôler le chevauchement des distributions. Deux centres C_1 et C_2 sont sélectionnés aléatoirement dans l'espace $CIELAB$ pour être à une distance couleur ΔE de 3.

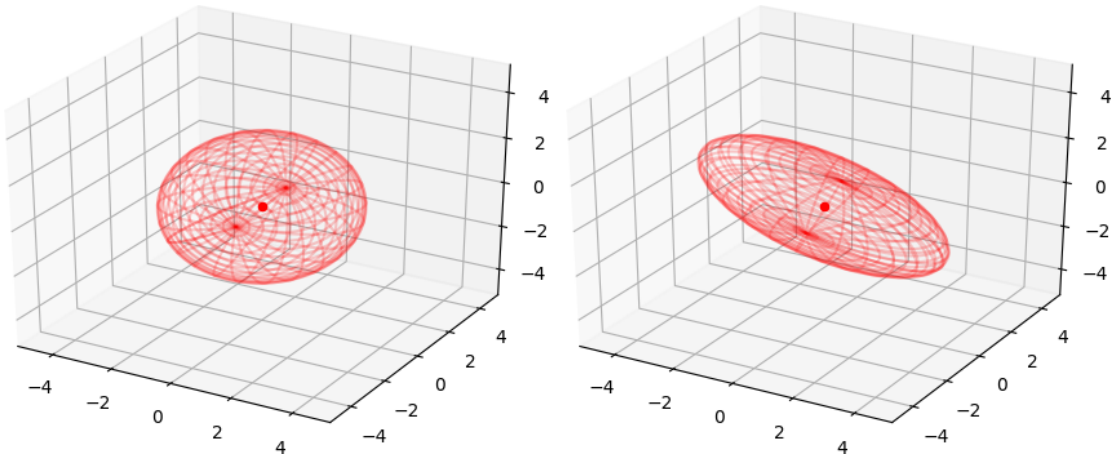
Taux de séparabilité

La complexité de détection du gradient est proportionnelle à l'intersection des ellipsoïdes relative aux distributions de Gaussienne $\mathcal{N}_3(C_1, \Sigma'_1)$ et $\mathcal{N}_3(C_2, \Sigma'_2)$.

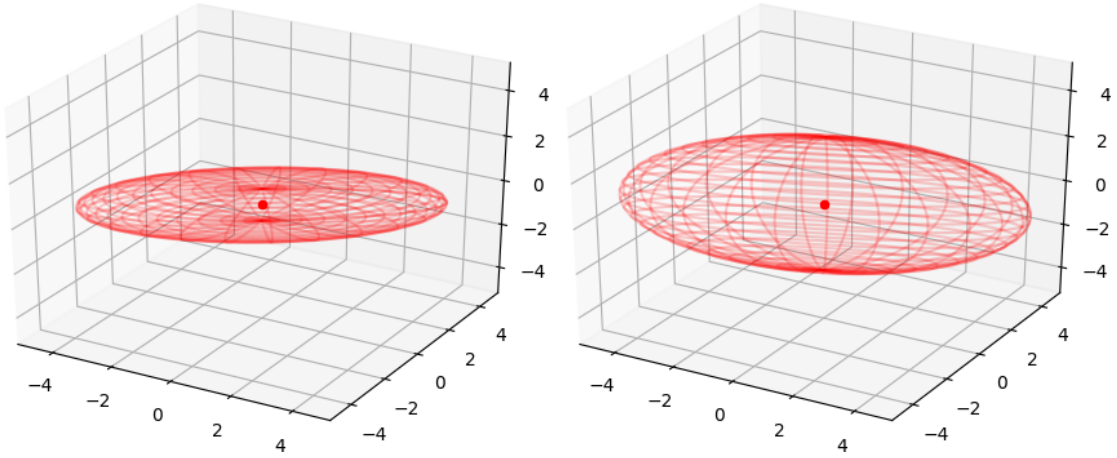
Nous proposons d'utiliser la distance algébrique normalisée entre deux enveloppes de distributions (eq. 1.24) pour définir le taux de séparabilité (r) comme illustré sur la figure 1.9. L'enveloppe de la distribution considérée est définie par la surface ellipsoïde à une distance de Mahalanobis de 3 unités de la moyenne de la distribution. La distance de 3 permet de contenir 99.7% des données à l'intérieur de l'enveloppe.

$$r = \frac{\overline{C'_1 C'_2}}{||C_1 C_2||_2} \quad \text{avec} \quad d_{\Sigma'_i}(C_i, C'_i) = 3, \quad \text{et} \quad \{C'_1, C'_2\} \in (C_1 C_2). \tag{1.24}$$

Un taux de séparabilité nul exprime un unique point d'intersection entre les deux ellipsoïdes. Puisque C'_i est à une distance de Mahalanobis de 3, les deux distributions ne se chevauchent que pour 0.3% des pixels couleur pour un taux de séparabilité nul. Une valeur négative indique une intersection plus importante, par opposition à une valeur positive qui ne montre aucune intersection



(a) $a = 1.0, \theta = (247; 49; 162)$ (b) $a = 1.8, \theta = (280; 298; 251)$



(c) $a = 3.0, \theta = (282; 117; 16)$ (d) $a = 3.9, \theta = (9; 110; 275)$

FIGURE 1.8 – **Exemples d'enveloppes de distributions**. L'enveloppe est à une distance de Mahalanobis de 3 du centre des distributions générées.

entre les deux enveloppes. Il peut cependant y avoir un léger chevauchement des distributions. La normalisation est relative à la distance entre les moyennes. Cela permet de comparer les résultats pour des distances aux centres différentes ou pour différents espaces couleur.

Une autre possibilité pour le taux de séparabilité aurait pu être la divergence de Kullback-Leibler. Cependant, la non-linéarité de sa formule rend la dynamique des résultats plus difficile à interpréter.

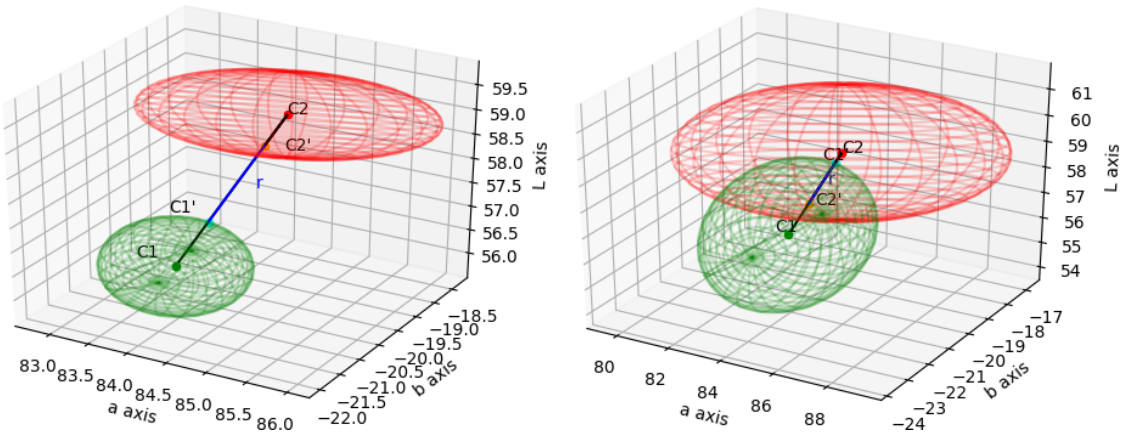
La base de données artificielles

Pour explorer un maximum du gamut [1] au cours de l'expérience, les paires de couleurs (C_1, C_2) sont choisies aléatoirement. Les coefficients multiplicateurs des distributions α_i sont calculés pour un taux de séparabilité r donné :

$$\alpha = \alpha_1 = \alpha_2 = \frac{(1-r)d_{\Sigma_1}(C_1, C_2)d_{\Sigma_2}(C_2, C_1)}{3(d_{\Sigma_1}(C_1, C_2) + d_{\Sigma_2}(C_2, C_1))}. \tag{1.25}$$

Le taux de séparabilité varie dans l'intervalle $[-1, +1]$. De plus, nous vérifions que les deux distributions couleur appartiennent au gamut de l'espace $CIERGB$. Les distributions couleur sont ensuite générées dans cet espace.

1. Gamut : ensemble des couleurs perceptibles ou reproductibles par un périphérique dans un espace de référence. Dans notre cas, le gamut correspond donc à l'ensemble de l'espace $CIELAB$.



(a) Distributions 1.8b (verte) et 1.8d (rouge) avec $r = 0.5$

(b) Distributions 1.8b (verte) et 1.8d (rouge) avec $r = -0.5$

FIGURE 1.9 – **Illustration du taux de séparabilité** (r) en fonction du chevauchement des deux distributions couleurs.

La figure 1.7 présente quatre images générées avec deux paires de couleurs (C_i, C_j).Pour faciliter la perception des différences, nous avons sélectionné un $\Delta E(C_i, C_j) = 12$. Les paramètres des distributions des figures 1.7a et 1.7b induisent une différence principalement sur la coordonnée luminance. Au contraire, les deux autres varient sur la composante teinte. Le taux de séparabilité sur la figure 1.7b (resp. 1.7d) est plus grand que sur la figure 1.7a (resp. 1.7c), car les distributions se chevauchent plus. Elles seront plus délicates à segmenter.

Critère de comparaison

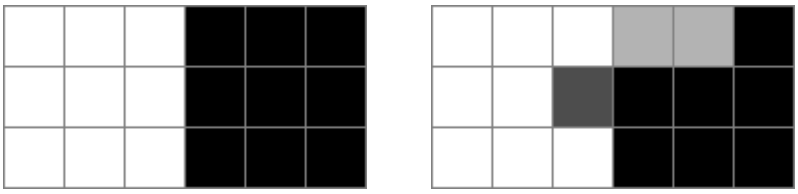
Pour évaluer la qualité d'un gradient, nous devons segmenter les images et utiliser un critère habituellement utilisé en classification. Chaque image test est composée de deux régions R_1 et R_2 (la figure 1.7 présente des images ($\mathcal{I}$) avec ces deux régions). Notre objectif est d'évaluer notre gestion de l'inter-dépendance des canaux, pas l'estimation de la direction. C'est pourquoi, nous nous sommes focalisés sur un bord vertical. Nous estimons le bord comme étant le maximum par ligne de l'estimation du gradient. Les vrais positifs ($\mathcal{TP}$), respectivement les vrais négatifs ($\mathcal{TN}$) sont les pixels affectés après détection de bord dans la région R_1, respectivement la région R_2. Les faux positifs ($\mathcal{FP}$) et les faux négatifs ($\mathcal{FN}$) correspondent aux pixels localisés du mauvais coté du bord après calcul. La figure 1.10 montre un résultat théorique de segmentation et un cas pratique pour deux régions 3×3. Nous utilisons le calcul de précision pour évaluer les performances :

$$Acc = \frac{\#\mathcal{TP} + \#\mathcal{TN}}{\#\mathcal{I}}, \tag{1.26}$$

Les résultats de précision sont entre 0% et 100%, 100% étant la réponse parfaite.

1.3.1.2 Impact de la forme de la distribution

Dans le monde réel, la forme des distributions couleur n'est pas toujours sphérique. Cette expérience évalue l'impact de la forme des distributions sur la détection de bords. Quatre formes ellipsoïdales permettent la création de 16 paires de distributions couleur. Ces distributions sont ensuite réparties sur l'espace couleur de manière à obtenir un $\Delta E = 3$ entre les moyennes des deux distributions comparées. La figure 1.11 présente la moyenne sur 10 tests pour chaque taux de séparabilité



(a) Segmentation théorique (b) Exemple de résultat pratique

FIGURE 1.10 – **Exemple de segmentation pratique comparé à la théorique**. Dans la figure 1.10b, le blanc correspond au vrai positif, le gris clair au faux positif, le gris foncé au faux négatif, et le noir au vrai négatif par comparaison avec 1.10a. Dans cet exemple, la précision est de $\frac{8+7}{18}$.

dans quatre cas représentatifs de paires des distributions couleur (les douze autres présentent des résultats similaires C).

Les résultats de la figure 1.11a évaluent l'extraction de bord sur des distributions presque sphériques, *i.e.* sans interdépendance entre les canaux. Comme attendu, l'approche de Di Zenzo présente de bons résultats, vient ensuite le FVG. Pour les seconds résultats, une des deux distributions est plus ellipsoïdale et orientée (fig. 1.11b). Dans ce cas, l'approche de Di Zenzo devient moins efficace. Ce résultat est accentué si la distribution est plus allongée (fig. 1.11c). La figure 1.11d illustre les résultats obtenus pour un cas plus générique : deux distributions ellipsoïdales et orientées. L'approche de Di Zenzo atteint ses limites dans ce cas.

1.3.1.3 Combinaison de formes de distributions couleur

Pour comparer les résultats à plus grande échelle, nous augmentons le nombre d'images à 500 pour chaque taux de séparabilité : cinquante paires de formes de distributions avec dix paires de couleurs par forme. Nous mesurons ensuite la moyenne de la précision pour les huit taux de séparabilité (tab. 1.1).

Le *full-vector gradient* réalise de meilleures performances que les autres gradients. Les différences sont moins importantes que celles observées dans la sous-section précédente. Cela est dû au paramètre aléatoire a qui crée les formes des distributions. La forme n'évolue pas linéairement de ronde à ellipsoïdale avec le paramètre a. Les formes plutôt rondes sont plus nombreuses que les formes vraiment ellipsoïdales parmi les 100 générées. Les résultats du gradient Sobel sont proches de ceux du *full-vector gradient*. Cela peut s'expliquer par les paires de couleurs orientées sur l'axe luminosité. Cette théorie est renforcée par les dires de Novak et Shafer Novak and Shafer (1987) (cité par exemple par Koschan (1995); Evans and Liu (2006); Wang and Shui (2016)) énonçant que 90% des bords couleur peuvent être analysés en niveaux de gris.

Pour illustrer les résultats de la table 1.1, nous présentons quelques cas d'extraction de bords pour des couleurs et taux de séparabilité différents (fig. 1.12). Chaque colonne présente les résultats de segmentation pour les quatre gradients. La première ligne présente les images à segmenter. La figure 1.12a présente la segmentation attendue.

La première colonne présente deux régions vertes, les variations se développent principalement sur un axe d'acquisition RGB. Cette configuration de distributions avantage le gradient de Di Zenzo. De plus, comme les variations sont majoritairement sur un axe, le gradient Sobel présente aussi une bonne précision.

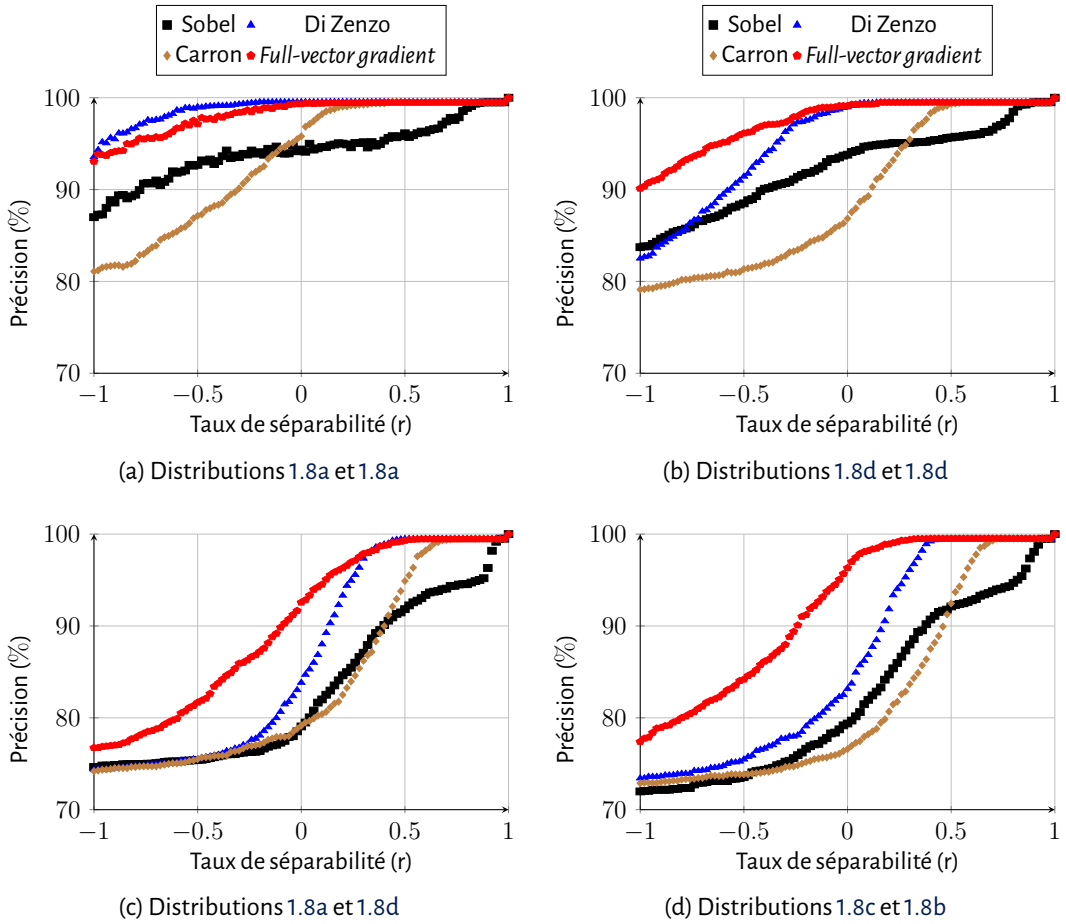
Dans le second cas, la couleur de la distribution est rose, autrement dit, du rouge peu saturé. Nous sommes donc proches de l'axe achromatique. Ici, une combinaison correcte des canaux est primordiale même si nous sommes dans un cas plus facile ($r = 0.5$). Cette nécessité explique les meilleurs résultats du full-vector gradient face aux autres.

TABLE 1.1 – **Moyennes et écarts-types pour des taux de séparabilité choisis**.

	Di Zenzo		FVG		Sobel		Carron	
r	Moy	E-T	Moy	E-T	Moy	E-T	Moy	E-T
-1	76.41	*0.14*	**79.41**	0.46	78.45	0.38	75.84	0.19
-0.5	78.57	0.30	**83.17**	0.62	81.36	0.61	76.97	*0.25*
-0.25	80.84	*0.47*	**86.22**	0.75	83.68	0.72	78.26	0.50
-0.15	82.17	0.66	**87.70**	0.77	84.75	0.84	78.96	*0.56*
0	84.78	*0.69*	**90.04**	0.81	86.58	0.96	80.48	0.79
0.15	88.02	*0.74*	**92.46**	0.84	88.62	1.07	82.66	0.85
0.25	90.59	*0.84*	**94.11**	0.85	90.08	1.12	84.48	0.92
0.5	96.58	0.66	**97.55**	0.75	93.31	1.09	90.94	*0.66*

Les résultats de la troisième colonne montrent la difficulté pour les différentes approches de travailler autour de l'axe achromatique. Dans ce cas, l'approche de Sobel développée en niveaux de gris est bien adaptée. Nous constatons que l'approche de Carron, qui inclut un axe intensité, est le deuxième meilleur score.

Enfin, le dernier cas, les deux distributions jaunes sont loin des canaux d'acquisition et de l'axe achromatique. Dans ce cas, l'importance de combiner les axes couleur correctement est nécessaire et explique le score du FVG.



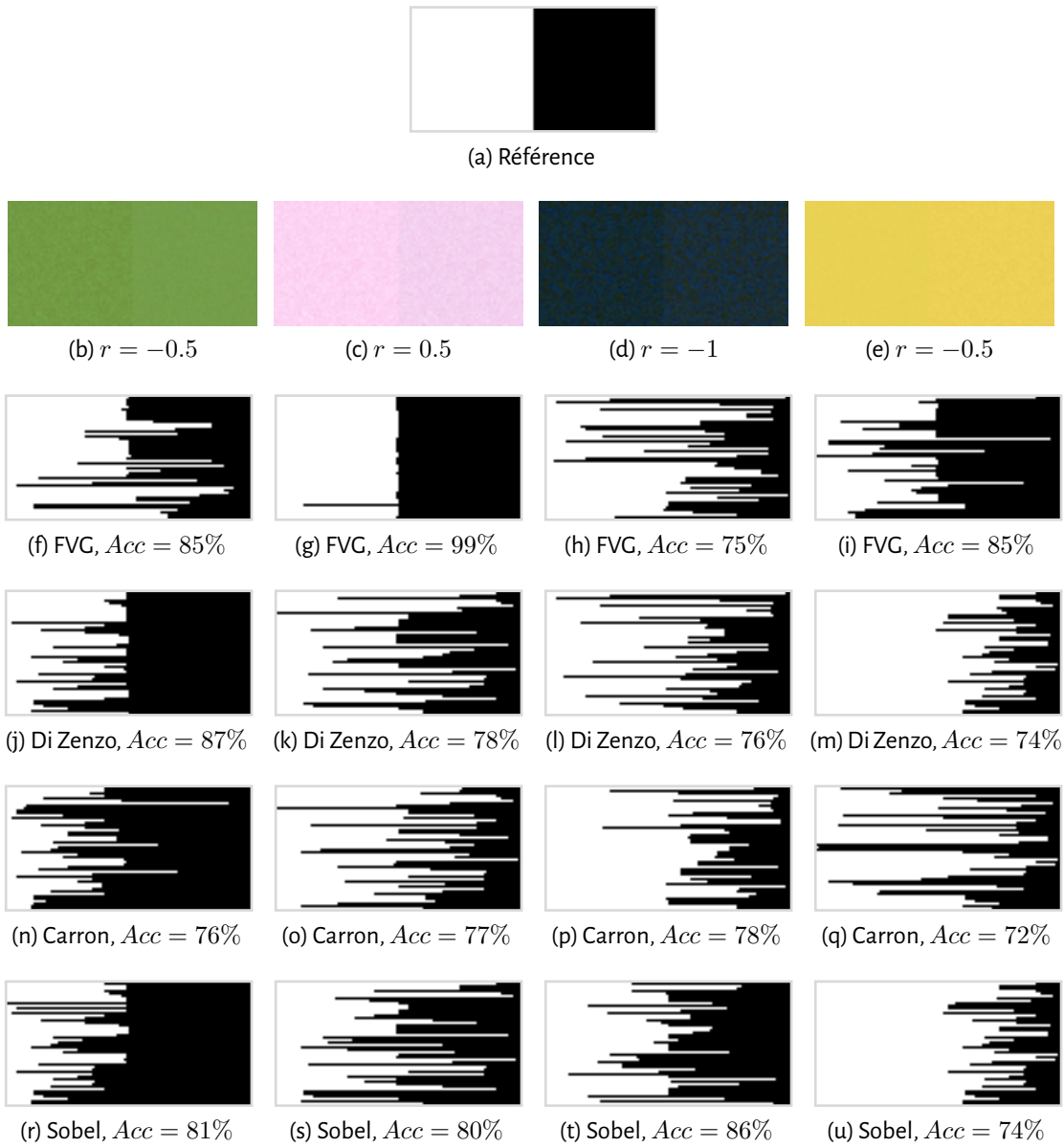
(a) Distributions 1.8a et 1.8a

(b) Distributions 1.8d et 1.8d

(c) Distributions 1.8a et 1.8d

(d) Distributions 1.8c et 1.8b

FIGURE 1.11 – **Exemples d'extraction de bords pour des images générées avec** $\Delta E = 3$. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



(a) Référence

(b) $r = -0.5$ (c) $r = 0.5$ (d) $r = -1$ (e) $r = -0.5$

(f) FVG, $Acc = 85\%$ (g) FVG, $Acc = 99\%$ (h) FVG, $Acc = 75\%$ (i) FVG, $Acc = 85\%$

(j) Di Zenzo, $Acc = 87\%$ (k) Di Zenzo, $Acc = 78\%$ (l) Di Zenzo, $Acc = 76\%$ (m) Di Zenzo, $Acc = 74\%$

(n) Carron, $Acc = 76\%$ (o) Carron, $Acc = 77\%$ (p) Carron, $Acc = 78\%$ (q) Carron, $Acc = 72\%$

(r) Sobel, $Acc = 81\%$ (s) Sobel, $Acc = 80\%$ (t) Sobel, $Acc = 86\%$ (u) Sobel, $Acc = 74\%$

FIGURE 1.12 – **Exemples de détection de bords pour des images générées avec un $\Delta E = 3$.** Les figures 1.12b, 1.12c, 1.12d et 1.12e servent à estimer le gradient. Les autres présentent les résultats de segmentation, la précision est donnée dans les sous-titres.

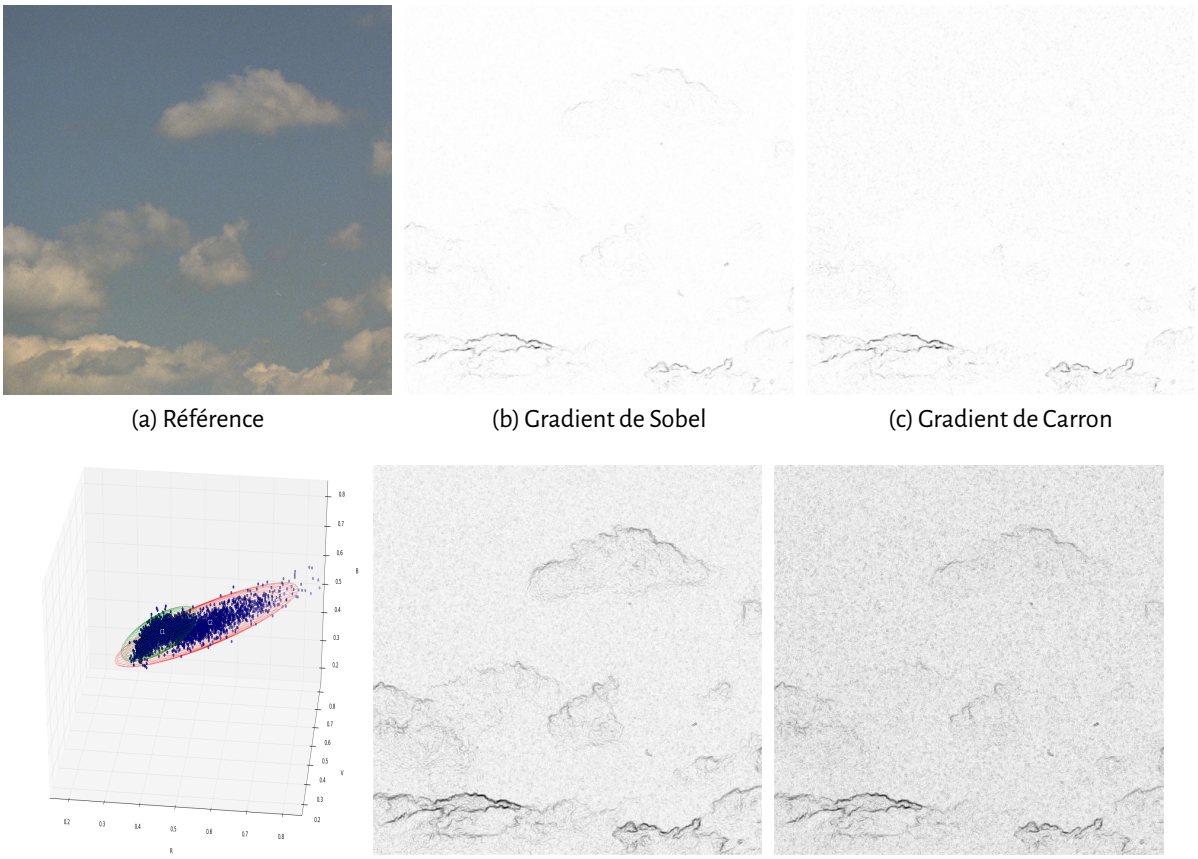
Une expérience a été réalisée avec des images générées dans $CIELAB$. Pour ne pas induire de confusion ces résultats sont présentés en annexe B.

1.3.2 Comparaison subjective des performances pour des images réelles

Dans cette section, nous comparons les gradients estimés sur des images du monde réel (base de données Stex Kwitt and Meerwald et VisTex MIT (1995)).

Obtenir une comparaison valide et objective des performances demande une base de données d'images avec des segmentations connues et valides. Nous avons aussi besoin de connaître la complexité couleur pour établir un taux de séparabilité. Cela n'est pas possible avec les bases de données existantes. C'est pourquoi, nous nous contenterons d'étudier subjectivement et de comparer les approches sur quelques images de deux bases de données. Pour atteindre le contenu couleur attendu, nous avons sélectionné des images du monde naturel qui présentent une complexité couleur plus grande que les objets manufacturés.

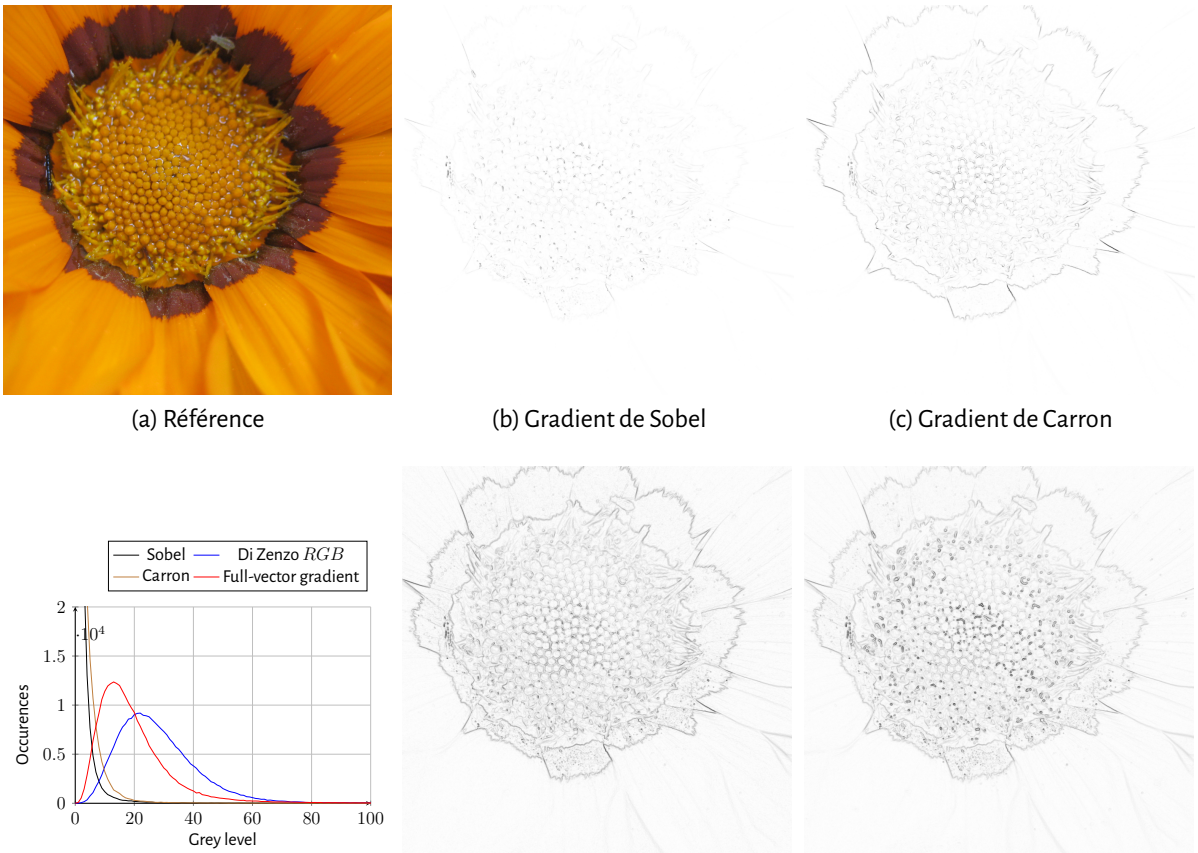
La première image sélectionnée (fig. 1.13a) est composée de nuages (gris et blanc) sur un ciel bleu non uniforme. Le gradient de Sobel, calculé en niveau de gris, ou l'approche de Carron devraient donner de bons résultats. Cependant ce n'est pas le cas. Même si la complexité couleur semble faible, il faut un gradient couleur plus adapté. L'histogramme couleur de l'image (fig. 1.13d) explique ces limites. Un modèle de mélange de Gaussiennes sépare la distribution couleur en deux classes. Les deux ellipsoïdes se chevauchent beaucoup. Ce modèle permet d'estimer un taux de séparabilité : -0.91. Cela explique que les méthodes de Sobel, Carron et Di Zenzo échouent. L'image résultat de Di Zenzo semble plus bruitée. Des bords manquent sur le nuage du haut et sur les petits en dessous. Le



(a) Référence (b) Gradient de Sobel (c) Gradient de Carron

(d) Histogramme couleur séparé en 2 classes (e) Full vector gradient (f) Gradient de Di Zenzo

FIGURE 1.13 – **Comparaison des gradients sur l'image nuage de la base VisTex**. La figure 1.13d représente l'histogramme couleur de l'image modélisée en deux Gaussiennes couleur. Cela évalue un taux de séparabilité $r = -0.91$. Les images de gradient sont inversées pour faciliter la perception, en revanche les dynamiques sont identiques pour les différents résultats.



(a) Référence (b) Gradient de Sobel (c) Gradient de Carron

(d) Histogrammes des gradients (e) Full vector gradient (f) Gradient de Di Zenzo

FIGURE 1.14 – **Comparaison d'une fleur de la base Stex**. La figure 1.14d représente l'histogramme des gradients. Les images de gradient ont été inversées mais ils ont la même dynamique expliquant l'aspect imperceptible des gradients de Sobel et de Carron.

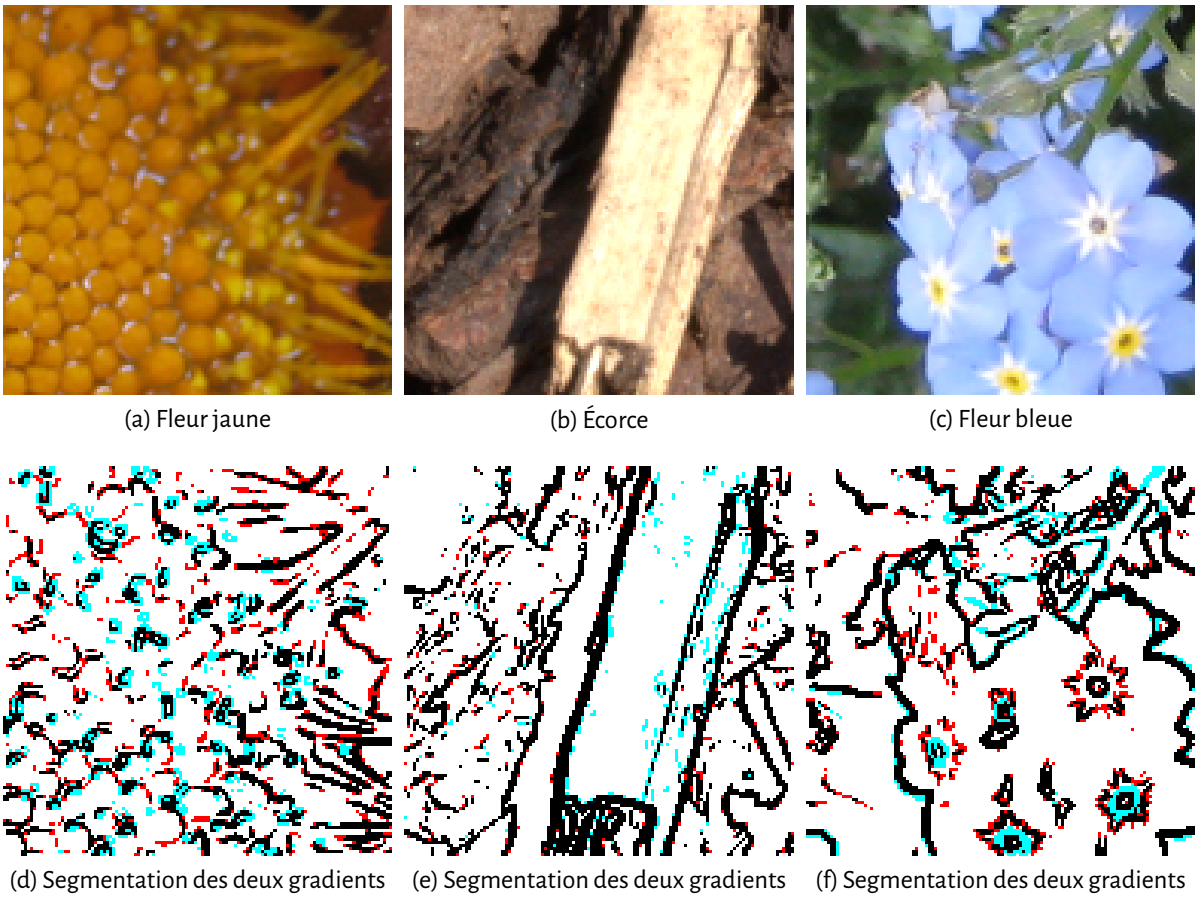
full-vector gradient extrait correctement ces bords. Grâce à la haute luminosité, les corrélations entre les canaux sont fortes. L'approche de Di Zenzo surestime les petites différences dans le ciel bleu et perd en discrimination dans la partie diffuse entre le ciel et le nuage. Ce résultat est cohérent avec les résultats des données synthétiques obtenus.

La deuxième image représente une fleur de la base Stex. Cette image est plus chromatique, principalement orange, avec une couronne brune et quelques pics de spécularité dus aux gouttes d'eau. Le gradient de Carron produit de meilleurs résultats que dans le cas précédent mais la dynamique reste faible (voir l'histogramme sur le figure 1.14d). À première vue, l'approche de Di Zenzo semble apporter plus d'information. Mais, les points extraits correspondent aux pics de spécularité observés au centre de la fleur. Ces pics réduisent la discrimination des pistils. Malgré ces différences, les deux approches présentent des résultats similaires.

La figure 1.15 présente trois résultats d'extraction de bords après avoir segmenté l'image de gradient. Le seuil appliqué pour la segmentation est le troisième quartile de l'histogramme du gradient de l'image. Une légende à trois couleurs permet de comparer les deux approches. Les pixels noirs sont les bords détectés par les deux gradients. Les pixels rouges correspondent aux détections du FVG lorsque le Di Zenzo ne détecte rien. Au contraire, les pixels cyans représentent les bords détectés par Di Zenzo mais pas par FVG.

Sur la figure 1.15d, les bords extraits par le FVG (noir et rouge) montrent les parties rondes au centre de la fleur et segmentent mieux les étamines sur la gauche. À l'inverse, Di Zenzo (noir et cyan) segmente mieux l'eau brillante.

Sur la figure 1.15e, l'approche de Di Zenzo surestime les variations dans la partie claire de l'écorce.



(a) Fleur jaune (b) Écorce (c) Fleur bleue

(d) Segmentation des deux gradients (e) Segmentation des deux gradients (f) Segmentation des deux gradients

FIGURE 1.15 – **Différences en extraction de bords entre le FVG et Di Zenzo**. Le seuil est placé au troisième quartile de l'histogramme du gradient de l'image. Les pixels noirs correspondent aux bords extraits par les deux approches, les cyans correspondent aux bords extraits par Di Zenzo mais pas par le FVG et les rouges sont les pixels extraits par le FVG mais pas par Di Zenzo.

Au contraire, le *full-vector gradient* est plus sensible dans les régions sombres.

La figure 1.15f est la plus intéressante. Le centre de la fleur est construit par zones concentriques, du brun au jaune, puis blanc et enfin bleu. Sur la fleur en bas à gauche de l'image, le gradient de Di Zenzo considère tout le cœur jaune comme un seul et unique bord, tandis que le *full-vector gradient* se concentre sur le bord entre le brun et le jaune. La segmentation du bord entre le blanc et le bleu est beaucoup plus complète avec le FVG. Nous pouvons dire que le *full-vector gradient* donne de meilleurs résultats que Di Zenzo.

1.3.3 Exemples d'applications de calcul de gradient

Pour illustrer la généricité du *full-vector gradient*, nous proposons deux expériences. La première compare la norme du gradient extraite depuis un capteur couleur, multi- et hyper-spectral. Dans la seconde, nous comparerons le gradient extrait dans le domaine couleur pour estimer l'impact potentiel de la déficience visuelle (CVD pour *color vision deficiency*) sur l'estimation du gradient.

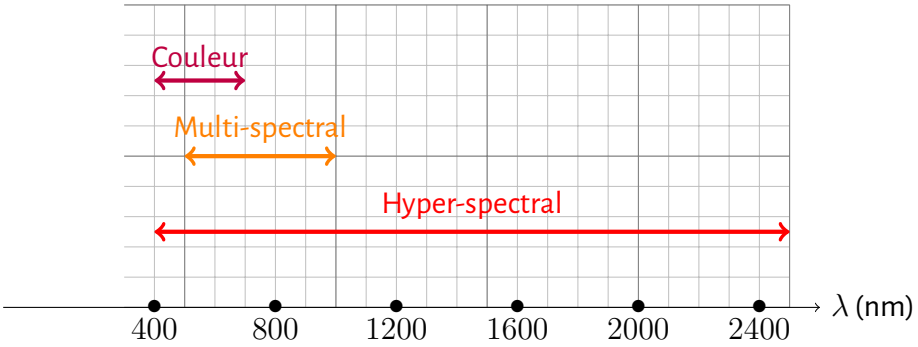
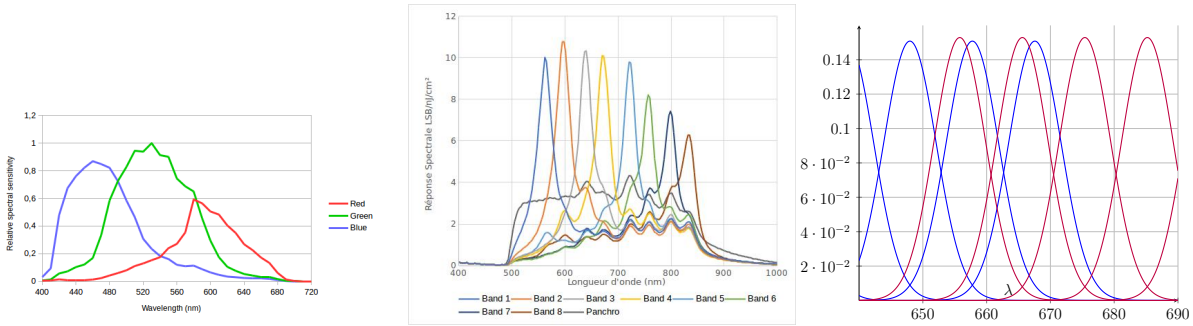


FIGURE 1.16 – **Domaine de perception des différents appareils** couleur, multi- et hyper-spectral présentés dans cette partie.

1.3.3.1 Vers le domaine physique,

Dans cette expérience, nous simulons la réponse d'un capteur couleur et celle d'un capteur multi-spectral depuis une image acquise avec un capteur hyper-spectral. L'image *Forest_x350_y2990* (fig. 1.20b) est une image aéroportée venant de la base de données *AVIRIS* Green et al. (1998). *AVIRIS* signifie *Airborne Visible InfraRed Imaging Spectrometer*, c'est un capteur 224 bandes allant de 366 nm à 2496 nm. L'appareil photo simulé pour l'image couleur est un Canon 500D, ses courbes de sensibilité spectrale sont présentées figure 1.17a. Pour le multi-spectral, l'appareil simulé est un capteur 9-bandes CMS-V par SILIOS Technologies, ses SSF sont présentées figure 1.17b. Les huit premiers canaux sont centrés respectivement en $\{561, 596, 638, 673, 722, 758, 801, 838\}$ nm. Le dernier est un filtre panchromatique.



(a) Capteur couleur (b) Capteur multi-spectral (c) Capteur hyper-spectral (zoom)

FIGURE 1.17 – **Fonctions de sensibilité spectrale de trois appareils d'acquisition**. Le premier possède un capteur trichromatique, c'est une Canon 500D. Le second est un capteur multi-spectral de huit canaux et un panchromatique (CMS-V) créé par SILIOS Technologies. Le troisième ne présente que huit de ces canaux (4 bleus sur le capteur couleur et 4 rouges du capteur proche infra-rouge).

Les images simulées pour le capteur couleur et multi-spectral le sont grâce à l'équation (1.27) :

$$C_i = \int_{\lambda_{min}}^{\lambda_{max}} S_i(\lambda) I(\lambda) d\lambda, \tag{1.27}$$

où C_i (resp. $S_i()$) est la valeur (resp. la SSF) du canal i. La valeur de l'image $I(\lambda)$ est le produit de la réflectance de la scène par l'illuminant. Les images d'*AVIRIS* sont des images de radiance, sans aucune information sur les conditions d'ensoleillement lors de la prise de vue.

Chaque capteur possède une matrice de Gram liée à ses propres SSF (fig. 1.17). La figure 1.18 présente les matrices de Gram des capteurs couleur et multi-spectral. Il y a chevauchement des SSF pour les deux capteurs. Le produit scalaire des produits croisés est donc non nul.

$$\begin{pmatrix} 20 & 23 & 7 \\ 23 & 80 & 38 \\ 7 & 38 & 55 \end{pmatrix}$$

(a) Canon 500 D

$$\begin{pmatrix} 3615 & 2741 & 1856 & 1831 & 1857 & 1489 & 1452 & 1475 & 2933 \\ 2741 & 4293 & 2751 & 2435 & 1879 & 1585 & 1570 & 1638 & 3131 \\ 1856 & 2751 & 3813 & 2939 & 2030 & 1783 & 1792 & 1617 & 3045 \\ 1831 & 2435 & 2939 & 3803 & 2393 & 2015 & 1827 & 1641 & 3023 \\ 1857 & 1879 & 2030 & 2393 & 3326 & 2356 & 1904 & 1662 & 2859 \\ 1489 & 1585 & 1783 & 2015 & 2356 & 2662 & 2070 & 1738 & 2427 \\ 1452 & 1570 & 1792 & 1827 & 1904 & 2070 & 2326 & 1837 & 2287 \\ 1475 & 1638 & 1617 & 1641 & 1662 & 1738 & 1837 & 1973 & 2101 \\ 2933 & 3131 & 3045 & 3023 & 2859 & 2427 & 2287 & 2101 & 3785 \end{pmatrix}$$

(b) CMS-V (SILIOS Technologies)

FIGURE 1.18 – **Matrices de Gram pour les capteurs couleur et multi-spectral**.

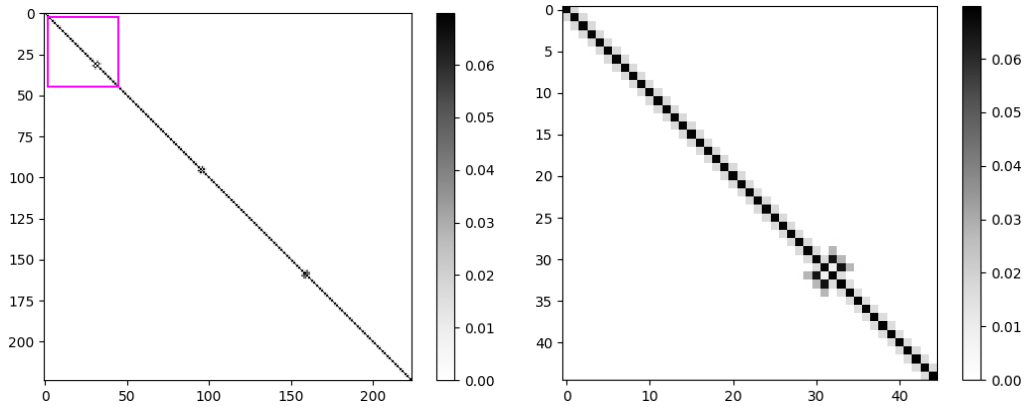
L'aire sous la courbe de sensibilité du rouge pour le capteur couleur (fig. 1.17a) est plus petite que pour les autres. Ainsi, son coefficient associé à la matrice de Gram (fig. 1.18a) est le plus petit (20) sur la diagonale. Le canal bleu chevauche plus le canal vert que le rouge, la corrélation entre le bleu et le vert (23) est donc plus grande que celle entre le bleu et le rouge (7).

Les canaux rouge et infrarouge ont une plus faible sensibilité (fig. 1.17b), ce qui implique que les valeurs de la matrice de Gram associée (fig. 1.18b) soient plus faibles que le reste de la diagonale. Deuxièmement, les dernières ligne et colonne (valeurs en bleu) ont des coefficients plus importants puisqu'elles correspondent au filtre panchromatique. Ce filtre peut être assimilé à l'intensité de l'image sur toutes les longueurs d'onde. Enfin, la sélectivité des filtres non panchromatiques, n'est pas très importante, donc les canaux ne se chevauchent majoritairement qu'avec leur voisinage (le filtre précédent et suivant). Cela correspond aux valeurs en magenta sur le matrice de Gram.

Pour le capteur hyper-spectral, nous avons seulement la mesure de largeur à mi-hauteur (FWHM pour *full width at half maximum*) pour chaque canal. La sensibilité spectrale des canaux est modélisée par une fonction Gaussienne, centrée sur la longueur d'onde du canal avec une déviation standard σ reliée à la FWHM par :

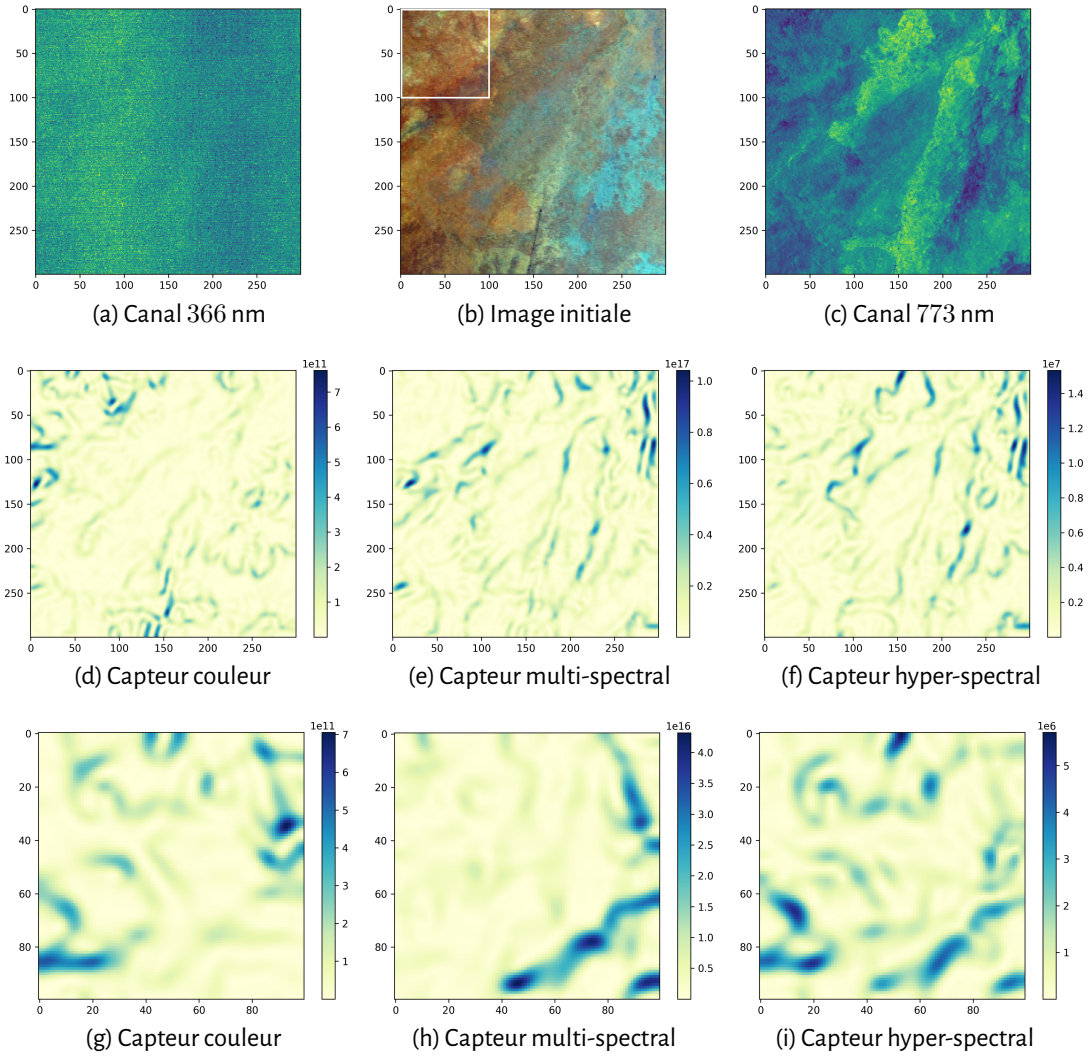
$$FWHM = 2\sqrt{2\ln(2)}\sigma. \tag{1.28}$$

Un extrait des courbes de sensibilité ainsi simulées est présenté figure 1.17c. La zone représentée correspond à un changement de capteur. Les quatre Gaussiennes bleues sont les derniers canaux du capteur couleur et les quatre rouges correspondent aux premiers du capteur proche infrarouge. Comme la FWHM est presque égale à la distance entre les longueurs d'ondes de deux canaux consécutifs, une Gaussienne ne chevauche significativement que les Gaussiennes associées au canal précédent et suivant. Nous nous attendons donc à obtenir une matrice de Gram presque diagonale (fig. 1.19). C'est effectivement vérifié. Cependant, l'imageur *AVIRIS* est constitué de trois capteurs : un dans le visible (366 à 668 nm), un dans le proche infrarouge (656 à 1263 nm), et un dans l'infrarouge (1253 à 2496 nm). Des discontinuités sont observables lors du léger recouvrement entre deux capteurs. Cela correspond aux plus gros carrés sur la figure 1.19.



(a) Spectre complet (b) Spectre du visible

FIGURE 1.19 – **Matrice de Gram pour le capteur hyper-spectral** (*AVIRIS*).



(a) Canal 366 nm (b) Image initiale (c) Canal 773 nm

(d) Capteur couleur (e) Capteur multi-spectral (f) Capteur hyper-spectral

(g) Capteur couleur (h) Capteur multi-spectral (i) Capteur hyper-spectral

FIGURE 1.20 – **Résultats pour trois types de capteur restreints au visible**. Le capteur couleur est un Canon 500D, le multi-spectral est un CMS-V par SILIOS Technologies et l'hyper-spectral est le capteur *AVIRIS*. Le carré blanc correspond à la zone agrandie des figures 1.20g, 1.20h et 1.20i.

Tout d'abord, nous comparons les gradients calculés à partir des capteurs restreints au visible, c'est à dire de 360 à 780 nm. La figure 1.20 présente l'image couleur et les gradients estimés pour les trois capteurs. Sur la première ligne, nous trouvons une extrapolation couleur de l'image initiale (fig. 1.20b) et deux canaux spécifiques : un bleu (fig. 1.20a) et un rouge (fig. 1.20c). La seconde présente les gradients obtenus par les trois capteurs restreints au visible pour la totalité de l'image (fig. 1.20d, 1.20e, 1.20f). Tandis que la troisième se focalise sur une zone de l'image (fig. 1.20g, 1.20h, 1.20i).

Nous observons que les dynamiques de la norme des gradients sont très différentes. Elles sont liées au nombre de canaux, mais aussi à la sélectivité de ceux-ci. Plus les canaux sont sélectifs, plus la matrice associée sera diagonale et plus l'amplitude de la norme sera petite. Ainsi, le capteur multi-spectral qui a une sélectivité faible et un canal panchromatique se retrouve avec une dynamique très haute.

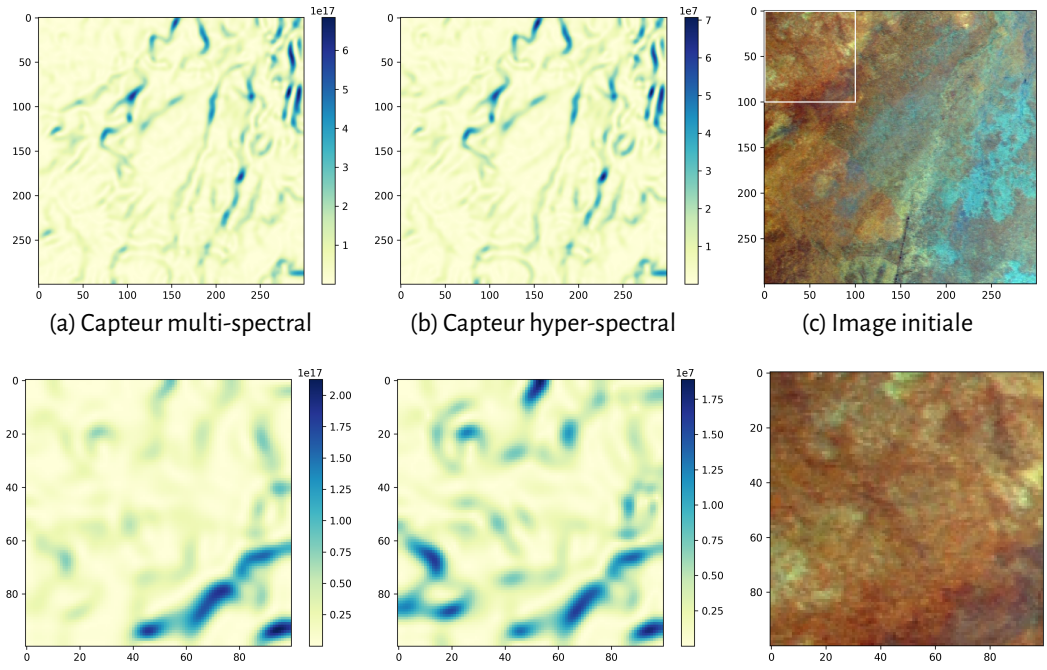
L'image couleur est très cohérente avec les gradients attendus de la figure 1.20b pour la zone de forêt en haut à gauche. Les deux autres gradients répondent moins fort dans cette zone. Cependant, les capteurs multi- et hyper-spectral détectent des normes de gradient très hautes dans le coin haut droit de l'image. Nous supposons que cela empêche de percevoir les informations dans le reste de l'image. Le zoom effectué (fig. 1.20g, 1.20h, 1.20i) confirme cette hypothèse. Même si plus d'informations sont visibles pour les trois gradients, elles ne sont pas identiques. Le capteur multi-spectral extrait moins d'information. Ceci est probablement dû aux informations manquantes dans les longueurs d'onde du bleu. Sans vérité-terrain, il est difficile d'aller plus loin dans l'analyse des résultats.

L'augmentation du nombre de canaux entraîne une meilleure discrimination. Par exemple, le coin en haut à droite présente des gradients de grande amplitude qui sont moins perçus par le capteur couleur. Si le gradient est fort sur quelques canaux seulement, un capteur couleur ne peut le mesurer : l'intégrale sur un large spectre lisse les variations. La rivière ou piste qui traverse l'image sur la diagonale allant vers le coin haut droit est plus visible avec le gradient du capteur hyper-spectral qu'avec le capteur multi-spectral. Ce n'est quasiment pas extrait avec le capteur couleur.

Les trois capteurs limités au domaine visible offrent des informations différentes. Cela encourage l'idée d'un capteur multi-échelles spectral. Il s'agit de combiner des canaux avec des sélectivités différentes pour obtenir des informations plus discriminantes.

Ensuite, nous prenons l'ensemble des capteurs multi- et hyper-spectral pour estimer le gain d'information. La figure 1.21 présente les gradients estimés sur les spectres complets des capteurs CMS-V (SILIOS Technologies) et *AVIRIS*. Puisque les longueurs d'onde dans l'infrarouge ont été ajoutées, la dynamique de la norme du gradient est respectivement 6 et 5 fois plus grande que lorsque les capteurs sont limités au visible. Les images de norme de gradient (fig. 1.21a, 1.21b) sont similaires. Cela montre l'importance de l'information dans l'infrarouge, qui est un fait connu pour les zones de végétations. Nous devons regarder dans les détails pour remarquer des différences locales (fig. 1.21d, 1.21e). Plusieurs éléments expliquent ces différences : le manque d'information acquise dans le bleu pour le capteur multi-spectral ; sa plage spectrale limitée à l'infrarouge contrairement à l'hyper-spectral ; la sélectivité du capteur hyper-spectral est beaucoup plus importante.

Cette première expérience a montré que le FVG est adapté pour calculer les gradients pour différents types de capteurs. Nous avons montré que les gradients estimés sont dépendants du contenu de la scène tout en prenant en compte avec exactitude les caractéristiques du capteur. Nous avons aussi expliqué pourquoi les matrices de Gram associées aux capteurs ne sont pas diagonales. Cela invalide l'hypothèse d'un espace d'acquisition orthogonal. Même si cette hypothèse peut être utilisée pour certains capteurs hyper-spectraux, utiliser la matrice de Gram permet de garder l'objectif métrologique.



(a) Capteur multi-spectral (b) Capteur hyper-spectral (c) Image initiale

(d) Capteur multi-spectral (e) Capteur hyper-spectral (f) Zoom de l'image

FIGURE 1.21 – **Résultats pour les capteurs multi- et hyper-spectral sur leurs spectres complets.** Le carré blanc correspond à la zone agrandie des figures 1.21d, 1.21e.

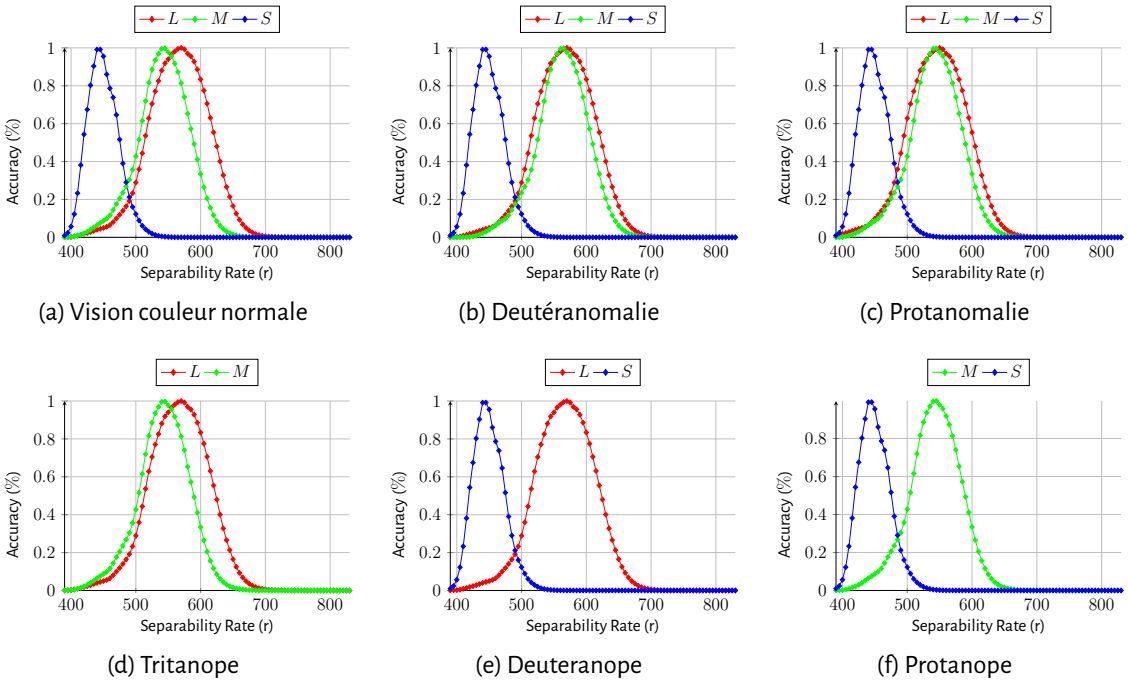
1.3.3.2 Et au delà !

Dans cette partie,nous nous demandons l'intérêt du calcul du gradient prenant en compte l'inter-corrélation des cônes dans l'analyse du système visuel humain (HSV pour *human visual system*). Considérant les fonctions d'appariement couleur (CMF pour *color matching function*) comme une SSF, nous proposons de comparer la détection de gradient entre une vision couleur normale et les visions couleur déficientes (CVD pour *color vision deficiency*) aussi appelé daltonisme en français. Cette étude se fait dans l'espace couleur défini par les cônes (LMS).

La CIE propose des CMF pour l'espace LMS pour un observateur standard à $2°$ et $10°$. Pour notre expérience, nous utilisons les courbes à $2°$ pour la vision normale des couleurs. Les CMF définissant les visions déficientes utilisent le modèle proposé par Shrestha (2016) (fig. 1.22). Pour les anomalies de cônes, les courbes de sensibilité se décalent. Pour la protanomalie, la sensibilité du cône L se décale vers celle du cône M (fig. 1.22c). Pour la deutéranomalie, la sensibilité du cône M glisse vers celle du cône L. Un déplacement de 20 nm est considéré comme une anomalie sévère. C'est le déplacement choisi. Pour les dichromates, la courbe associée au cône manquant est simplement enlevée. Le cône L est manquant chez les protanopes, le M chez les deutéranopes et le S chez les tritanopes. Nous appliquons le FVG avec ces CMF.

Pour estimer l'image perçue, nous utilisons la même méthode que dans la sous-section 1.3.3.1. Les CMF de la figure 1.22 jouent le rôle de SSF et nous appliquons l'équation (1.27). Pour mesurer le gradient, nous avons besoin de calculer les matrices de Gram associées à chaque vision couleur (tab. 1.2). Les valeurs sur la diagonale correspondent à la norme des courbes de sensibilité, elles sont donc identiques sur toutes les matrices.

Pour les dichromates, les courbes de sensibilité sont les mêmes à l'exception du cône manquant. Pour obtenir la matrice de Gram associée, à partir de la matrice de vision normale des couleurs, on enlève la ligne et la colonne liées au cône manquant. La matrice associée à la vision deutéranope (le cône M manque) correspond à la figure 1.22b. Comme le graphique le montre, les intégrales des deux courbes entraînent des aires très différentes. Ceci est cohérent avec la différence de valeurs sur la diagonale.



(a) Vision couleur normale (b) Deutéranomalie (c) Protanomalie

(d) Tritanope (e) Deuteranope (f) Protanope

FIGURE 1.22 – **Courbes de sensibilité pour les différentes visions**. La première ligne correspond aux visions trichromatiques, la seconde aux dichromates.

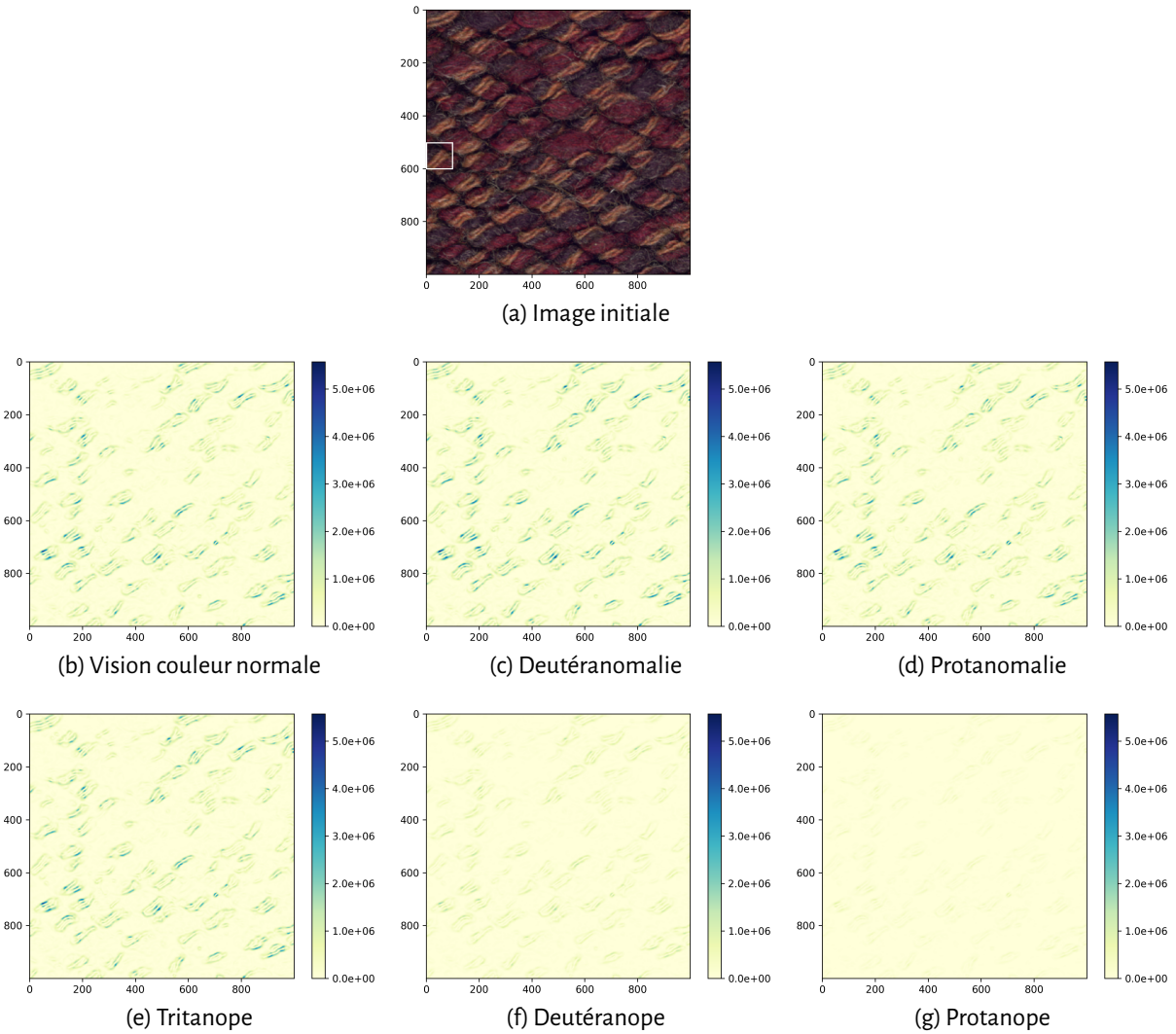
TABLE 1.2 – **Matrices de Gram pour les différentes visions couleur**

	Normal	Cône anormal	Cône manquant
Cône L	$\begin{pmatrix} 85.487 & 68.859 & 4.305 \\ 68.859 & 66.054 & 6.876 \\ 4.305 & 6.876 & 41.166 \end{pmatrix}$	$\begin{pmatrix} 85.487 & 74.211 & 8.392 \\ 74.211 & 66.054 & 6.876 \\ 8.392 & 6.876 & 41.166 \end{pmatrix}$	$\begin{pmatrix} 66.054 & 6.876 \\ 6.876 & 41.166 \end{pmatrix}$
Cône M		$\begin{pmatrix} 85.487 & 74.211 & 4.305 \\ 74.211 & 66.054 & 3.320 \\ 4.305 & 3.320 & 41.166 \end{pmatrix}$	$\begin{pmatrix} 85.487 & 4.305 \\ 4.305 & 41.166 \end{pmatrix}$
Cône S			$\begin{pmatrix} 85.487 & 68.859 \\ 68.859 & 66.054 \end{pmatrix}$

Pour les cônes anormaux, la corrélation entre les cônes L et M est très grande car leurs courbes associées coïncident presque entièrement. Comparée à la vision normale, la protanomalie, *i.e.* L vers M, (respectivement la deutéranomalie, *i.e.* M vers L) implique une corrélation plus forte (respectivement plus faible) avec le cône S. C'est pourquoi le coefficient entre L et S 8.392 (respectivement M et S 3.320) est plus haut (respectivement plus bas) que la valeur pour la vision normale 4.305 (respectivement 6.876).

Pour montrer l'impact de cette modélisation des CMF pour la CVD, nous utilisons une image de la base de données *HyTexila* Khan et al. (2018). Elle contient des images prises dans des conditions supervisées (distance et illuminant) avec un capteur hyper-spectral 186 bandes (405 nm à 996 nm). Nous utilisons l'image *textile_18_red*. Les figures 1.23a et 1.25a présentent une représentation RGB de l'image utilisée, respectivement entière et zoomée.

Les résultats de normes de gradients pour les six visions sont présentés figure 1.23. Nous avons choisi ici d'utiliser la même dynamique pour les six images. Pour les trichromates (fig. 1.23b, 1.23c, 1.23d), les gradients estimés présentent une grande amplitude pour les pixels correspondant à un changement entre les laines jaune, violette et rouge.



(a) Image initiale

(b) Vision couleur normale (c) Deutéranomalie (d) Protanomalie

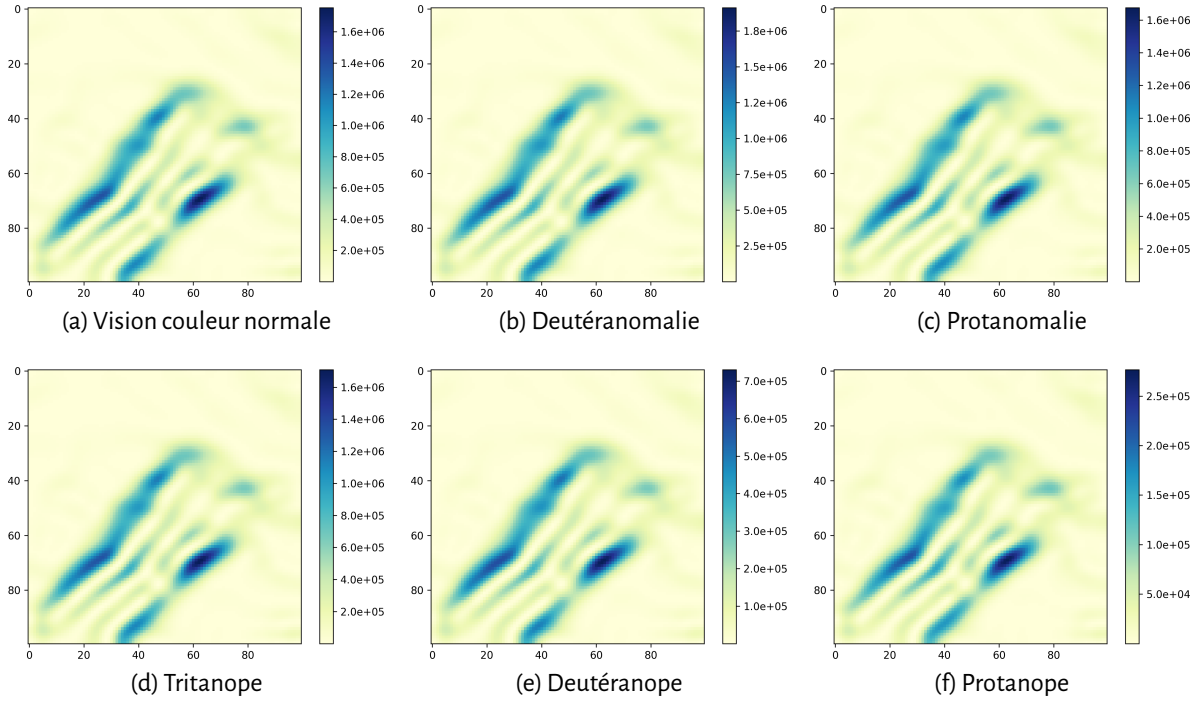
(e) Tritanope (f) Deutéranope (g) Protanope

FIGURE 1.23 – **Normes des gradients estimés pour les différentes visions**. La carré blanc correspond à la zone d'intérêt des figures 1.24 et 1.25.

L'image de norme de gradient pour la vision tritanope (fig. 1.23e) a une amplitude plus grande que pour les autres dichromates. Les tritanopes n'ont pas le cône S, *i.e.* l'information bleue. Le jaune est une combinaison de rouge et de vert, le violet de rouge et de bleu. Ceci est cohérent avec les résultats. La figure 1.23g présente les normes des gradients les plus faibles. La vision protanope reçoit de l'information des cônes M et S (couleurs bleue et verte). Différencier du rouge et du violet ou du rouge et du jaune est plus compliqué sans le cône L.

Pour afficher la zone d'intérêt (fig. 1.24), nous choisissons d'adapter l'échelle de dynamique par figure. Toutes les images sont similaires à l'exception de leurs échelles. La plus petite correspond à la vision protanope, puis vient le deutéranope. La vision tritanope et protanomalie ont des échelles similaires proches de celle de la vison normale. L'échelle la plus grande est obtenue pour la deutéranomalie. Le système visuel humain est principalement basé sur le contraste. Donc, bien qu'avec des dynamiques différentes, la perception de toutes ces visions pour cette scène est très similaire. Ceci est cohérent avec le faible handicap causé par les CVD.

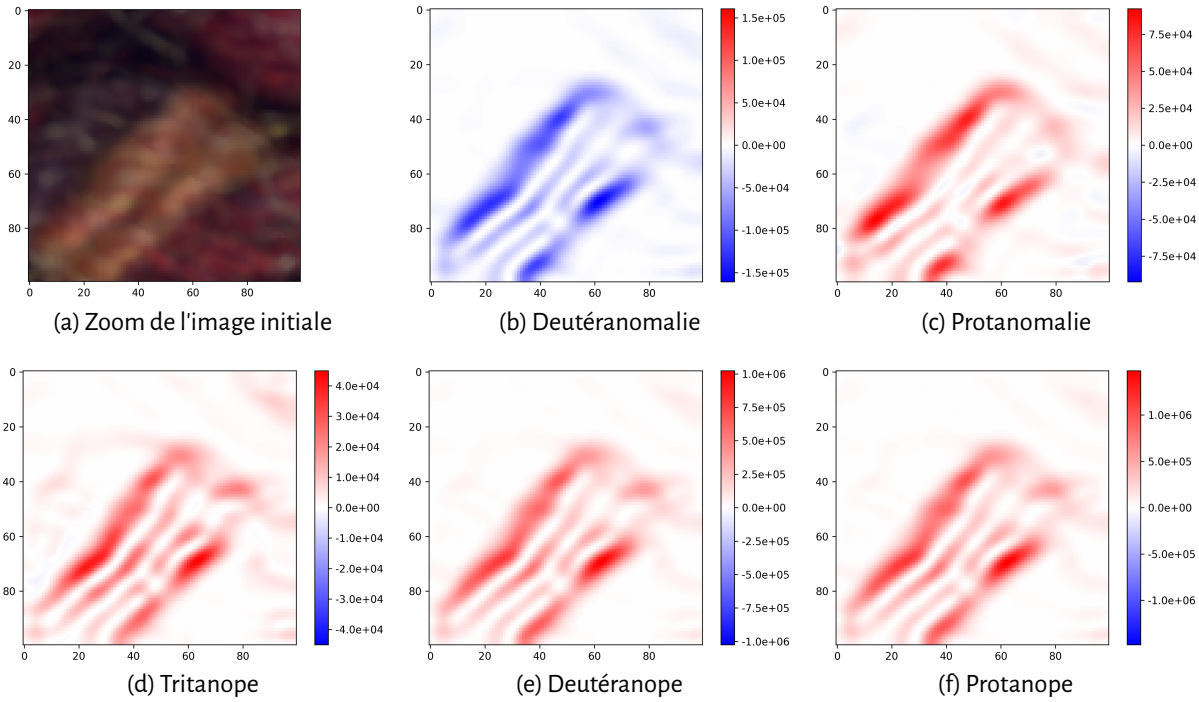
Pour mieux comprendre l'impact d'une vision déficiente, nous avons calculé la différence entre la vision normale et les visions déficientes. Les résultats sont présentés sur la figure 1.25. Les visions déficientes auxquelles il manque un cône (fig. 1.25f, 1.25e, 1.25d) sont très similaires exceptée l'échelle. La plus petite différence est obtenue pour les tritanopes. Les couleurs présentes dans l'image sont mieux discriminées avec les cônes L et M. La deutéranomalie donne des normes de gradients plus grandes que la vision normale. Comme le cône M est décalé vers le cône L, ceux-ci se chevauchent sur le pic



(a) Vision couleur normale (b) Deutéranomalie (c) Protanomalie

(d) Tritanope (e) Deutéranope (f) Protanope

FIGURE 1.24 – **Zoom des différentes normes de gradients estimés pour les différentes visions couleur**.



(a) Zoom de l'image initiale (b) Deutéranomalie (c) Protanomalie

(d) Tritanope (e) Deutéranope (f) Protanope

FIGURE 1.25 – **Différences entre les visions couleur normale et déficientes**. Les valeurs en bleu indiquent que la vision déficiente a une norme supérieure à celle normale. Au contraire, en rouge, la norme du gradient de la vision normale est plus grande que la déficiente.

de réflectance de la laine jaune. La protanomalie génère les gradients les plus éloignés de la vision normale. Certaines zones ont des normes de gradients plus faibles notamment la laine jaune. Alors que d'autres ont des normes plus grandes, elles correspondent aux zones très éclairées/brillantes de l'image.

1.4 Synthèse

L'estimation du gradient couleur en traitement d'image est une question ouverte depuis plus de trente ans déjà. Nous avons ici un objectif métrologique de reproductibilité et d'indépendance des gradients mesurés. Avec cet objectif, nous avons repris la définition du gradient pour obtenir une formulation qui nous convenait. L'utilisation de la matrice d'auto-corrélation proposée par Di Zenzo (1986) était très intéressante. Cependant, la supposition d'un espace orthogonal était limitante. Nous avons donc utilisé la matrice de Gram pour prendre en compte l'auto-corrélation des canaux et donc l'information intrinsèque aux capteurs. Pour cela nous mesurons le produit scalaire entre les courbes de sensibilité spectrale des canaux dans l'espace des fonctions intégrables.

Nous avons ensuite mis en regard l'état de l'art avec cette méthode. Bien que les méthodes estimées dans les espaces transformés interdisent la validité métrologique, nous avons présenté l'approche de Carron and Lambert (1994) qui imite la perception humaine en prenant en compte un gradient teinte lorsque la saturation est forte et valorisant l'information luminosité le reste du temps. L'expression de Di Zenzo (1986) s'insère dans une famille qui considère les espaces RGB comme orthogonaux. Les gradients dit non-linéaires ou morphologiques, bien qu'intéressants ne répondent pas bien aux distributions aléatoires. Nous avons choisi l'approche niveau de gris proposée par Sobel and Feldman (1968) pour comparer nos résultats.

Afin de valider la proposition de gradient couleur réalisée dans ce chapitre, nous avons d'abord appliqué ce gradient sur des images simulées de trois types de capteurs (hyper, multi-spectral et couleur). Nous observons des différences en fonction des capteurs utilisés. Les résultats appuient l'analyse spectrale multi-échelles des images. Nous suggérons aussi l'utilisation du FVG dans l'espace perceptuel LMS. Il s'agit là d'une ouverture sur le domaine perceptuel qui appelle de nombreuses autres questions.

Nous proposons ensuite de nous comparer avec l'état de l'art. Les résultats se séparent en deux classes : analyse objective au travers de deux expériences menées avec notre base de données artificielle et analyse subjective sur des images réelles. L'hypothèse de cette base artificielle, nous le rappelons, suppose que la difficulté d'extraction d'une norme de gradient de forte amplitude est liée à l'entrelacement des distributions couleur que nous essayons de séparer. Le FVG donne de meilleurs résultats que les autres gradients analysés. Plus les distributions sont oblongues et d'orientations différentes, *i.e.* plus nous sommes dans les conditions réelles, meilleurs sont les résultats. L'analyse subjective montre l'avantage des méthodes de Di Zenzo et du FVG. Si les deux approches présentent des points forts différents, la dernière segmentation présente une supériorité du FVG.

Ce chapitre propose de se plonger dans l'univers des gradients couleur. Nous n'avons fait que survoler l'ensemble des possibilités qu'ils offrent. En effet, nous nous sommes essentiellement concentrés sur la norme du gradient couleur proposé (FVG). Son orientation et sa couleur sont deux informations tout aussi importantes. Le prochain chapitre, va permettre de mettre en œuvre l'orientation estimée par cette méthode avec une proposition de détecteur couleur de points d'intérêt.

Résumé

- Démonstration des limites de la proposition de Di Zenzo supposant implicitement l'espace couleur utilisé comme orthogonal.
- Adaptation de la proposition de Di Zenzo rendant la formulation indépendante aux nombres de canaux.
- Proposition d'une formulation prenant en compte les caractéristiques physiques du capteur.
- Mise en place d'un protocole de validation du pouvoir discriminant des gradients couleur sur une base de données artificielles.
- Ouverture sur l'adaptation du *full-vector gradient* à une mesure perceptuelle sur les visions couleur déficientes.

Détecteurs de points d'intérêt

Contenu

2.1 État de l'art 53
2.1.1 Détecteurs de coins 54
2.1.1.1 Détecteurs de coins basés gradients 54
2.1.1.2 Détecteurs de coins basés motif 57
2.1.2 Détecteurs basés *blob* 58
2.1.3 Détecteurs basés régions 60
2.2 Détecteurs de points d'intérêt couleur 61
2.2.1 Adaptation aux multi-échelles 62
2.2.1.1 Choix du filtre de dérivation 62
2.2.1.2 Vers le multi-échelles 63
2.2.2 Adaptation à la couleur 64
2.2.3 Mise en forme 64
2.2.3.1 Extraction du gradient 65
2.2.3.2 Fonction réponse 66
2.2.4 Sélection des “bons” points d'intérêt 67
2.3 Preuve de l'approche 69
2.3.1 Base de données artificielles 71
2.3.1.1 Sensibilité des détecteurs 71
2.3.1.2 Distance aux points réels 72
2.3.1.3 Influence de la complexité sur le rappel et la précision 76
2.3.2 Base de données Oxford 78
2.3.2.1 Protocole de l'expérimentation 78
2.3.2.2 Résultats 81
2.3.3 Comparaison point d'intérêt/point visuellement saillant 86
2.3.3.1 Protocole expérimental 86
2.3.3.2 Résultats 88
2.4 Synthèse 93

Les tâches de haut-niveau de type indexation, classification ou reconstruction $3D$ utilisent principalement deux outils l'un après l'autre : le détecteur puis le descripteur. Le détecteur de point d'intérêt a pour objectif de déterminer les points pertinents de l'image. Un point pertinent est un point dont le voisinage est suffisamment spécifique pour pouvoir être retrouvé dans une autre image. L'objectif du descripteur est de pouvoir reconnaître les points d'intérêt trouvés par le détecteur dans chacune des images. Pour cela, le voisinage du point d'intérêt détecté est utilisé pour décrire son environnement. Cette description doit permettre d'apparier les points d'une image à l'autre. Le détecteur et le descripteur ont donc des objectifs complémentaires bien que très différents. Nous avons choisi de bien marquer cette différence en les décrivant dans deux chapitres séparés. Nous nous intéressons dans celui-ci au détecteur.

Revenons à la notion de points pertinents. La pertinence du point vient évidement de son voisinage. Celui-ci peut mettre en évidence que le point est le sommet d'un coin, une jonction. Le voisinage peut aussi représenter une zone uniforme qui tranche avec son contour. L'élément clé des points détectés est bien de pouvoir être retrouvé dans une scène similaire ou sous un autre point de vue. Dans ce sens, attention à ne pas confondre avec les points saillants en référence à la saillance visuelle. En effet, les points saillants sont des points d'intérêt au sens perceptuel du terme, ils seront ceux qui vont être perçus rapidement et par la plupart des observateurs. Même si certains des points saillants correspondront à la définition de point pertinent, ce ne sera pas le cas de tous. Cette correspondance est, de plus, très dépendante de la tâche demandée à l'observateur. La concordance entre les points pertinents et les points saillants, devient importante lorsque le résultat de détection des points doit être jugé par un humain puisque dans ce cas, cet humain s'attendra à trouver une cohérence entre points pertinents et saillants.

Il existe différents type de points d'intérêt, nous trouvons les coins, les *blob* par exemples. Ici l'appellation coin et *blob* est générique. Le mot coin fait donc référence aux formes de type coins, croix forme en T etc. Un *blob* correspond ici à une zone relativement uniforme présentant un contraste avec l'extérieur de la zone. Dans la suite du chapitre, ces mots garderont le sens plus large que nous venons de présenter pour faciliter la lecture.

L'un des premiers détecteurs de points d'intérêt est celui proposé par Harris and Stephens (1988). Ils cherchent à extraire des coins et se basent sur la matrice d'auto-corrélation des dérivées spatiales que nous avons vu dans le chapitre précédent. La figure 2.1 reprend les étapes élémentaires proposées par les auteurs pour la création d'un détecteur recherchant les coins. Les coins sont des zones de gradient à forte amplitude, la première étape cherche donc à les extraire. La seconde prend en compte les interactions spatiales grâce à la matrice d'auto-corrélation spatiale. Harris and Stephens (1988) justifient la troisième étape comme un filtrage nécessaire pour diminuer le bruit du à leur choix de filtre dérivateur. Ils mesurent ensuite une fonction indiquant la probabilité que le pixel mesuré représente un coin. Les maxima locaux de cette fonction sont les coins présents dans l'image.

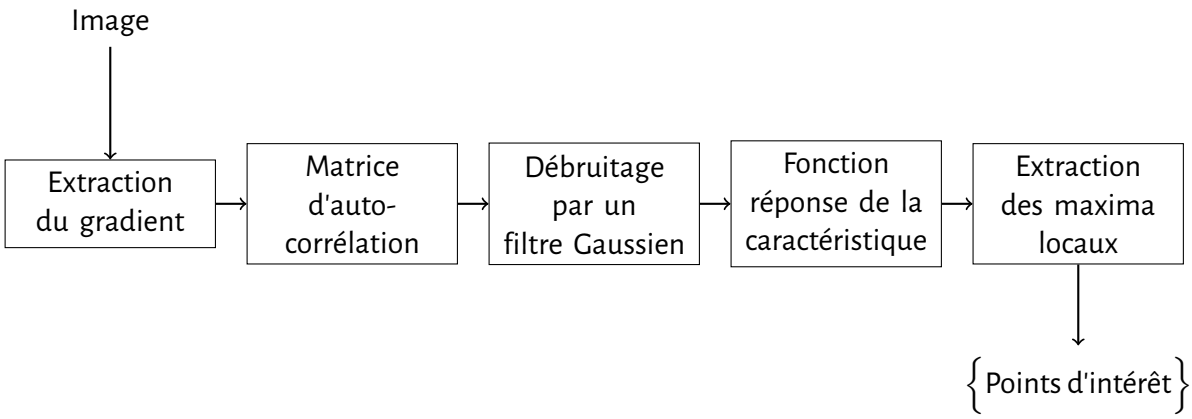


FIGURE 2.1 – **Décomposition du détecteur de Harris** en étapes élémentaires.

Nous proposerons dans ce chapitre d'adapter ce détecteur à une analyse couleur multi-échelles pour l'extraction de coins. Cela nous amènera à nous interroger sur la nécessité et le placement de ces étapes dans notre contexte générique multi-canaux respectant la physique de l'acquisition.

L'ensemble du chapitre s'organise comme suit : tout d'abord une section présentant l'état de l'art. Cet état de l'art justifie la proposition d'une nouvelle écriture plus générique et précise. Nous expliquons celle-ci dans la section suivante. Enfin, nous présentons des résultats de ce détecteur dans trois contextes. Tout d'abord, nous étudions les détecteurs avec la base de données artificielles basées sur des nuages de points Gaussiens $3D$ présentée dans le chapitre précédent. Puis nous nous comparons à l'état de l'art avec la base Oxford (référence en terme de comparaison de détecteurs de points d'intérêt). Enfin nous comparerons ce détecteur au résultat obtenu lors d'une expérience psychovisuelle utilisant un outil de suivi du regard sur des images d'œuvres médiévales.

2.1 État de l'art

Tuytelaars et al. (2008) proposent une vue d'ensemble des détecteurs de point d'intérêt. Ils les classent en quatre familles principales : les détecteurs de coins, ceux de *blob*, les détecteurs de régions et les détecteurs efficients. Les premiers permettent d'extraire les coins. Les seconds détectent des zones approximativement constantes correspondant aux *extrema* d'une dérivée seconde. La troisième famille extrait des régions d'intérêt. Les auteurs proposent une quatrième famille basée sur l'objectif d'efficacité calculatoire du détecteur, et non sur sa méthode. Nous avons choisi de ne garder que les trois premières car la classification de la dernière est différente des trois premières et chaque élément de cette famille peut entrer dans une des familles précédentes.

Tuytelaars et al. (2008) mettent aussi l'accent sur certains abus de langage, notamment l'appellation point d'intérêt lorsque nous allons décrire ce point grâce à son voisinage. Il soulignent aussi l'appellation “invariance” qui est un abus de langage. En effet, le point d'intérêt détecté ne doit pas être invariant aux transformations mais produire le même résultat quelles que soient les transformations. Par exemple, nous souhaitons qu'un coin détecté dans une image soit de nouveau détecté dans cette même image tournée de $90°$. Le point lui aussi aura tourné de $90°$. On parle de covariance si la transformation commute avec la fonction de détection. Compte-tenu des habitudes de langage et de la confusion que peut impliquer le terme covariance, nous continuerons par la suite à employer le terme “invariances”.

Ces invariances sont très recherchées pour rendre le détecteur (ou le descripteur) plus robuste. En effet, la notion d'invariances est aussi vieille que celle des attributs et des opérateurs de traitement des images. L'ambition de l'outil ultime serait d'être invariant à toute transformation/modification de l'objet à analyser. Mais ce faisant, l'outil d'analyse deviendrait incapable de discriminer des variantes de l'objet. Donc le choix des invariances à conserver est profondément lié à la nature du traitement.

Mikolajczyk et al. (2005) proposent un comparatif des détecteurs de points d'intérêt. Ils comparent des méthodes de chacune des familles et s'intéressent à la détection multi-échelles. Nous utiliserons les deux critères qu'ils proposent lors de l'analyse de nos propositions.

Lindeberg est un auteur spécialisé dans l'analyse multi-échelles pour la détection de points d'intérêt. Dans Lindeberg (1994), l'auteur argue que l'utilisation de filtres Gaussiens de différentes variances est la méthode la plus adaptée pour une étude générique de la taille d'une zone d'intérêt. L'argumentaire s'appuie sur les travaux de Young qui modélisa le cortex visuel des mammifères par des dérivées de Gaussiennes. Lindeberg présente aussi dans cet article d'autres méthodes pour l'analyse multi-échelles utilisant une décomposition en arbre quaternaire ou en ondelettes de l'image. Lindeberg (1998) propose une méthode d'analyse multi-échelles pour l'extraction de points d'intérêt. L'image est convoluée avec un filtre de type Laplacien de Gaussienne pour différentes valeurs de variances.

En couleur

La détection de points d'intérêt, qu'il agisse de détection de coins ou de régions se base sur une rupture d'uniformité dans l'image. L'hypothèse faite par la plupart des auteurs est de considérer qu'une rupture suffisamment forte pour être un point d'intérêt sera nécessairement détectée en niveaux de gris. C'est pourquoi la plupart des détecteurs de points d'intérêt présents dans la littérature ne s'intéressent pas à la couleur. Or, revenons sur l'hypothèse initiale s'intéressant à la rupture d'uniformité. Quelle définition pour rupture d'uniformité ? Prenons par exemple une photo de champ de coquelicots en fleur. La luminosité de l'image sera quasiment uniforme sur tout le champ. Aucun gradient franc n'apparaîtra donc en niveaux de gris. Or, excepté pour un daltonien, tout le monde percevra le gradient entre les fleurs rouges et les tiges/feuilles vertes qui sont adjacentes à la fleur ? McIlhagga and Mullen (2018) montrent que les bords couleurs sont perçus par les humains. L'intégration de la couleur semble donc une étape importante de la détection de points d'intérêt, pour des images de complexité supérieure à la plupart des images multimédia.

Gouet et al. (2000) proposent une évaluation des détecteurs couleurs. Les détecteurs sont évalués sur la répétabilité de la détection des points d'intérêt et dans la localisation de ceux-ci. Certains auteurs (Montesinos et al. (1998); Ruzon et al. (2001); Forssén (2007)) ont fait des propositions de détecteurs couleur.

La figure 2.2 présente une organisation de l'état de l'art des détecteurs couleurs selon la nature du détecteur et le type d'approche couleur utilisé. Nous les retrouvons dans les trois catégories présentées précédemment : coin, *blob* et régions ainsi que les sous-catégories déjà existantes en niveaux de gris. La prise en compte de l'information couleur dans ce schéma se lit au niveau du type d'approche utilisé par les auteurs : principalement CMA ou $CCMA$. L'approche marginale (CMA pour *colour marginal approaches*), employée par la plupart des auteurs, calcule chaque canal comme un niveaux de gris avant de les combiner pour extraire l'information souhaitée. D'autres auteurs incorporent la corrélation entre les canaux à l'aide d'une approche inter-canaux ($CCMA$ pour *cross-channel marginal approaches*). Derrière cette écriture, nous retrouvons la volonté de prendre en considération l'aspect complexe, vectoriel de la couleur. À notre connaissance aucune approche vectorielle de la couleur n'a été proposée pour l'extraction de points d'intérêt.

2.1.1 Détecteurs de coins

La famille de détecteurs basés sur l'extraction de coins, points selles, points en T se répartit en deux sous-familles : la détection basée gradient et celle basée comparaison de pixels.

2.1.1.1 Détecteurs de coins basés gradients

Un des détecteurs les plus connus de la sous-famille basée gradient est Harris and Stephens (1988). Ce détecteur se base sur la matrice d'auto-corrélation des dérivées spatiales premières . Le premier à utiliser les matrices d'auto-corrélation pour la détection de points d'intérêt est Moravec (1980). Cette matrice d'auto-corrélation est définie par :

$$ACor = \begin{pmatrix} \left\| \frac{\partial I(x,c)}{\partial x_1} \right\|_2^2 & \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle \\ \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle & \left\| \frac{\partial I(x,c)}{\partial x_2} \right\|_2^2 \end{pmatrix}. \tag{2.1}$$

Nous pouvons faire le lien ici avec la matrice de Di Zenzo (1986) qui propose une mesure de gradient couleur à partir de cette formulation. La matrice de l'équation (2.1) s'applique en niveaux de gris dans

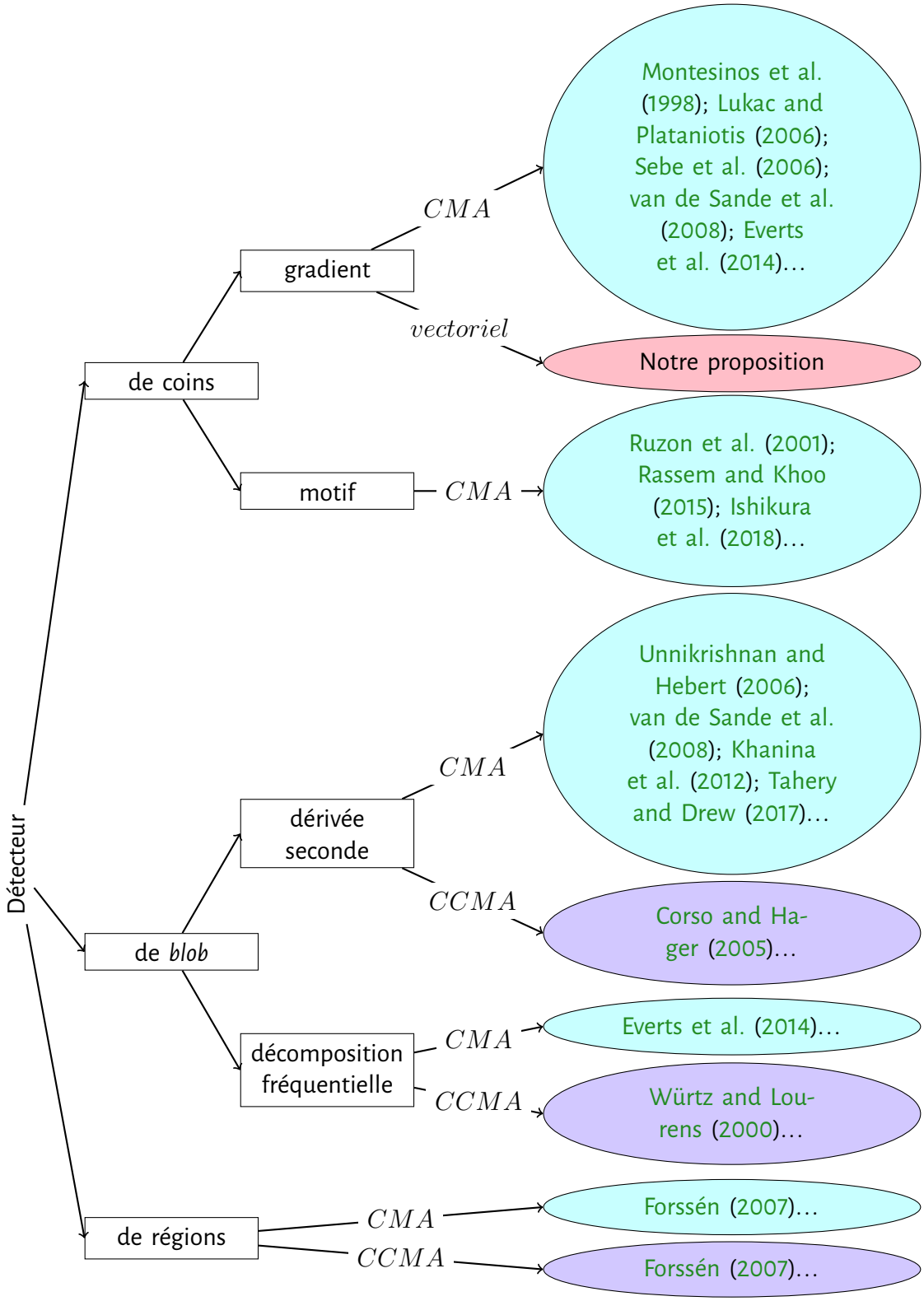


FIGURE 2.2 – **État de l'art des détecteurs couleur** organisé en fonction du type des attributs du détecteur et de la nature du traitement couleur effectué : CMA *combined marginal approaches*, *i.e.* une analyse marginale ; $CCMA$ *cross-channel marginal approaches*, l'approche prend en compte les interactions entre canaux.

la proposition de Harris and Stephens (1988). Cependant, ils s'appuient sur la même notion de dérivée conduisant à la notion de rupture lorsque le gradient est de forte amplitude.

Harris and Stephens (1988) expliquent que selon le rapport de grandeur entre les deux valeurs propres, nous sommes en présence d'une zone uniforme, de bords ou de coins :

$$\begin{aligned} \text{zone uniforme} &: \lambda_+ \approx \lambda_- \approx 0; \\ \text{bord} &: \lambda_+ \gg \lambda_-; \\ \text{coin} &: \lambda_+ \approx \lambda_- \gg 0. \end{aligned} \tag{2.2}$$

En effet, un bord correspond à un gradient de forte amplitude dans la direction opposée au bord et un gradient de faible amplitude dans le sens du bord, donc à deux valeurs propres de grandeurs très différentes. Une zone uniforme est une zone sans variation donc avec un gradient faible dans toutes les directions : deux valeurs propres proches de zéro. Au contraire, un coin se définit par deux bords de directions différentes induisant deux valeurs propres très supérieures à zéro.

Ils définissent une fonction permettant d'évaluer leur relation et fixent un seuil au dessus duquel le pixel sera qualifié de point d'intérêt :

$$\begin{aligned} &\det(ACor) - k\,\mathrm{tr}(ACor)^2 \\ &\text{avec} \quad \det(ACor) = \lambda_+\lambda_-, \\ &\text{et} \quad \mathrm{tr}(ACor) = \lambda_+ + \lambda_-, \end{aligned} \tag{2.3}$$

où k est un paramètre empirique non défini dans Harris and Stephens (1988). Montesinos et al. (1998) proposent d'utiliser la valeur $k = 0.04$ sans plus de commentaires.

De nombreuses adaptations à ce détecteur ont été publiées. Shi and Tomasi (1993) utilisent la matrice d'auto-corrélation mais ne conservent que la valeur propre minimale en justifiant que si celle ci est suffisamment grande, alors comme le prédisent Harris and Stephens (1988), nous sommes en présence d'un coin. Mais comment se définit simplement “suffisamment grande” et de façon générique. Forlenza et al. (2012), plus récemment, emploient le détecteur de Harris pour une application embarquée en temps réel dans un drone.

Le détecteur de Harris and Stephens (1988) n'est invariant qu'à la rotation. C'est pourquoi Mikolajczyk and Schmid (2001) proposent une adaptation multi-échelles de ce détecteur. Ils extraient les points d'intérêt détectés par la méthode de Harris à chaque échelle choisie. Ils sélectionnent ensuite les points à conserver et la taille à leur associer grâce à la méthode d'extraction multi-échelles basée sur les Laplaciens de Gaussiennes (présentée dans la sous-section suivante). Nous avons schématisé leur proposition dans la figure 2.3. Deux approches antagonistes (*blob* et coins) sont combinées sur le même point d'intérêt de façon empirique. Est-ce pertinent ? Ce détecteur multi-échelles, pour autant efficace, est souvent utilisé et communément appelé Harris-Laplace. Afin d'améliorer encore l'efficacité de ce détecteur Mikolajczyk and Schmid (2002) proposent une version invariante aux transformations affines.

En couleur

Parmi les premiers auteurs à proposer un détecteur de points d'intérêt couleur, nous retrouvons Montesinos et al. (1998). Ils utilisent la formulation de Harris and Stephens (1988) en exploitant l'expression vectorielle de la couleur proposée par Di Zenzo (1986), avec les limitations qu'elle implique. Ils travaillent dans l'espace RGB. D'autres auteurs, ont employé cette méthode sur d'autres espaces couleur : Sebe et al. (2006) exploitent l'espace des couleurs opposées et celui des quotients de couleur présenté par Gevers and Smeulders (1999) ; van de Weijer and Gevers (2004) utilisent l'espace des invariants couleur. Le chapitre 9 du livre Lukac and Plataniotis (2006) est dédié à ce type d'approche. Everts et al. (2014) adaptent cette approche couleur du détecteur de coins de Harris and Stephens

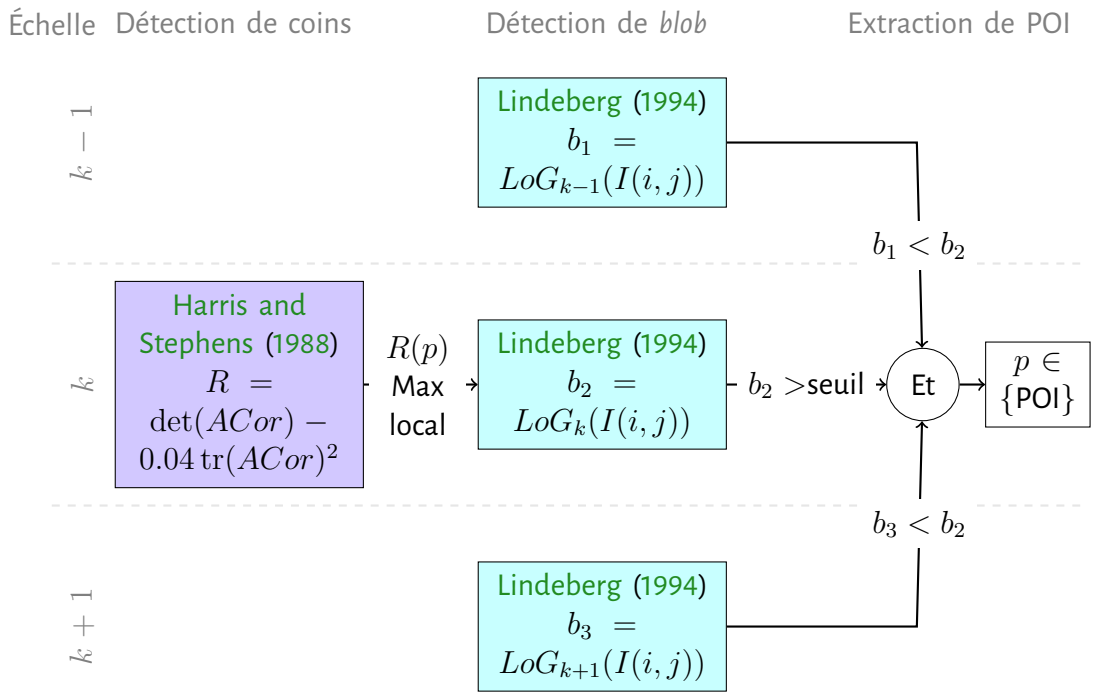


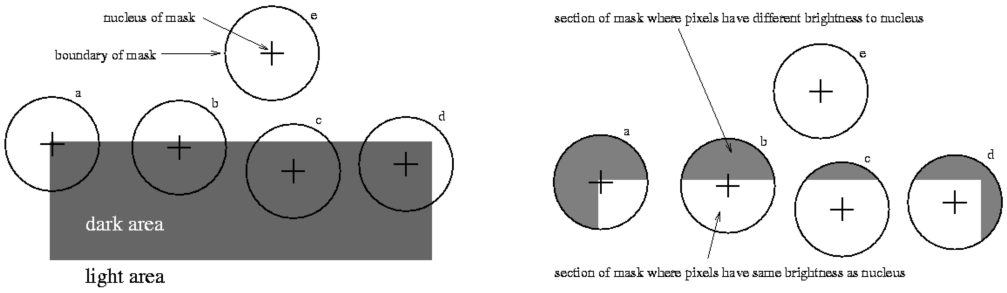
FIGURE 2.3 – **Schéma de fonctionnement du détecteur multi-échelles Harris-Laplace** proposé par Mikolajczyk and Schmid (2001).

(1988) à la vidéo en ajoutant la couche temps. Le calcul couleur se fait dans l'espace des couleurs opposées pour approcher l'espace des invariants couleur en utilisant les ratios canal opposé sur canal luminosité et obtenir plus facilement une invariance aux ombres.

2.1.1.2 Détecteurs de coins basés motif

Les détecteurs de coins basés "comparaison de pixels" cherchent à trouver un "motif" de type coin (et assimilés) sous la forme d'un arrangement de pixel voisins.

Smith and Brady (1997) introduisent les détecteurs de coins basés comparaison de pixels. Ils proposent SUSAN pour *similar univalue segment assimilating nucleus*. L'idée de ce détecteur est de comparer les pixels appartenant à un noyau (support spatial) à la valeur du pixel au centre du noyau. En fonction du nombre de pixels de valeur similaire au centre nous sommes face à une zone uniforme, un bord ou un coin. La figure 2.4 illustre le propos. Cette méthode permet de détecter les bords et les coins. La recherche de minimum permet d'extraire les coins.



(a) Différents position du noyau de test (b) Comparaison au centre

FIGURE 2.4 – **Illustration du fonctionnement de SUSAN**. Images extraites de Smith and Brady (1997)

Pour améliorer le temps de calcul, Rosten and Drummond (2006) s'inspirent de SUSAN et proposent FAST (*features from accelerated segment test*). La comparaison se fait uniquement sur la bordure du noyau. Ce n'est pas sans rappeler les *local binary pattern* : comparaison du pixel central à une couronne distante de d. Les coins seraient alors les motifs avec un seul changement de valeur binaire (à la rotation prés). Pour accélérer encore la détection, les auteurs mesurent les quatre points cardinaux de cette bordure. Si ceux-ci ne répondent pas aux critères pour être qualifiés de coins alors le point est automatiquement éliminé. Attention ce critère de rapidité peut exclure certains type de coins. Un coin de type croix peut échouer au test des points cardinaux tout en restant un coin qui devrait être détecté. Mair et al. (2010) visent à améliorer encore la vitesse de calcul. Ils proposent de réduire le diamètre du noyau, puis effectuent un apprentissage pour déterminer l'ordre de comparaison des pixel. *Center surround extrema features* (CENSURE), introduit par Agrawal et al. (2008), utilisent des formes spécifiques pour faire une analyse similaire à celles des précédents auteurs. Ils emploient le principe des images intégrales dans le but d'accélérer le temps de calcul.

Une fois les coins détectés, Rosin (1999) proposent d'analyser leurs contenus. Ils permettent d'étudier l'ouverture, l'arrondissement d'un coin.

En couleur

Ruzon et al. (2001) utilisent des histogrammes de valeur de pixels sur des zones d'intérêt pour extraire les bords. Ils effectuent ces mesures dans l'espace $CIELAB$. Ils proposent une méthode similaire pour extraire des coins, moyennant une normalisation des histogrammes représentant l'intérieur et l'extérieur du coin. Ishikura et al. (2018) emploient le détecteur CENSURE de façon marginale sur l'espace $CIELAB$.

Rassem and Khoo (2015) s'inspirent du détecteur niveaux de gris utilisant les histogrammes d'intensité locaux pour l'adapter à la couleur. Pour cela, ils quantifient les couleurs par canal et concatènent les histogrammes pour ensuite extraire les points d'intérêt comme effectué en niveaux de gris. Ils effectuent les calculs sur différents espaces couleur : RGB, HSV ou encore l'espace des couleurs opposées.

Ces approches simplifient l'analyse couleur par un traitement marginal ou encore par quantification de l'espace couleur. Cela induit que nous perdons de l'information et le respect de la mesure physique.

2.1.2 Détecteurs basés *blob*

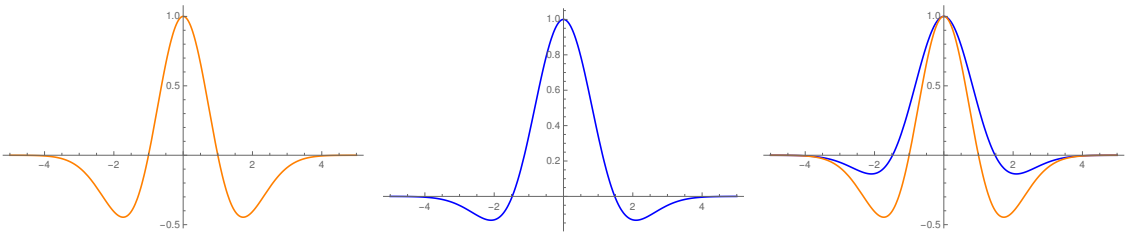
La famille de détecteurs de points d'intérêt basés sur les *blob* s'appuie sur les propriétés des dérivées secondes. Les minima et maxima locaux vont correspondre à des *extrema* locaux dans l'image initiale. C'est-à-dire au centre d'une région. Les dérivées ne sont généralement effectuées qu'après un lissage Gaussien induisant une taille implicite de la région ainsi mise en avant. Pour une Gaussienne d'écart-type σ, le *blob* détecté sera de taille $\sqrt{n\sigma}$, l'image I appartenant à $\mathbb{R}^n$.

Le premier détecteur de ce type est le Laplacien de Gaussienne (LoG pour *Laplacian of Gaussian*). La première étape consiste à appliquer un filtre Gaussien sur l'image :

$$I(x) = I(x) * G(x, \sigma), \tag{2.4}$$

où G est une Gaussienne de variance σ. Le Laplacien se définit comme la somme des dérivées secondes sur chacun des axes de l'image. La dérivation peut être indépendamment effectuée sur l'image ou sur la fonction de convolution, ce qui permet de simplifier nettement les calculs. Les *extrema* du Laplacien sont ensuite extraits comme points d'intérêt. Lindeberg (1998) propose cette méthode pour l'analyse multi-échelles.

Une approximation du Laplacien de Gaussienne est la différence de Gaussienne (DoG pour *difference of Gaussian*). Il s'agit ici d'effectuer le produit de convolution avec deux Gaussiennes de variances différentes, puis de calculer la différence entre les deux résultats. La figure 2.5 représente, en dimension un, le Laplacien de Gaussienne et la différence de Gaussienne ainsi que l'écart entre les deux. Si l'allure des courbes est similaire, la différence de Gaussienne a des coefficients négatifs beaucoup plus faible que le Laplacien de gaussienne. Cela implique que le calcul de dérivation sera moins marqué et entraînera une perte de discrimination pour les DoG. Un autre problème posé par ce type d'approximation DoG est la différence ou le rapport nécessaire entre les différents σ_i pour conserver une forme la plus proche possible du LoG. Quelle relation permet de calculer le second écart-type pour approximer le Laplacien ? Lowe (1999) emploie les différences de Gaussienne comme détecteur de



(a) Laplacien de Gaussienne ($\sigma = 1$) (b) Différence de Gaussienne($\sigma_1 = 1$ et $\sigma_2 = 1.2$) (c) Comparaison

FIGURE 2.5 – **Différence entre LoG et DoG**.

points d'intérêt avant de présenter son descripteur SIFT. Park et al. (2014) s'inspirent de la différence de Gaussienne et l'appliquent à des dérivées d'ordre supérieur.

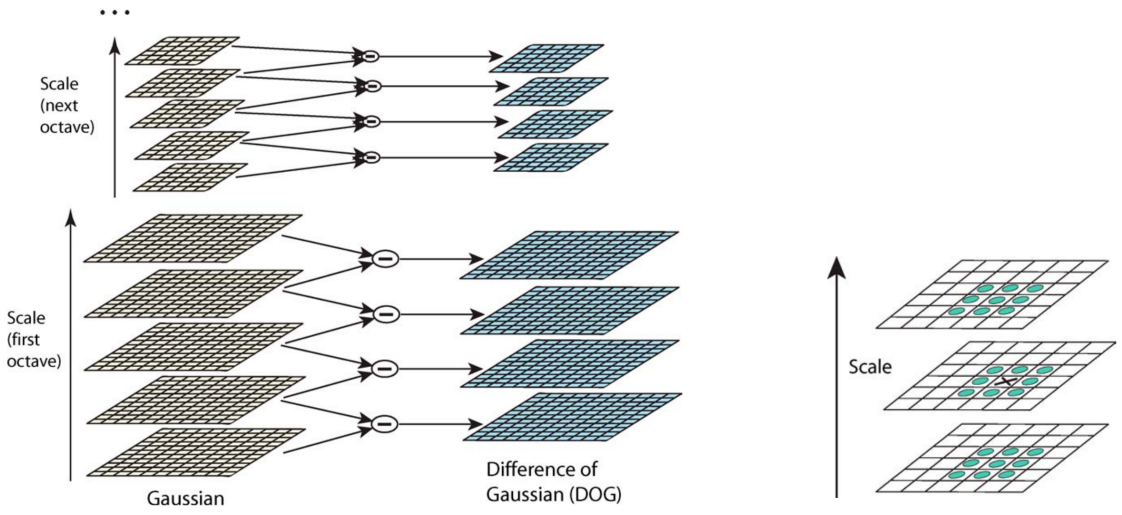
Afin d'améliorer le temps de calcul, Bay et al. (2006) utilisent les ondelettes de Haar pour approximer les différences de Gaussiennes. Fauqueur et al. (2006) proposent un détecteur de coins s'appuyant sur une décomposition en ondelettes. Terzić et al. (2015) utilisent la décomposition en ondelettes et la justifient comme étant une approximation du fonctionnement des cellules corticales pour l'extraction de points d'intérêt.

Lowe (1999) propose une analyse multi-échelles de points d'intérêt basée sur les différences des Gaussiennes. L'étude multi-échelles se fait à travers plusieurs octaves correspondant à une échelle de détection. Pour chacune de ces octaves, $k + 1$ filtres sont utilisés sur l'image permettant de mesurer k DoG (fig. 2.6a). Un point d'intérêt est détecté sur un *extremum* local d'une différence de Gaussienne lors de l'analyse à échelle unique. Ici un pixel est considéré comme point d'intérêt si la mesure de DoG à l'échelle k_i de l'octave est *extremum* non seulement sur les voisins spatiaux mais aussi sur les voisins de l'échelle $k_i - 1$ et $k_i + 1$ (fig. 2.6b).

Les approches avec filtrage Gaussien ont le désavantage d'adoucir voire éliminer les bords présents dans l'image. Pour limiter la perte d'information contenue par le bords des objets, une famille de détection multi-échelles basé sur un filtrage non-linéaire est apparue récemment. Alcantarilla et al. (2012) s'appuient sur l'équation de diffusion de Perona and Malik (1990) pour proposer un filtrage non-linéaire de l'image puis détecter les points d'intérêt à l'aide de la matrice Hessienne sur chaque échelle. Feng et al. (2016) emploient la même méthode mais proposent un algorithme différent d'optimisation du filtrage non-linéaire.

En couleur

Corso and Hager (2005) reprennent le principe des Laplaciens de Gaussienne pour les utiliser sur des projections d'un espace couleur. Ils présentent des résultats sur l'espace RGB ou celui des couleurs opposées. Unnikrishnan and Hebert (2006) calculent le produit des LoG sur chacun des canaux comme une fonction permettant d'extraire les points d'intérêt. Ils proposent aussi l'emploi du



(a) Principe des octaves (b) Localisation de l'*extremum*

FIGURE 2.6 – **Illustration du fonctionnement multi-échelles des DoG**. Images extraites de Lowe (2004)

déterminant d'une matrice 3×3 contenant verticalement les canaux RGB et horizontalement les dérivées de l'ordre 0 à 2.

Tahery and Drew (2017) proposent de combiner une approche de type Hessienne couleur (somme des déterminants du Hessien par canal) avec l'approche de Harris and Stephens (1988) en couleur. Les approches basées sur les dérivées premières et celles sur les dérivées secondes ne mettent pas en avant les mêmes caractéristiques, la combinaison de telles approches est-elle pertinente ? Khanina et al. (2012) proposent une extension couleur du Hessien pondérant les canaux à l'aide d'une valeur moyenne dépendant du Laplacien marginal de chaque canal. Cela implique que le résultat est dépendant du contenu de l'image. Le Hessien résultant est de type *blob* si le rapport entre les deux valeurs propres est supérieur à un seuil reprenant ici l'idée de détection des bords de Harris and Stephens (1988). Quelques articles proposent une extraction de points d'intérêt sur des images hyper-spectrales. Mukherjee et al. (2009) réduisent la dimension à l'aide d'une analyse en composante principale, puis réalisent une combinaison empirique des données pour extraire les points d'intérêt. Dorado-Munoz et al. (2012) proposent d'adapter la méthode des différences de Gaussiennes pour des images hyper-spectrales en employant un ordonnancement lexicographique. Cet ordre favorise la première longueur d'ondes au détriment des autres.

Everts et al. (2014) proposent un détecteur basé sur les filtres de Gabor calculés de façon marginale sur les canaux de l'espace des couleurs opposées. S'inspirant du fonctionnement du système visuel des primates, Würtz and Lourens (2000) utilisent la décomposition grâce aux filtres de Gabor pour modéliser les cellules corticales. Ils exploitent cette décomposition sur l'espace des doubles opposés, en marginal, pour extraire des points d'intérêt. Si la seconde proposition essaye de prendre en compte la corrélation entre les canaux, grâce à l'espace des doubles opposés, ce n'est pas le cas de la première. La méthode de Würtz and Lourens (2000) analyse un système visuel et adapte des outils pour s'en approcher sans apporter plus de preuve à leur modèle.

2.1.3 Détecteurs basés régions

Ce type de détecteurs extrait des régions d'intérêt et non des points d'intérêt. Ces régions sont souvent modélisées par des ellipses impliquant une invariance aux transformations affines.

Tuytelaars and Van Gool (2004) proposent deux méthodes permettant d'extraire des régions d'intérêt. Dans les deux cas, la région s'initialise pour un point puis s'étend pour obtenir la région complète. La première proposition initialise les régions avec les points d'intérêt détectés par la méthode de Harris and Stephens (1988). Puis elle propage la région en suivant les bords. Cette méthode n'est pas très stable : les bords n'étant pas toujours visibles ou détectables d'une image à l'autre. La seconde méthode s'initialise sur les *extrema* locaux en intensité de l'image. Elle parcours ensuite la zone angulairement cherchant l'*extremum* d'une fonction de coût sur chaque rayon. La zone ainsi délimitée est approximée par une ellipse qui devient la région d'intérêt .

La méthode MSER (*maximally stable extremal regions*), proposé par Matas et al. (2004), se base sur une ligne de partage des eaux. Cela crée des zones de pixels uniformes et connectés. La région est qualifiée de maximale si la différence entre les zones détectées pour différentes lignes de partage des eaux est suffisamment faible.

Mikolajczyk et al. (2005) ont montré, dans leur étude comparant les détecteurs, que MSER était sensible au flou. Une proposition de Perdoch et al. (2007) pallie ce problème. Ils ne conservent que les chemins fermés et calculent l'ellipse approchant comme caractéristique de la zone d'intérêt. Martins et al. (2016) ajoutent une carte de saillance pour la détection MSER afin d'améliorer la couverture des zones détectées sur l'image.

Les approches basées région détectent des régions sans contrainte de taille. Elles sont donc multi-échelles de façon intrinsèque.

En couleur

Forssén (2007) propose deux adaptations couleur du détecteur de régions MSER. La première s'inspire de la proposition de Corso and Hager (2005) et calcule les MSER de façon marginale sur l'espace des couleurs opposées. La deuxième adaptation essaye d'étendre la méthode de ligne de partage des eaux à la couleur. Pour cela, ils modélisent les données RGB par une loi de Poisson, puis utilisent la distance du χ^2 entre le pixel d'intérêt et les pixels déjà présents dans la région. Cette famille d'approche nécessite de définir la notion d'uniformité dans une image couleur. Toutes ces approches s'affranchissent de traiter la question sous son aspect physique et/ou perceptuel.

2.2 Détecteurs de points d'intérêt couleur

Nous proposons dans cette section une adaptation du détecteur de Harris au multi-échelles et à la couleur. En effet, l'état de l'art nous a montré le manque d'approches respectant la contrainte de métrologie ou prenant en compte la couleur de façon vectorielle.

La figure 2.7 fait le parallèle entre la proposition de Harris and Stephens (1988), son adaptation en niveaux de gris dans la librairie *OpenCv* et notre proposition. Les étapes en bleu sur le schéma ne sont pas modifiées d'une proposition à l'autre. Nous remarquons que certaines étapes diffèrent entre la proposition initiale et l'implémentation dans *OpenCV*. Cela nous a amené à nous questionner sur l'objectif de ces étapes et comment les adapter le plus rigoureusement possible. Ces étapes sont en jaune sur la figure 2.7. En vert, nous trouverons les corrections que nous avons apportées au détecteur afin de prendre en compte la couleur et d'extraire une fonction réponse multi-échelles.

Nous présenterons tout d'abord le choix du filtre dérivateur que nous avons fait pour l'adaptation multi-échelles, puis l'insertion de la couleur à travers l'emploi du *full-vector gradient* présenté dans le chapitre précédent. Ces deux étapes correspondent à l'extraction du gradient. Ensuite, nous présenterons l'étape de mise en forme présentée dans la figure 2.7. Enfin, nous présenterons les choix de sélection de points d'intérêt à conserver et le poids à leur assigner.

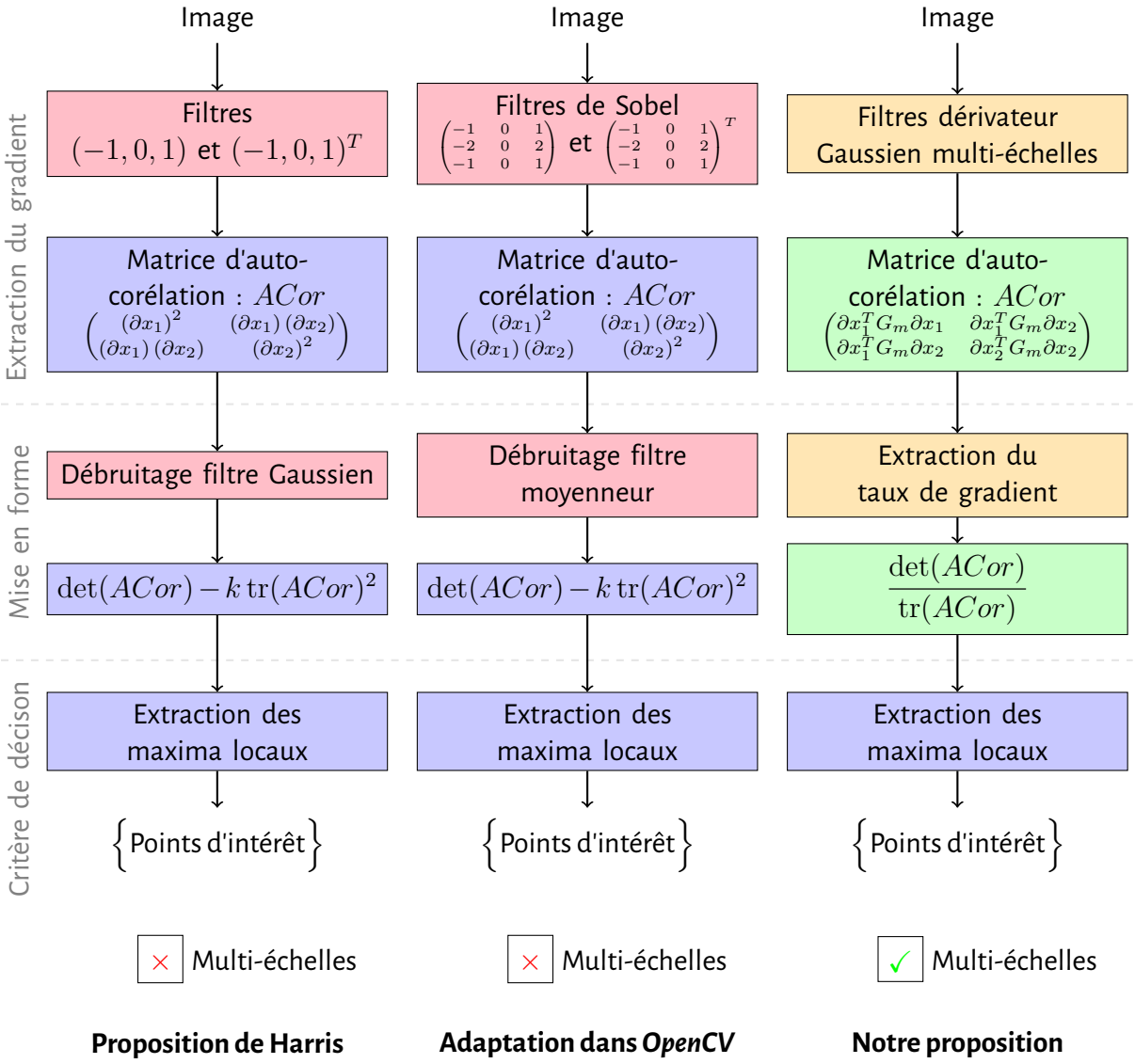


FIGURE 2.7 – **Comparaison des étapes de calcul du détecteur de Harris** entre la proposition initiale de Harris and Stephens (1988), l'implémentation dans *OpenCV* et notre proposition.

2.2.1 Adaptation aux multi-échelles

2.2.1.1 Choix du filtre de dérivation

La première étape de la détection de coins pour le détecteur de Harris and Stephens (1988) est l'extraction des gradients verticaux et horizontaux. Harris propose deux approximations de dérivées centrées, l'adaptation dans *OpenCV* utilise les filtres de Sobel : le produit d'un filtre dérivateur avec un filtre intégrateur orthogonal. Afin d'être le plus générique possible, nous choisissons de modéliser la dérivée par un filtre dérivateur de type Gaussien :

$$\frac{\partial Filter}{\partial x_i} = \begin{pmatrix} \\ \\ \\ \end{pmatrix} \times \begin{pmatrix} \\ \\ \\ \end{pmatrix}^T . \tag{2.5}$$

Celui-ci est constitué de la dérivée d'une Gaussienne selon l'axe souhaité et d'une Gaussienne dans le sens perpendiculaire à celle-ci, pour limiter le bruit qui pourrait être présent dans le signal. Mathématiquement, son écriture est :

$$\frac{\partial Filter}{\partial x_i} = \left(\frac{\partial}{\partial x_i} G(x_i)\right) \times \left(G(x_i)\right)^T, \quad \text{où} \quad G(x_i) = \exp\left(\frac{-x_i^2}{2\sigma^2}\right). \tag{2.6}$$

Une Gaussienne contient 99.7% des données sur un intervalle de six écarts-types autour de la moyenne de la Gaussienne. Le support spatial du filtre est donc choisi pour conserver ces 3σ de part et d'autre du centre de notre filtre. Cela nous donne sa taille :

$$(6\sigma + 1) \times (6\sigma + 1). \tag{2.7}$$

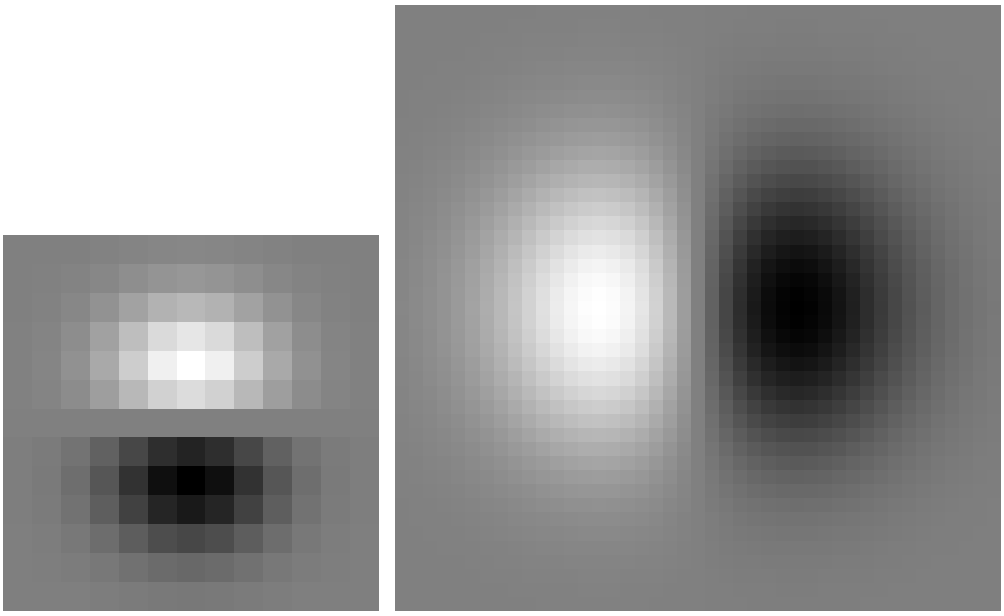
Puisque nous sommes dans le domaine de l'image, le centre sera nécessairement un pixel. Pour garder un filtre symétrique par rapport au centre, nous ajoutons le pixel central ($+1$ dans la taille du filtre). La construction proposée nous donne ainsi accès à une écriture naturelle du processus multi-échelles.

2.2.1.2 Vers le multi-échelles

Nous obtenons une analyse multi-échelles en sélectionnant différentes valeurs d'écarts-types pour les fonctions Gaussiennes impliquant des tailles de support différentes. La figure 2.8 présente les images de deux tailles de filtre dans des orientations différentes.

De nombreux articles (Lindeberg (1994); Lowe (1999),etc.) sélectionnent les écarts-types de façon à ce que leurs rapports soit constants. Lindeberg (1994) écrit :

This result can also be obtained directly from scale invariance and the Pi-theorem.



(a) Filtre dérivateur vertical avec $\sigma = 2$

(b) Filtre dérivateur horizontal avec $\sigma = 7$

FIGURE 2.8 – **Exemples de filtres dérivateurs Gaussiens**.

Cette phrase suggère que le rapport entre les aires qui correspondent à la taille des filtres soit constant et supérieur à 1. L'aire étant induite par un rapport entre les tailles de filtre $k\sigma$. Naturellement entre deux échelles consécutives, le rapport d'aire amène à $k_i = \sqrt{cst}k_{i+1}$. La première valeur de constante possible est donc 2, d'où le choix de 1.4 proposé dans certains articles et *OpenCV*. Même si la proposition de Lindeberg (1994) est intéressante intellectuellement, nous nous interrogeons sur la nécessité de respecter ou non ce rapport entre les échelles de filtres. Quel que soit le rapport choisi, il est connu, l'aire des filtres également, une normalisation est donc facilement accessible.

Dans cette première approche d'un détecteur de points d'intérêt, nous ne faisons aucune hypothèse sur les tailles de filtres. Les écarts-types s'incrémentent de un en un pour obtenir de nombre d'échelles souhaité. Le nombre d'échelles et la taille des filtres dépendent de la taille de l'image étudiée et du type d'information que nous souhaitons extraire de l'image. Pour définir justement ces paramètres, il faudrait effectuer des expériences sur des images théoriques contenant des objets de tailles différentes.

2.2.2 Adaptation à la couleur

Après l'extraction du gradient, Harris and Stephens (1988) utilisent la matrice d'auto-corrélation spatiale des gradients extraits. Di Zenzo (1986) utilise cette même matrice et propose une adaptation couleur. Dans le chapitre précédent, nous avons proposé d'adapter l'écriture de Di Zenzo (1986) afin de prendre en compte la corrélation entre les canaux. Nous proposons ici de reprendre cette écriture, permettant une adaptation couleur, multi et hyper-spectrale de la détection de coins. Pour cela, nous rappelons la définition de la nouvelle matrice d'auto-corrélation :

$$ACor = \begin{pmatrix} \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_1} \right\rangle & \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle \\ \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle & \left\langle \frac{\partial I(x,c)}{\partial x_2}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle \end{pmatrix}. \tag{2.8}$$

Le produit scalaire utilisé entre les dérivées spatiales marginales est incomplet. Pour prendre en compte la non-orthogonalité des canaux, le produit scalaire s'écrit à l'aide de la matrice de Gram :

$$\left\langle \frac{\partial I(x,c)}{\partial x_k}, \frac{\partial I(x,c)}{\partial x_l} \right\rangle = \frac{\left(\frac{\partial I(x,c)}{\partial x_k}\right)^T G_m \left(\frac{\partial I(x,c)}{\partial x_l}\right)}{\|G_m\|_\infty}, \tag{2.9}$$

où la matrice de Gram se définit par :

$$Gm = \begin{pmatrix} \|s_0\|_2^2 & \langle s_0, s_1\rangle_2 & \dots & \langle s_0, s_m\rangle_2 \\ \langle s_1, s_0\rangle_2 & \|s_1\|_2^2 & \dots & \langle s_1, s_m\rangle_2 \\ \vdots & \vdots & \ddots & \vdots \\ \langle s_m, s_0\rangle_2 & \dots & \langle s_m, s_{m-1}\rangle_2 & \|s_m\|_2^2 \end{pmatrix}. \tag{2.10}$$

Le produit scalaire associé à la matrice de Gram est mesuré à l'aide des courbes de sensibilité spectrale $S_i()$ associées à chaque canal s_i :

$$\langle s_i, s_j\rangle_2 = \int_{\lambda_{min}}^{\lambda_{max}} S_i(\lambda) S_j(\lambda) d\lambda. \tag{2.11}$$

2.2.3 Mise en forme

L'analyse du résultat recherché nous indique que le filtre permet d'identifier des zones correspondant aux “coins” recherchés, et de ne pas considérer les zones de bords linéaires ou curvilignes.



(a) Image initiale (b) Rapport avant filtre intégrateur (c) Rapport après filtre intégrateur

FIGURE 2.9 – **Illustration de la nécessité du filtre intégrateur**.

Clairement l'objectif de cette étage de traitement est d'établir les zones pour lesquelles le rapport entre la longueur du gradient et la surface du filtre est maximal. L'aspect intégral permet de considérer à la fois ce rapport longueur/surface et la norme du gradient en chaque position spatiale.

Harris and Stephens (1988), ainsi que les auteurs leur succédant, indiquent qu'à l'issu de la phase de calcul, un filtrage passe-bas est nécessaire. Pourtant, que ce soit dans les implémentations standards, dont celle d'*OpenCV* ou dans des articles plus récents, nous remarquons qu'un filtre passe-bas est déjà intégré dans le calcul du gradient. Un filtrage additionnel n'aurait du sens, que si sa fréquence de coupure était plus faible, donc pas une moyenne simple ou Gaussienne.

La deuxième question que nous évoquons dans cette section est la nécessité de la constante k définie empiriquement dans le calcul de la fonction réponse des méthodes reprenant l'écriture de Harris and Stephens (1988).

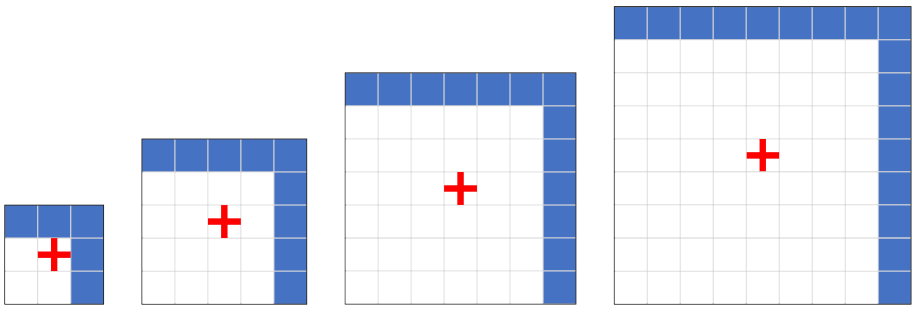
2.2.3.1 Extraction du gradient

La figure 2.7 présente une étape de débruitage à la suite du calcul de la matrice d'autocorrélation. Suite au choix de filtre que nous avons fait cette étape semblait inutile à notre approche.

L'expérience nous a appris que cette étape ne peut être supprimée du processus. Cependant, elle n'a rien d'une étape de débruitage. Il s'agit ici d'extraire les coins de la fonction réponse. En effet, un filtre de débruitage est un filtre passe-bas permettant d'extraire l'information moyenne présente sur la fenêtre du filtre. Ce filtrage permet ici de sommer les gradients présents dans la fenêtre et d'obtenir ainsi une valeur plus grande sur les zones de coin puisque la quantité de gradients de fortes amplitudes sera supérieure à une zone ne contenant qu'un bord droit par exemple.

Nous avons choisi pour cette étape d'utiliser un filtre intégrateur de la taille du filtre dérivateur employé. La figure 2.9 présente l'image de la fonction réponse avant et après le filtre intégrateur que nous employons. Si l'image de la figure 2.9b semble extraire majoritairement les bords de l'image initiale (fig. 2.9a), ce n'est pas le cas de l'image après filtrage (fig. 2.9c). Sur cette dernière, les zones correspondant aux dix coins attendus dans l'image étoile donnent des valeurs plus élevées que le reste de l'image.

Il y a deux limites à cette extraction de coins. Tout d'abord, puisque nous cherchons les maxima, les coins détectés seront décalés avec la taille des filtres comme le montre la figure 2.10a. La deuxième limite concerne les zones détectées. Une grande taille du filtre dérivateur peut être vue comme la perception de l'objet à une distance plus grande. Avec cette hypothèse, le centre de l'étoile devient le centre d'une forme en $*$ et est donc un coin à détecter (fig. 2.10b). Mais les deux familles se doivent-elles d'être exclusives ?



(a) Décalage du points d'intérêt avec la taille du filtre



(b) Somme de plusieurs coins pour une grande taille de filtre

FIGURE 2.10 – **Limites de l'utilisation du filtre intégrateur**.

Les filtres dérivateurs associés à ces filtres intégrateurs permettent d'obtenir une fonction réponse dont les ordres de grandeur dans les échelles sont similaires. Nous pouvons donc fixer un seuil identique pour chaque échelle. En faisant l'hypothèse que toute image contient des points d'intérêt, nous choisissons ce seuil à 20% du maximum de la fonction réponse sur toutes les échelles.

2.2.3.2 Fonction réponse

La méthode de Harris and Stephens (1988) utilise cette fonction réponse :

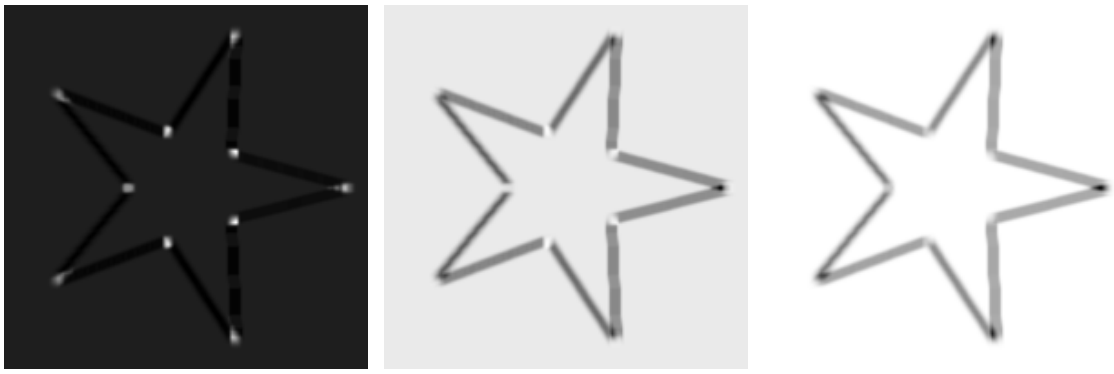
$$f_R = \det(ACor) - k\,\mathrm{tr}(ACor)^2, \tag{2.12}$$

où k est une constante mal définie. La figure 2.11 montre les résultats de cette fonction réponse pour différentes valeurs de k. D'une valeur à l'autre, nous pouvons ou non extraire les informations souhaitées. Dans le cas de cette image test, la valeur $k = 0.04$ donne des résultats satisfaisants. Cependant, avec des images réelles et différentes tailles de filtres, les réponses obtenues ressemblent davantage à la figure 2.11c, qui ne permet pas d'extraire les coins. En fonction des valeurs propres extraites de la matrice d'auto-corrélation, la constante k à choisir varie beaucoup.

Nous avons cherché une formulation qui nous permette de nous absoudre de cette constante. L'hypothèse initiale de Harris and Stephens (1988) pose que les coins sont les points où les deux valeurs propres sont grandes. Sous cette hypothèse le produit des valeurs propres sera supérieur à leur somme. Nous proposons donc la fonction réponse suivante, qui ne nécessite aucune constance à définir empiriquement :

$$f_R = \frac{\det(ACor)}{tr(ACor)}. \tag{2.13}$$

En plus d'une fonction réponse, nous pouvons ajouter l'angle entre un vecteur propre associé à la plus grande valeur propre, et l'axe horizontal comme orientation principale de la zone d'intérêt.



(a) Réponse pour $k = 0.04$ (b) Réponse pour $k = 0.2$ (c) Réponse pour $k = 0.4$

FIGURE 2.11 – **Instabilité de la constante** k.



(a) Image réelle (987 points détectés) (b) Image de synthèse (62 points détectés)

FIGURE 2.12 – **Résultat sans sélection de points d'intérêt**.

Rappelons le calcul d'angle :

$$\theta = \begin{cases} \text{Indéfini si } \left(\left\| \frac{\partial I(x,c)}{\partial x_1} \right\|^2 - \left\| \frac{\partial I(x,c)}{\partial x_2} \right\|^2 \right)^2 + (\tau)^2 = 0 \\ \text{sign}\,(\tau) \arcsin \left(\dfrac{\lambda_+ - \left\| \frac{\partial I(x,c)}{\partial x_1} \right\|^2}{2\lambda_+ - \left\| \frac{\partial I(x,c)}{\partial x_1} \right\|^2 - \left\| \frac{\partial I(x,c)}{\partial x_2} \right\|^2} \right) \\ \text{avec } \tau = \left\langle \frac{\partial I(x,c)}{\partial x_1}, \frac{\partial I(x,c)}{\partial x_2} \right\rangle \text{ sinon.} \end{cases} \tag{2.14}$$

2.2.4 Sélection des “bons” points d'intérêt

À cette étape, notre détecteur de points d'intérêt donne les résultats de la figure 2.12. Les quinze filtres sont de taille $\sigma_i = i$. Tous les maxima locaux sont gardés. Tout d'abord nous observons bien la délocalisation du point d'intérêt avec la croissance de l'échelle. Le centre de l'étoile devient un point d'intérêt pour les très grandes échelles comme nous l'avons mentionné plus haut. Nous pouvons aussi noter la très bonne répétabilité du détecteur ainsi conçu à travers les échelles.

Il faut maintenant, sélectionner les “bons” points d'intérêt. Nous avons déjà dit que la localisation était plus précise sur les plus petites échelles ce qui est confirmé par la figure 2.12. Cependant, la taille de la zone d'intérêt n'est pas nécessairement celle associée à la plus petite échelle.

Dans l'état de l'art (Lowe (2004); Bay et al. (2008), etc.), la sélection de la bonne échelle se fait sur une recherche de maximum sur plusieurs échelles (fig. 2.6). Le choix de trois échelles consécutives dont le pixel central est maximum ne met pas en avant la bonne répétabilité du détecteur dans les échelles. En effet, un point détecté sur six échelles consécutives, pourra être détecté sur une à deux échelles. S'il est détecté deux fois l'information est redondante, s'il ne l'est qu'une fois comment être sûr qu'il soit qualifié de point d'intérêt à fort potentiel.

Nous souhaitons donc effectuer une autre sélection. Compte-tenu de la dérive de la localisation avec l'augmentation de l'échelle, conserver la position de détection de la plus petite échelle semble logique. Le nombre de fois où un pixel est détecté dans les échelles doit être un indicateur de la capacité à détecter, de nouveau, ce point dans une autre configuration d'image. C'est donc une information que nous souhaitons conserver. Pour cela, il est important de pouvoir suivre le point d'intérêt dans les échelles malgré sa dérive spatiale selon celles-ci. Parmi les critères possibles de sélection, nous pouvons prendre en compte l'angle associé à la zone d'intérêt qui reste fixe sur les échelles associées au même point initial. Une fois chaque point traqué dans les échelles, nous pouvons regarder quel centre est maximum sur les échelles pour déterminer l'échelle de la zone d'intérêt.

Les critères que nous souhaitons utiliser pour la sélection des points d'intérêt sont :

— point d'intérêt détecté s'il est retrouvé sur trois échelles ou plus,

— localisation du point d'intérêt au centre de la plus petite échelle détectée,

— suivi dans les échelles à l'aide de l'orientation de la zone d'intérêt,

— taille de la zone associée à l'échelle où la fonction réponse est maximum sur les échelles,

— poids du point d'intérêt associé au nombre d'échelles où il est détecté.

Nous qualifierons cette méthode de multi-échelles puisque c'est l'analyse des points dans les échelles qui mène à la sélection.

Nous proposons une autre méthode de sélection basée sur la suppression des superpositions des zones d'intérêt après repositionnement des points d'intérêt. Compte tenu de la dérive liée à la taille du filtre, les points d'intérêt appartenant au même coin sont repositionnés sur les coordonnées du point d'intérêt de plus petite taille, celui-ci étant le plus précis. Ensuite, le critère de sélection de cette méthode est le suivant : lorsque deux centres sont à une distance inférieure aux plus faibles des diamètres, alors le point de plus faible valeur de fonction réponse est supprimé.

TABLE 2.1 – **Synthèse de sélection des points** pour les deux approches proposées.

	Harris$_{ME}$
Localisation	du point de plus petite taille de la séquence multi-échelles,
Taille	du plus grand point de la séquence multi-échelles,
Poids	nombre d'échelles de la séquence multi-échelles,
Angle	associé à la plus petite taille de la séquence multi-échelles.

	Harris$_{S_2}$
Localisation	du point de plus petite taille de la séquence multi-échelles,
Taille	du point de poids maximum sur la séquence multi-échelles,
Poids	valeur de la fonction de réponse,
Angle	associé au point.

Les résultats selon ces deux méthodes de sélection sont présentés figure 2.13. Tous les points de l'étoile sont détectés ainsi que son cœur pour les plus grandes échelles. La méthode multi-échelles

extrait aussi certaines formes en T des branches de l'étoile. Pour la détection par suppression des superpositions sur l'image réelle, les petites échelles détectent les coins nets ainsi que les forme en T. Ces "coins" ne sont détectés qu'à une échelle et sont donc perdus dans la détection multi-échelles (fig. 2.14). La sélection par suppression des superpositions interdit la plupart des chevauchements entre points d'intérêt, ce qui n'est pas le cas de la seconde méthode. Compte-tenu des critères de sélection, la détection multi-échelles extrait des points d'intérêt de tailles plus grandes. Les zones de la peinture imitant la spécularité sont détectées à des échelles plus grandes.

Nous pouvons nous demander quelle approche est la meilleure. Si la détection par suppression des superpositions localise et conserve la majorité des coins nets, y compris dans les zones colorées, la pertinence de tel points d'intérêt dans une autre vue de l'image se pose. Un point stable dans les échelles semble plus robuste lors d'un changement de point de vue par exemple. Nous avons choisi de conserver la taille de filtre maximale pour la détection multi-échelles, mais une taille moyenne entre la minimum et la maximum ne serait-elle pas plus "juste".

2.3 Preuve de l'approche

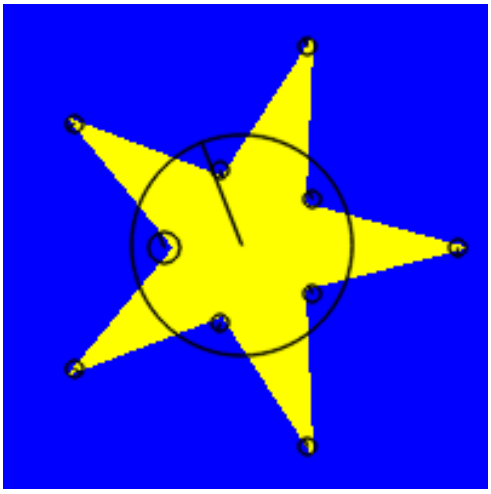
Nous comparons notre approche couleur avec quatre détecteurs niveaux de gris, deux détecteurs de *blob*, et deux détecteurs de coins. Les détecteurs de *blob* sont : les différences de Gaussiennes implémentées dans le descripteur SIFT proposé par Lowe (1999) (nommé DoG par la suite), et l'approximation via les filtres de Haar introduite par Bay et al. (2006) (nommé Haar dans cette section). Les détecteurs de coins sont : la proposition de Mikolajczyk and Schmid (2001) combinant le détecteur de Harris avec une approche multi-échelles de type LoG (noté H-L pour Harris-Laplace) et le détecteur AGAST présenté par Mair et al. (2010) basé sur le détecteur FAST. Nous étudierons le détecteur couleur présenté précédemment avec les deux méthodes de sélection des points d'intérêt : méthode multi-échelles (Harris$_{ME}$) et méthode par suppression des superpositions (Harris$_{S_2}$). Pour le calcul du gradient, nous utilisons les courbes de sensibilité à 10° de l'espace $CIERGB$ comme dans le chapitre précédant.

TABLE 2.2 – **Approches utilsiées**.

détecteur	**type**	**Caractéristiques**
DoG	*blob*	Utilisé dans SIFT Lowe (1999)
Haar	*blob*	Utilisé dans SURF Bay et al. (2006)
Harris-Laplace	coins	Harris + multi-échelles LoG Mikolajczyk and Schmid (2001)
AGAST	coins	détecteur de motifs Mair et al. (2010)
Harris$_{ME}$	coins	Conserve les points détectés sur au moins 3 échelles
Harris$_{S_2}$	coins	Conserve le point de fonction réponse maximum dans un diamètre donné



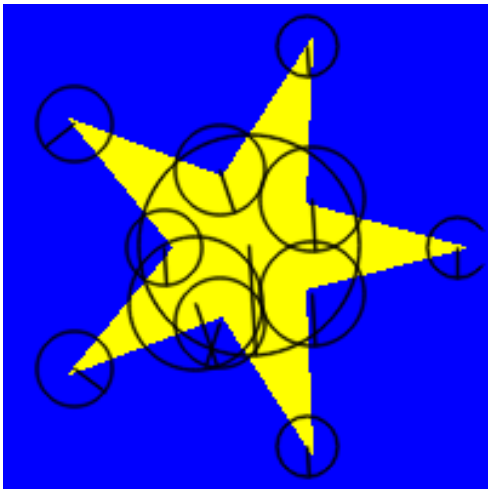
(a) Méthode de détection par suppression des superpositions sur une image réelle (340 points détectés)



(b) Méthode de détection par suppression des superpositions sur une image de synthèse (11 points détectés)

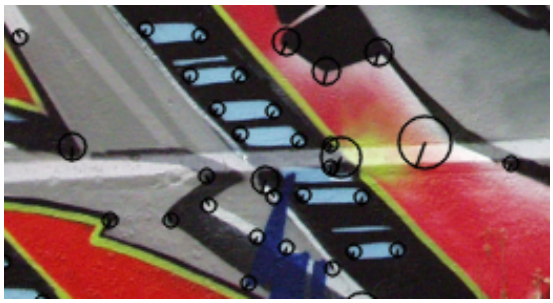


(c) Méthode multi-échelles de détection sur une image réelle (97 points détectés)



(d) Méthode multi-échelles de détection sur une image de synthèse (12 points détectés)

FIGURE 2.13 – **Résultat avec sélection de points d'intérêt**. La méthode par suppression des superposition conserve plus de points d'intérêt de petites tailles. La méthode multi-échelles autorise la superposition des points d'intérêt.



(a) Méthode de détection par suppression des superpositions sur une image réelle

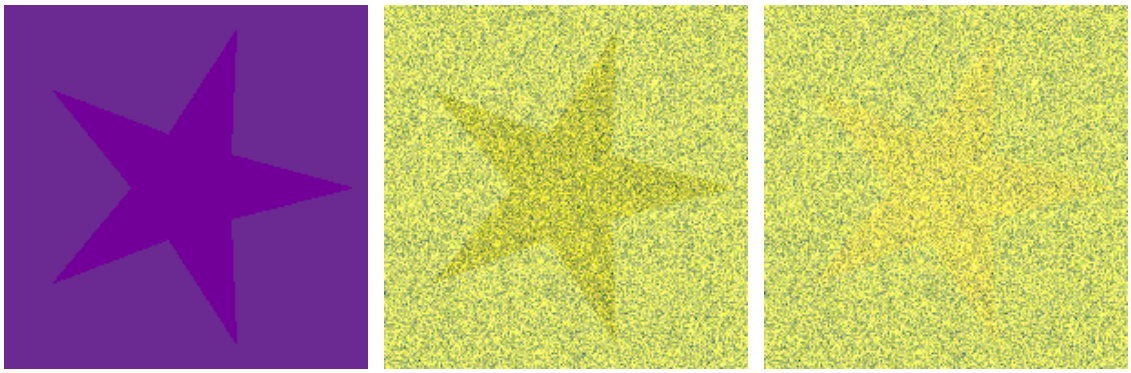


(b) Méthode multi-échelles de détection sur une image réelle

FIGURE 2.14 – **Zoom des résultats de sélection pour l'image réelle**. Beaucoup des points d'intérêt détectés sur les très petites échelles avec la méthode par suppression des superpositions sont perdus lors de la détection multi-échelles.

2.3.1 Base de données artificielles

Pour évaluer l'apport du détecteur de points d'intérêt couleur que nous proposons, nous reprenons pour partie la base de données artificielles présentée dans le chapitre 1. Afin d'avoir des coins à détecter, nous utilisons un masque d'étoile à cinq branches pour sélectionner l'apparition de la première ou de la seconde distribution couleur. La figure 2.15 présente des images exemples. La première est une image bicolore avec un taux de séparabilité $r = 1$. Les deux suivantes présentent un taux de séparabilité $r = -1$ où les distributions se chevauchent fortement. En fonction de l'orientation de la paire couleur, il est difficile voire impossible de distinguer l'étoile sans la connaître.



(a) Coins faciles à extraire ($r = 1$) (b) Coins difficiles à extraire ($r = -1$) (c) Coins impossibles à extraire ($r = -1$)

FIGURE 2.15 – **Exemples d'images artificielles pour la détection de points d'intérêt**.

Nous effectuerons trois tests sur cette base de données. Tout d'abord nous mesurerons la sensibilité des détecteurs en évaluant le ΔE minimum pour extraire des points d'intérêt. Ensuite, nous évaluerons les distances entre les points détectés et les points théoriques. Nous ne conserverons que la distance minimale à chaque point exact afin de mesurer le rappel des détecteurs. Enfin, nous réaliserons une expérience similaire à celle du gradient pour regarder l'impact de la complexité de l'image sur la précision de détection des détecteurs comparés.

2.3.1.1 Sensibilité des détecteurs

Avant toute chose, les détecteurs ayant généralement un seuil de détection, nous voulons déterminer quel ΔE choisir pour la création des paires couleur afin que les comparaisons soit pertinentes. Nous avons utilisé l'implémentation d'*OpenCV* pour ces détecteurs niveau de gris (NdG) avec les paramètres par défaut d'où l'importance de déterminer le seuil de détection.

Pour cela, nous nous plaçons dans un cas facile ($r = 1$), deux aplats, l'un formant l'étoile, l'autre le fond (fig. 2.15a). Pour 100 paires de couleurs par ΔE, nous mesurons le pourcentage d'images ayant permis d'identifier des points d'intérêt par l'approche étudiée. Les résultats sont donnés figure 2.16. Compte tenu du seuil de détection dépendant des maxima de la fonction réponse, dès que le ΔE est supérieur à 0, les deux approches couleur détectent systématiquement des points d'intérêt. Ce n'est pas le cas des autres détecteurs dont le seuil est un paramètre à choisir lors de la détection. Pour ces quatre détecteurs NdG, il faut un ΔE supérieur à 5 pour extraire des points sur certaines paires couleur uniquement. Même lorsque le ΔE devient très grand, certaines paires couleur ne permettent pas la détection de points d'intérêt en niveaux de gris. La figure 2.17 montre des cas de réussite et d'échec de détection pour des images dont les paires de couleurs ont un $\Delta E = 30$. La figure 2.17a présente un cas où tous les détecteurs extraient des points d'intérêt. Les détecteurs Haar et AGAST détectent des points d'intérêt sur l'image associée à la figure 2.17b mais pas DoG et Harris-Laplace.

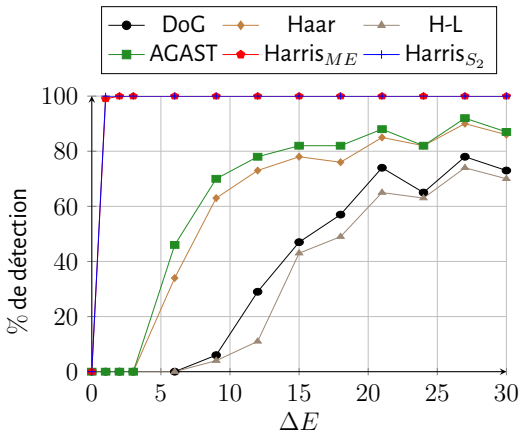


FIGURE 2.16 – **Détection des points d'intérêt en fonction de l'écart couleur** pour $r = 1$.



(a) Détection pour les détecteurs NdG (b) Détection pour certains détecteurs NdG (c) Pas de détection pour les détecteurs NdG

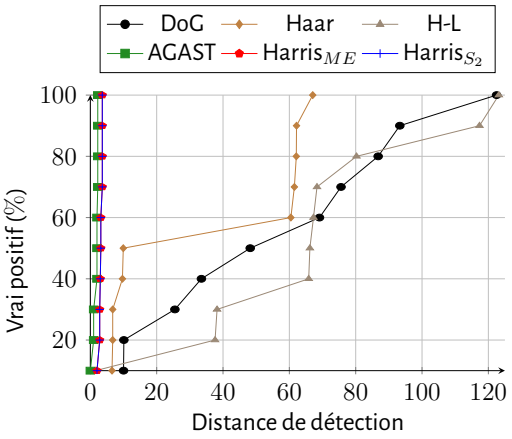
FIGURE 2.17 – **Exemples d'images pour** $\Delta E = 30$. Les images ne jouant que sur des différences de teintes ou de saturation impliquent un changement quasi-nul sur la couche luminosité interdisant aux détecteurs NdG d'extraire des points d'intérêt.

La différence de luminosité entre les deux couleurs de la figure 2.17b est plus faible que celle de la figure 2.17a ce qui explique la non extraction de certains détecteurs. La paire de couleur de la figure 2.17c ne diffère presque pas sur le plan luminosité, il est donc impossible pour des détecteurs NdG d'extraire des points d'intérêt.

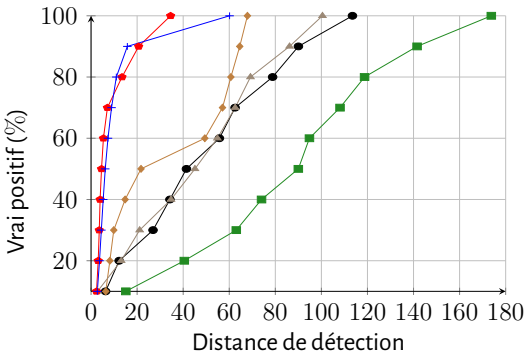
Le ΔE n'est plus considéré comme valide au delà de 20, nous devons donc choisir une valeur inférieure. De plus, un écart plus grand revient à comparer des images de couleur suffisamment différente pour limiter la complexité. Afin, de pouvoir tout de même nous comparer aux détecteurs NdG, nous choisissons pour les expériences suivantes $\Delta E = 15$.

2.3.1.2 Distance aux points réels

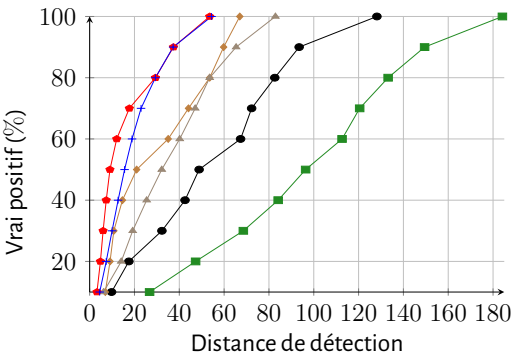
Pour évaluer la qualité des détecteurs, la première idée est de compter le nombre de points d'intérêt détectés à une distance donnée du coin réel de l'image synthétique. Il est difficile d'évaluer la taille de la zone d'intérêt associée à un coin, notamment sur les images d'étoiles de la base artificielle. La zone doit-elle être de taille différente sur une branche ou sur l'intersection de deux branches ? Nous choisissons donc de ne pas faire de choix arbitraire de distance qui serait lié à cette taille de zone d'intérêt. Nous avons choisi de mesurer la distance minimum depuis un coin issu du générateur, aux points d'intérêt déterminés par le détecteur étudié. Nous pouvons ainsi afficher une courbe



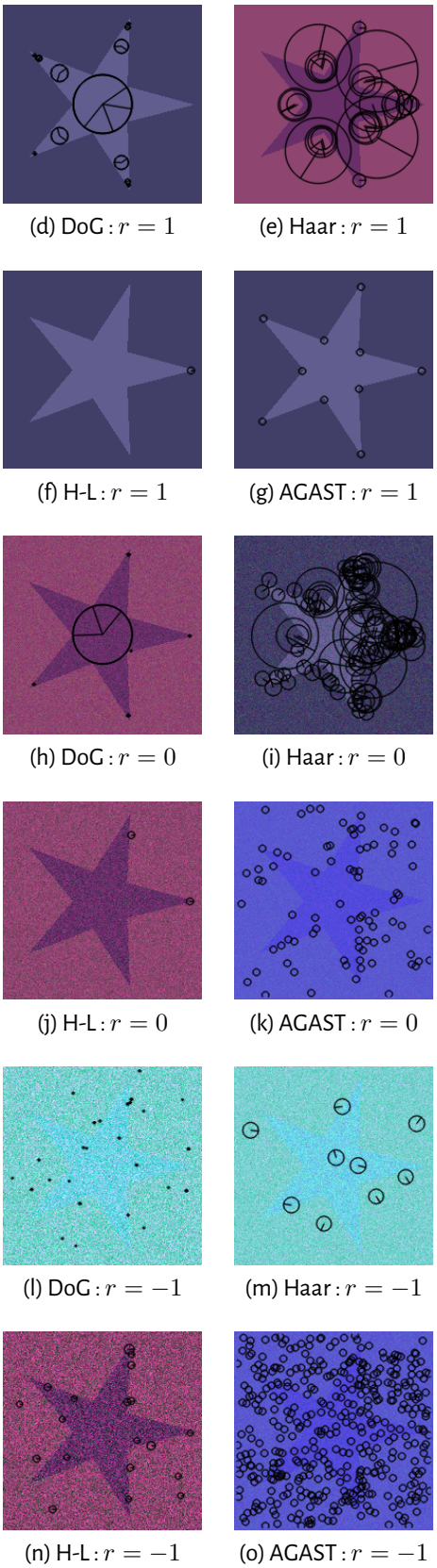
(a) Différence entre les détecteurs pour $r = 1$



(b) Différence entre les détecteurs pour $r = 0$



(c) Différence entre les détecteurs pour $r = -1$



(d) DoG : $r = 1$ (e) Haar : $r = 1$

(f) H-L : $r = 1$ (g) AGAST : $r = 1$

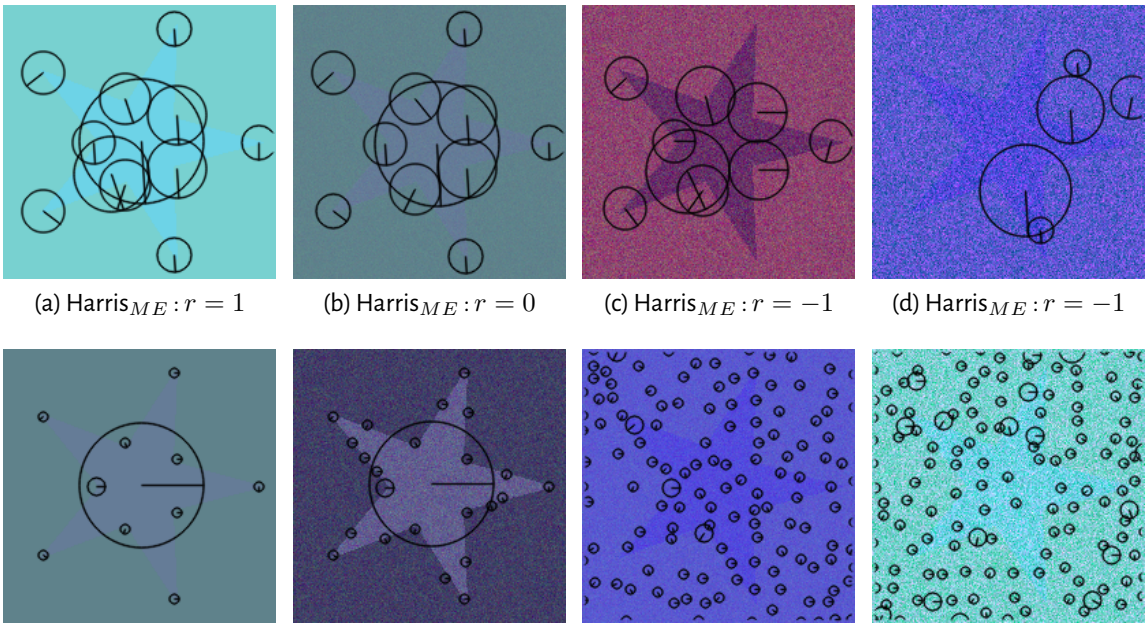
(h) DoG : $r = 0$ (i) Haar : $r = 0$

(j) H-L : $r = 0$ (k) AGAST : $r = 0$

(l) DoG : $r = -1$ (m) Haar : $r = -1$

(n) H-L : $r = -1$ (o) AGAST : $r = -1$

FIGURE 2.18 – **Rappel en fonction de la distance aux points exacts** et images exemples pour les détecteurs niveaux de gris.



(a) Harris$_{ME} : r = 1$ (b) Harris$_{ME} : r = 0$ (c) Harris$_{ME} : r = -1$ (d) Harris$_{ME} : r = -1$

(e) Harris$_{S_2} : r = 1$ (f) Harris$_{S_2} : r = 0$ (g) Harris$_{S_2} : r = -1$ (h) Harris$_{S_2} : r = -1$

FIGURE 2.19 – **Images de points d'intérêt détectés pour les détecteurs couleur**.

indiquant le pourcentage de points d'intérêt détectés correctement, en fonction de la distance maximale autorisée aux points exacts. Ainsi nous mesurons le rappel des points détectés par rapport aux points exacts pour une distance donnée.

Le rappel correspond ici au nombre de points à une distance inférieure à D divisé par le nombre de points à trouver (10). La précision correspond au même nombre divisé par le nombre points d'intérêt extraits par le détecteur. En résumé, le calcul du rappel et de la précision est donné par :

$$rappel = \frac{\#\mathcal{E}_D}{\#\mathcal{E}_{ex}}, \tag{2.15}$$

$$precision = \frac{\#\mathcal{E}_D}{\#\mathcal{E}_{det}}, \tag{2.16}$$

$$\begin{aligned} \text{avec} \quad & \mathcal{E}_{ex} = \{p | \, p \text{ soit un point exact}\}, \\ & \mathcal{E}_{det} = \{p | \, p \text{ soit un point détecté}\}, \\ & \mathcal{E}_D = \{p \in \mathcal{S}_{det} | \, d(p, \mathcal{E}_{ex})^{1} < D\}. \end{aligned}$$

Le calcul du rappel se fait pour 100 paires de couleurs et 4 paires de distributions. Nous avons ensuite calculé la probabilité du nombre de vrais positifs à une distance donnée. Pour les détecteurs NdG, les paires de couleurs ne permettant pas l'extraction de points d'intérêt ont été éliminées. La figure 2.18 présente les résultats pour trois valeurs de taux de séparabilité ($r = 1, r = 0$ et $r = -1$). La figure illustre aussi des résultats de détection pour les détecteurs NdG aux trois niveaux de complexité.

Lorsque les images sont des aplats, le détecteur AGAST donne la plus faible valeur de distance : la détection est parfaite (fig. 2.18g). La détecteur AGAST retrouve les motifs pour lesquels il a été conçu. Ce résultat n'est valable que pour environ 80% des images, les autres ne permettant pas la détection de points d'intérêt avec ce détecteur. L'aspect "bruité" réduisant sa probabilité de détection d'un motif donné. Dès que le taux de séparabilité augmente la distance nécessaire pour trouver les points croit

1. Par définition $d(p, \mathcal{E}) = \inf_{e \in \mathcal{E}} d(p, e)$.

TABLE 2.3 – **Statistiques du nombre de points extraits** en fonction des méthodes utilisées et du taux de séparabilité.

	Nb de POI	DoG	Haar	AGAST	H-L	Harris$_{ME}$	Harris$_{S_2}$
$r=1$	% Images>0	47%	76%	81%	38%	100%	100%
	Moyen	27	27	10	4	12	11
	Maximum	46	37	10	12	12	12
	Mininimum	4	7	10	1	12	12
$r=0$	% Images>0	55.25%	81.75%	99%	38.5%	100%	100%
	Moyen	25	70	1445	6	10	80
	Maximum	83	385	4688	65	12	140
	Minimum	1	1	1	1	5	17
$r=-1$	% Images>0	80%	92%	100%	55%	100%	100%
	Moyen	41	183	3320	15	9	123
	Maximum	111	784	4831	118	13	149
	Minimum	1	1	12	1	2	90

très rapidement. Sachant que les images artificielles sont de taille 200×200, une distance supérieure à 180 pour trouver le dernier coin interroge sur la position du point.

Le détecteur Haar détecte rapidement (distance inférieure à 30) cinq coins. Les coins suivants demandent une distance nettement plus grande. Dans le cas de la figure 2.18e, le détecteur détecte aisément l'intersection des branches mais rate des coins.

Les deux autres détecteurs NdG, répondent de façon similaire, la distance nécessaire pour augmenter le rappel croit progressivement. Les figures 2.18d et 2.18h mettent en avant la nature détecteur de *blob* de ce détecteur qui extrait des zones uniformes et non les coins de l'étoile. Lorsque la complexité est plus grande seules des échelles de l'ordre du bruit détectent des points d'intérêt. La figure 2.18f ne détectant qu'un seul point explique l'allure de la courbe du détecteur Harris-Laplace pour $r=1$. Lorsque le taux de séparabilité diminue, le nombre de points augmente mais la localisation reste erratique.

Dès que les distributions ne sont pas réduites à deux aplats [2], les détecteurs couleurs nécessitent une distance beaucoup plus faible que les détecteurs NdG pour extraire tous les points d'intérêt attendus. La figure 2.19 illustre des résultats de détection pour les détecteurs couleur. Puisque le détecteur AGAST ne détecte qu'environ 80% des points dans le cas d'aplats, les détecteurs couleurs sont donc les plus efficaces sur cette expérience. Une distance de 60 pixels pour retrouver tous les points lorsque la complexité est forte ($r=-1$) reste quand même très grande compte tenu des dimensions des images. Sélectionner une distance conditionne le nombre de points qui pourront être détectés dans l'expérience suivante. Une distance trop petite implique nécessairement l'échec de certains détecteurs, une trop grande n'a aucun sens. Les figures 2.19g, 2.19h, 2.19c et2.19d illustrent des détections pour le plus fort taux de séparabilité. Même avec une complexité forte, certaines paires de couleurs et de distributions permettent d'extraire les 10 coins attendus. D'autres cas en détectent moins. D'une manière générale, les détecteurs couleur sont plus spécifiques. Ils détectent de façon plus précise les points recherchés et ont moins tendance à faire de fausse détection. Par opposition les détecteurs NdG sont très sensibles aux changements couleur, ce que montre la table 2.3.

Notons que sur les cinq paires de couleurs illustrant la détection, seules deux permettent aux détecteurs NdG d'extraire des points (fig. 2.18d et 2.18e). Avec la complexité plus de paires permettent la détection.

Notons que nous ne recherchons ici que les vrais positifs sans nous intéresser au nombre de faux positifs engendrés, c'est-à-dire au nombre de points d'intérêt total détectés ne correspondant

2. Aplat : surface, plage de couleur uniforme.

à aucun coin de l'étoile. La table 2.3 récapitule le nombre d'images qui a servi au calcul du rappel, le nombre moyen de points trouvés dans les images ainsi que les nombres minimal et maximal de points détectés. Premièrement, toutes les images proposées ont pu être traitées par les deux détecteurs couleur. Par opposition, les détecteurs NdG ne sont pas toujours capables de sortir des points d'intérêt des images proposées. Dans le cas de Harris-Laplace, la capacité à détecter peut descendre à moins d'une image sur deux.

Deuxièmement, rappelons que nous attendions dix coins à détecter. Dans les cas simples ($r = 1$), et selon les échelles, des configurations particulières peuvent être trouvées par les détecteurs couleur (grande échelle spatiale : fig. 2.19a et 2.19e).

Lorsque les deux distributions sont complètement disjointes ($r = 1$), le détecteur AGAST joue parfaitement son rôle de détection de coins, mais pas sur toutes les images (81% seulement). Dès que la complexité croit, il sur-détecte les points d'intérêt (jusqu'à 3320 pour $r = -1$).

Le détecteur Haar est lui majoritairement en sur-détection, mais n'est capable d'extraire des points que dans 24% des cas triviaux ($r = 1$). Le fait que son nombre de point augmente avec la complexité (r tendant vers -1), indique qu'il détecte du bruit.

DoG et Harris-Laplace clôturent cette analyse, avec une capacité encore plus faible à détecter bien qu'ils soient sensibles au bruit.

Pour conclure sur cette table, revenons sur la comparaison entre les deux détecteurs couleur. La différence entre l'approche multi-échelles et suppression des superpositions réside dans la conservation des points d'intérêt cohérents sur plusieurs échelles (Harris$_{ME}$). Ce filtrage multi-échelles permet de ne conserver qu'un nombre de points réduits mais fiables. En revanche,il peut conduite à une sous-détection lorsque le taux de séparabilité diminue (complexité croissante).

La figure 2.20 présente les courbes de rappel selon la distance de détection, pour différents taux de séparabilité (r). En première constatation, nous remarquons que les approches couleur (Harris$_{ME}$ et Harris$_{S_2}$) présentent les meilleurs comportements face aux autres approches. Comme attendu, la performance de détection diminue avec la difficulté de discrimination (r tendant vers -1). Cependant, même dans les cas les plus difficiles, plus de 60% des points détectés le sont à une distance inférieure à 20 pixels. Par opposition, il faut atteindre une distance de 60 pixels pour les autres approches, ou 40 pixels pour les approches Harris-Laplace ou Haar.

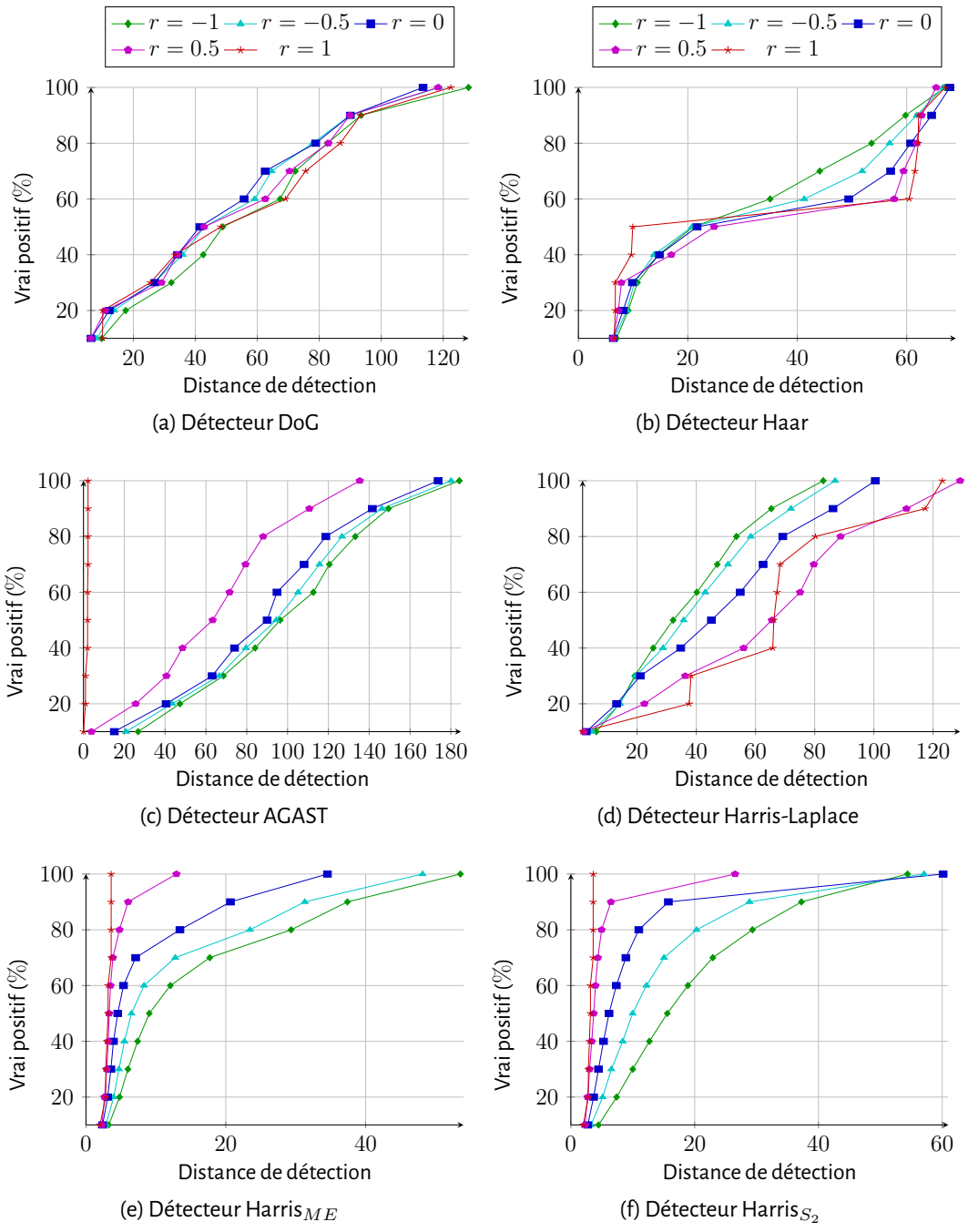
Quand le taux de séparabilité est nul (courbes bleues), les deux distributions sont disjointes de 99.7%. Sans être un cas trivial, cela est un cas proche de nombreuses réalités. Dans ce cas, les approches couleur trouvent 80% des points recherchés dans une zone proche (moins de 10 pixels). En revanche, la distance de recherche passe à 80 pixels pour l'approche DoG, mise à mal par son aspect *blob*. De même, l'approche AGAST, qui cherche des motifs particuliers, perd son potentiel discriminant avec la complexité de ces distributions très rapidement (dès que $r < 1$). Clairement, elle nécessite des pré-traitements pour être adaptée (filtre passe-bas). Le détecteurHarris-Laplace n'est pas meilleur que le détecteur DoG.

2.3.1.3 Influence de la complexité sur le rappel et la précision

Nous choisissons de prendre une distance de $D_{max} = 30$ pixels pour cette expérience. Nous pouvons ainsi mesurer le rappel et la précision. Cette distance correspond à 15% de la largeur de l'image de synthèse et permet de conserver 50% de rappel pour quatre détecteurs sur six.

Nous avons mesuré ces deux valeurs pour 100 paires de couleurs et seulement 4 paires de distributions. Nous effectuons ensuite la moyenne des valeurs uniquement sur les images ayant permis la détection de points. La figure 2.21 présente les résultats obtenus. Les courbes de rappel (fig. 2.21a) correspondent aux résultats précédents.

Ces courbes représentent le nombre de points détectés proches des points exacts. Le détecteur AGAST détecte le bon nombre de points pour des complexités faibles ($r = 1$) puis chute rapidement



(a) Détecteur DoG

(b) Détecteur Haar

(c) Détecteur AGAST

(d) Détecteur Harris-Laplace

(e) Détecteur Harris$_{ME}$

(f) Détecteur Harris$_{S_2}$

FIGURE 2.20 – **Courbes de rappel** selon la distance de détection et le taux de séparabilité pour les différents détecteurs.

dès $r = 0.5$. Les trois autres détecteurs niveaux de gris conservent un rappel a peu près constant quelle que soit la complexité. Les deux détecteurs couleur répondent comme attendu avec un rappel à 100% pour des images peu complexes, puis chutent légèrement avec la complexité pour obtenir un rappel d'environ 80% pour $r = -1$.

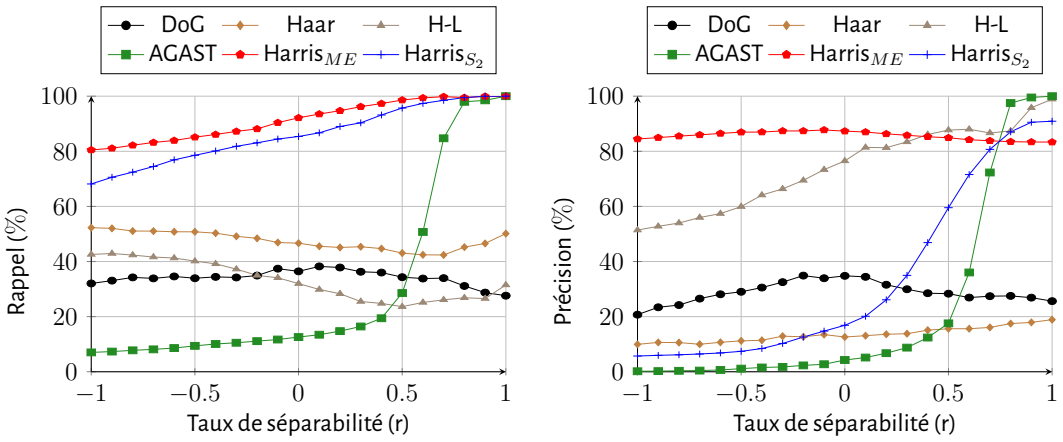
Les résultats de précision expriment le nombre de points détectés proches de points exacts par rapport au nombre de points détectés. Deux possibilités pour obtenir un score proche de zéro, peu de points proches des points exacts ou trop de points détectés. Le détecteur AGAST détecte beaucoup de points dès que l'image devient complexe. Cela explique la chute de la précision dès $r = 0.5$. Le détecteur Haar détecte aussi de plus en plus de points avec la complexité donc sa précision diminue même s'il détecte toujours cinq coins sur les dix. Le détecteur DoG reste à peu près constant sur toute la plage de complexité. En effet, le nombre de points détectés varie peu et la figure 2.20a montrait déjà la constance de ce détecteur. La précision du détecteur Harris-Laplace décroît avec la complexité. Les résultats de rappel augmentent avec la complexité, laissant entendre une meilleure détection, or la chute de la précision indique qu'il s'agit du nombre de points d'intérêt détectés qui augmente impliquant une plus forte probabilité d'être proche d'un point exact. Si le comportement des deux détecteurs couleur était similaire sur le rappel, ce n'est pas le cas pour la précision. En effet, le détecteur Harris$_{S_2}$ augmentant son nombre de détection avec le taux de séparabilité, sa précision chute avec la complexité des images. Au contraire, le détecteur Harris$_{ME}$ est presque constant sur toute la plage de complexité. Le choix de sélection sur plusieurs échelles implique une diminution du nombre d'échelles sur lequel sont trouvés les points avec l'augmentation du bruit. Ainsi, le nombre de points aura tendance à diminuer avec la complexité, ce qui explique qu'il soit le seul détecteur dont la précision augmente avec le taux de séparabilité.

Nous pouvons aussi noter que les détecteurs couleurs ont été calculés avec 15 tailles de filtres dont des filtres allant jusqu'à la taille du centre de l'étoile. Choisir un nombre de filtre et de taille plus faible aurait probablement amélioré la précision, notamment pour les complexités faibles.

2.3.2 Base de données Oxford

2.3.2.1 Protocole de l'expérimentation

Mikolajczyk and Schmid (2004) ont proposé un protocole avec une base de données pour évaluer d'une part un détecteur de points d'intérêt et d'autre part un descripteur de points d'intérêt. Ils s'inspirent des travaux de Schmid (1996) cité par Gouet et al. (2000) présentant un premier système



(a) Rappel (b) Précision

FIGURE 2.21 – **Influence de la complexité sur le rappel et la précision** pour $D_{max} = 30$.

d'évaluation de la qualité d'un détecteur de points d'intérêt. Mikolajczyk et al. (2005) proposent une base de données d'images réelles pour effectuer ces mesures.

Nous ne reprenons, dans l'expérience de Mikolajczyk and Schmid (2004), que l'analyse de répétabilité des points d'intérêt. Le détecteur ne cherche qu'à détecter le point d'intérêt. Aucun appariement ne peut être mesuré sans faire intervenir un descripteur de points d'intérêt. L'intervention d'un descripteur peut influencer les résultats en fonction de l'adéquation du descripteur avec le détecteur. C'est pour cette raison que nous ne nous intéressons qu'à la mesure de qualité d'un détecteur à ce stade.

Base de données

Cette base de données est constituée de huit séries de six images. Les images que nous avons utilisées sont présentées en annexe E. Chaque série s'intéresse à une transformation particulière : flou, luminosité, cisaillement, etc. La première image correspond à l'image de "qualité", c'est-à-dire l'image la plus nette ou avec la transformation permettant la meilleure lisibilité de la scène exploitée. Pour chacune des autres images, l'homographie permettant de replacer la première image dans le plan des suivantes a été calculée et est fournie avec la base.

Nous exploiterons les images mettant en œuvre : le flou sur des objets manufacturés et sur des arbres (séries *bikes* et *trees*), du cisaillement sur des objets manufacturés (séries *graffiti* et *wall*), un changement de luminosité dû à la capture de la photo (série *leuven*) et un changement de qualité de compression (série *ubc*). Les autres séries ne sont pas conservées car impliquant des images en niveaux de gris ou presque.

Par la suite d'autres séries d'images ont été insérées dans la base de données par différents auteurs (Strecha et al. (2008); Heinly et al. (2012)). Nous avons choisi de ne pas présenter leurs résultats car ils sont similaires à ceux présentés, les images étant principalement des objets manufacturés sur lesquels les détecteurs de l'état de l'art sont compétents. Le contenu spatio-chromatique est de complexité faible.

Critère de qualité du détecteur

Mikolajczyk and Schmid (2004) ont proposé de comparer les performances des détecteurs selon la répétabilité de la détection du point et la surface commune entre les points détectés. Nous choisissons d'appliquer ces tests sur nos détecteurs pour permettre une comparaison selon les termes de la littérature.

La répétabilité se mesure simplement en calculant le pourcentage de points communs aux deux images, détectés aux mêmes endroits dans les deux images. Cela nécessite de connaître la transformation d'une image à l'autre : $h_{i\to j}$, l'homographie transformant l'image i dans le même plan que l'image j. Pour chaque scène, il y a cinq vue/"transformations" incluant tout ou partie de la scène d'origine. Le nombre de points détectés peut donc varier entre la scène et les différentes vues. Par la suite, nous considérerons uniquement les points dont la position spatiale peut appartenir à la fois à la scène et à la vue.

Typiquement dans la figure 2.22, la figure 2.22a correspond à la scène et la figure 2.22d à une des cinq vues. Les points bleus détectés dans la scène sont projetés en vert dans la vue (fig. 2.22b). Ce code couleur sera conservé par la suite. De façon inverse, les points détectés dans la vue (fig. 2.22d) sont projetés dans l'image scène en magenta. Une partie a nécessairement été supprimée.

Le protocole de test vise à estimer le nombre de points d'intérêt commun à la vue et la scène. Soit p_i, un point détecté dans la scène et p_j son homologue potentiel dans la vue. Soit $h_{S\to V}$, la transformation homographique liant la scène et la vue. Un point p_R est considéré comme détecté dans les



(a) Scène avec les points détectés sur la scène (b) Vue avec les points détectés sur la scène



(c) Scène avec les points détectés sur la vue (d) Vue avec les points détectés sur la vue

FIGURE 2.22 – **Exemple de points d'intérêt communs entre la scène et une vue** pour le détecteur DoG.

deux images si :

$$\|p_i - h(p_j)\|_2 \leq \tau, \tag{2.17}$$

$$\text{avec} \quad p_i \in \mathcal{I}_{S\to V} = \{p_i \in I_S | h(p_i) \in I_V\}, \quad p_j \in \mathcal{I}_{V\to S} = \{p_j \in I_V | h^{-1}(p_j) \in I_S\}. \tag{2.18}$$

L'erreur de localisation de $\tau = 1.5$ pixels est sélectionnée ainsi car, d'après Mikolajczyk and Schmid (2004), elle n'empêche pas l'appariement correct des points d'intérêt suite à sa description. Cette distance correspond à une 8-connexité, c'est-à-dire que le point transformé ne peut s'écarter que d'un pixel tout autour du point visé.

La répétabilité est donc donnée par :

$$rep = \frac{\#\{p_R\}}{\min\left(\#\mathcal{I}_{S\to V}, \#\mathcal{I}_{V\to S}\right)}. \tag{2.19}$$

Nous obtenons ainsi un taux de points communs localisés aux mêmes endroits dans les deux images.

Le deuxième critère s'intéresse à la surface commune entre les deux points d'intérêt détectés comme communs aux deux images. En effet, compte tenu de la taille de la zone d'intérêt, nous pouvons évaluer les ellipses obtenues après transformation par homographie et mesurer le recouvrement de celles-ci. Mikolajczyk and Schmid (2004) considèrent le point d'intérêt comme correctement détectés dans les deux images si l'erreur de recouvrement est inférieure à un seuil. Les auteurs

TABLE 2.4 – **Statistique des points détectés sur la base Oxford**.

	$\#\{\text{POI}\}$	DoG	Haar	AGAST	H-L	Harris$_{ME}$	Harris$_{S_2}$
bikes 1000×700	Scène	3384	5773	11965	778	68	380
	$\mathcal{I}_{S\leftrightarrow V_1}$	1727	3742	3366	527	48	146
	$\mathcal{I}_{S\leftrightarrow V_1}$	373	1563	381	152	30	37
trees 1000×700	Scène	13270	9011	35028	5000	120	1431
	$\mathcal{I}_{S\leftrightarrow V_1}$	11255	8858	33536	4788	88	1376
	$\mathcal{I}_{S\leftrightarrow V_5}$	2912	5265	8800	1525	106	448
graffiti 800×640	Scène	2678	4796	7680	1439	98	350
	$\mathcal{I}_{S\leftrightarrow V_1}$	2108	3865	5992	1224	80	316
	$\mathcal{I}_{S\leftrightarrow V_5}$	1360	1956	4674	611	26	137
wall 1000×700	Scène	10358	10418	40690	5000	131	1640
	$\mathcal{I}_{S\leftrightarrow V_1}$	10067	7482	26288	4617	79	1173
	$\mathcal{I}_{S\leftrightarrow V_5}$	5137	3751	14098	1997	43	618
leuven 900×600	Scène	2461	5114	12514	891	53	275
	$\mathcal{I}_{S\leftrightarrow V_1}$	2108	4360	10092	579	50	268
	$\mathcal{I}_{S\leftrightarrow V_5}$	1148	2220	4968	336	47	202
ubc 800×640	Scène	5570	6108	22100	1295	52	433
	$\mathcal{I}_{S\leftrightarrow V_1}$	5570	6001	21260	1295	52	433
	$\mathcal{I}_{S\leftrightarrow V_5}$	3245	3770	5248	1295	43	425

font aussi remarquer la nécessité d'utiliser une échelle commune de transformation des points d'intérêt des deux images afin d'éviter un biais de mesure si la taille associées aux points de la première image est très petite ou très grande.

Les détecteurs que nous étudions sont invariants à l'échelle mais pas aux transformations affines. Le critère final d'erreur de recouvrement est mesuré ainsi :

$$Err_{rec} = 1 - E_h^2 \frac{\min(\sigma_i^2, \sigma_j^2)}{\max(\sigma_i^2, \sigma_j^2)}, \tag{2.20}$$

où la paire (p_i, p_j) répond au critère de répétabilité du point et E_h correspond à l'échelle induite par l'homographie. Seule la taille de la zone d'intérêt (σ_i pour le point p_i) est pertinente puisque nous n'avons pas d'information sur la forme de la zone d'intérêt.

C'est une mesure qui est très intéressante pour les détecteurs multi-échelles. Comme le proposent Mikolajczyk et al. (2005), nous ne conservons que les points ayant une erreur de recouvrement inférieure à 0.4. Cela permet d'obtenir une mesure similaire à la répétabilité bien que plus robuste.

2.3.2.2 Résultats

La table 2.4 reprend le nombre de points d'intérêt détectés dans la scène initiale (ligne grisée), ainsi que le nombre de points communs entre la scène et la vue comparée pour la première et la dernière vues.

Les figures 2.23 et 2.24 présentent les résultats pour les taux de répétabilité et de recouvrement. Premièrement, nous pouvons constater que le comportement des mesures de répétabilité et d'erreur de recouvrement sont très proches. Pour les images et transformations considérées, l'erreur de recouvrement n'apporte pas d'information supplémentaire.

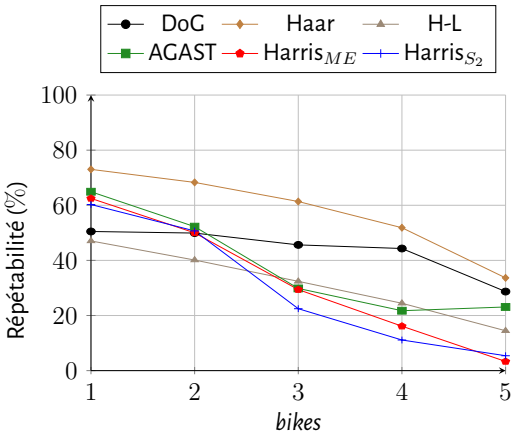
Le test de répétabilité a été crée en même temps que le détecteur Harris-Laplace. Au final, ce détecteur n'est le plus performant que dans le contexte du flou sur les objets manufacturés (images



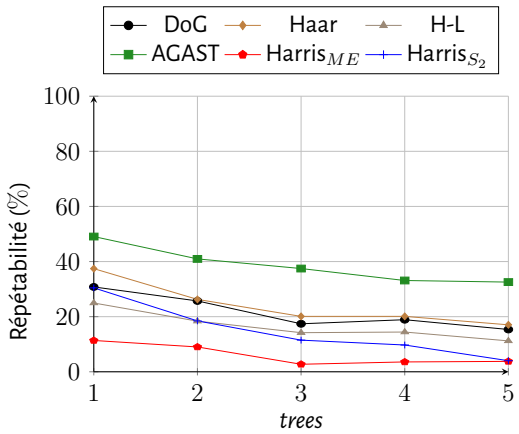
(a) Flou sur objets manufacturés



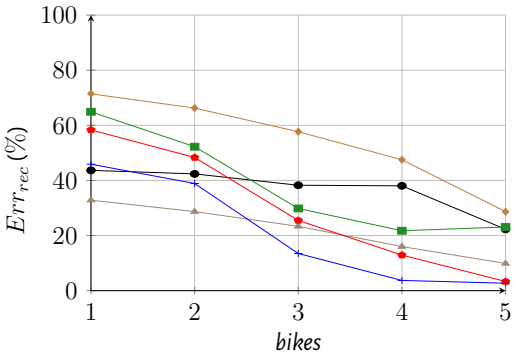
(b) Flou sur objets naturels



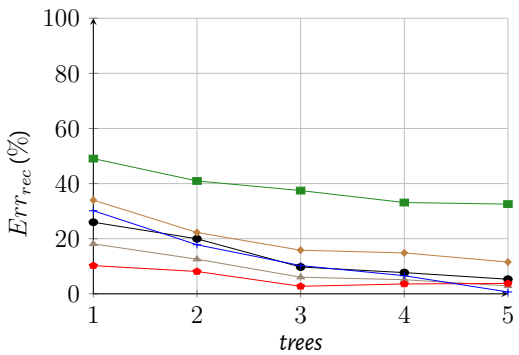
(c) Répétabilité



(d) Répétabilité



(e) Erreur de recouvrement

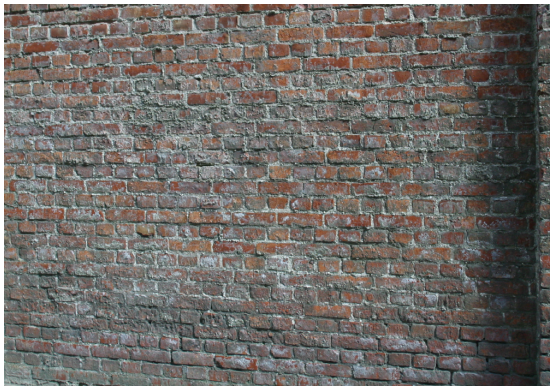


(f) Erreur de recouvrement

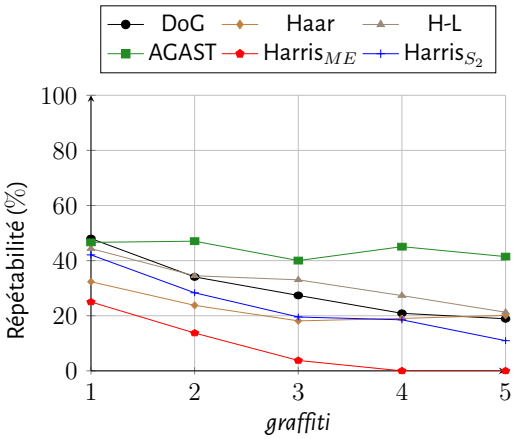
FIGURE 2.23 – **Résultats de répétabilité des détecteurs sur la base Oxford**.



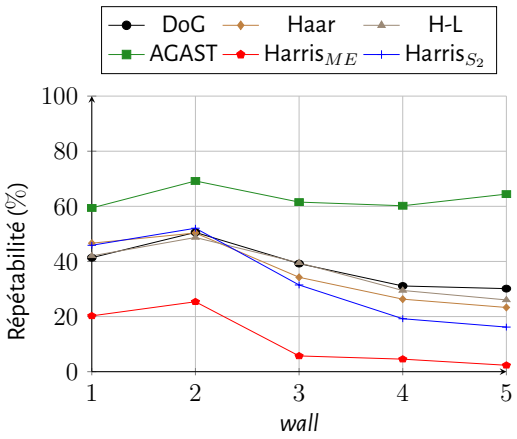
(a) Flou sur objets manufacturés



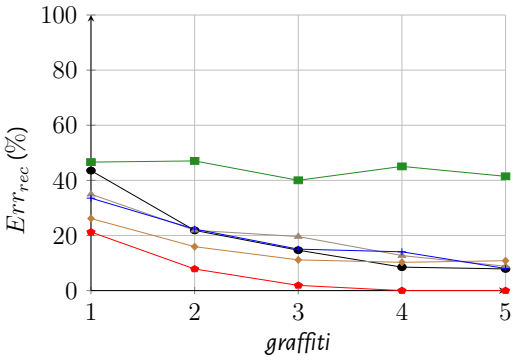
(b) Flou sur objets naturels



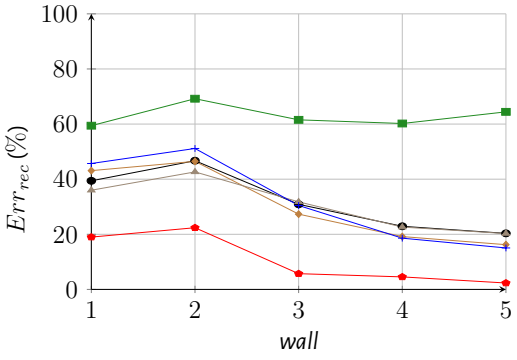
(c) Répétabilité



(d) Répétabilité



(e) Erreur de recouvrement



(f) Erreur de recouvrement

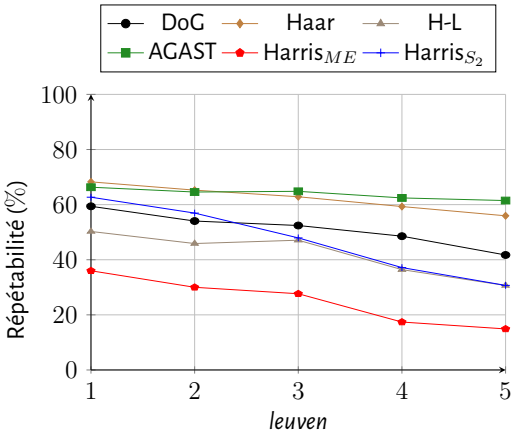
FIGURE 2.24 – **Résultats de répétabilité des détecteurs sur la base Oxford**.



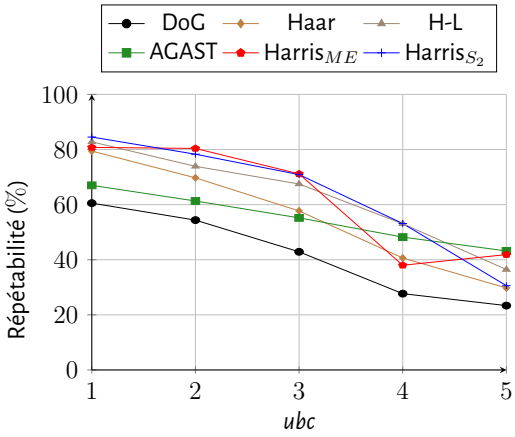
(a) Changement de luminosité



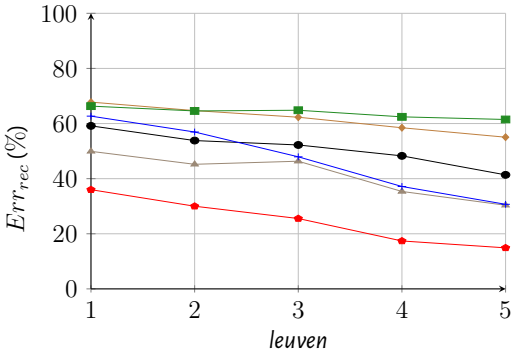
(b) Changement du taux de compression



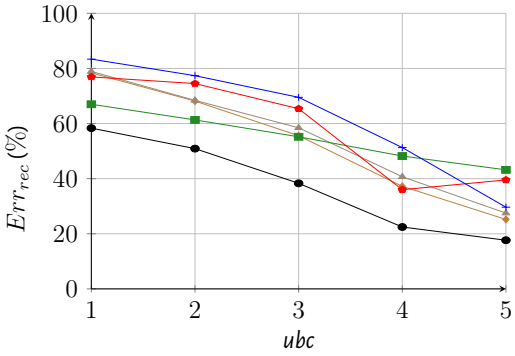
(c) Répétabilité



(d) Répétabilité

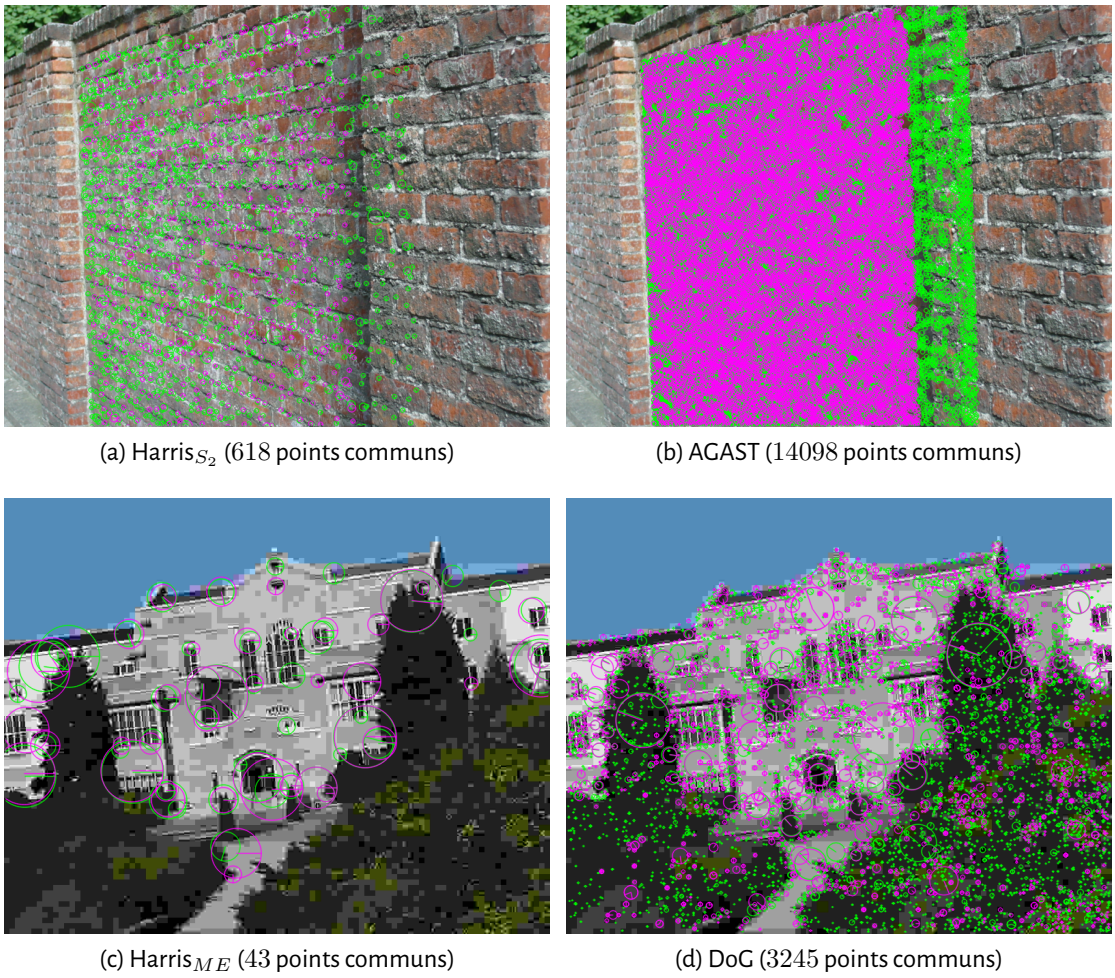


(e) Erreur de recouvrement



(f) Erreur de recouvrement

FIGURE 2.25 – **Résultats de répétabilité des détecteurs sur la base Oxford**.



(a) Harris$_{S_2}$ (618 points communs) (b) AGAST (14098 points communs)

(c) Harris$_{ME}$ (43 points communs) (d) DoG (3245 points communs)

FIGURE 2.26 – **Impact du nombre de points d'intérêt détectés** sur le critère de répétabilité. Sur chaque image, les points conservés de la première image (meilleure qualité) sont affichés en vert. En magenta apparaissent les points conservés de l'image affichée (ici la plus déformée).

bikes). Le raffinement multi-échelles réduit le nombre de points d'intérêt et résous l'incertitude liée au flou (2.4). Dans le contexte d'objets manufacturés, la faible complexité chromatique lui permet de se passer de l'information couleur. Dans le cas d'objets naturels (fig. 2.23d), son manque de considération pour la couleur lui nuit.

Le détecteur AGAST, basé sur la détection de motifs particuliers, semblerait être un candidat solide pour ce test (fig. 2.23d, 2.24c, 2.24d et 2.25c). Cependant, le succès du taux de répétabilité doit être analysé avec les données de la table 2.4. En effet, la bonne répétabilité est obtenue grâce à une très forte sur-détection (au moins 14098 points communs extraits pour la série *wall*, fig. 2.24d). Sur la déformation par cisaillement (série *wall*), la grande quantité de points d'intérêt détectés par le détecteur AGAST explique le bon résultat de répétabilité. Plus il y a de points, plus la probabilité de détecter un point à une distance de 1.5 pixel est forte. Cette probabilité augmente si les points communs sont cherchés sur une image où la partie commune est de plus petite taille.

Les approches DoG et Haar ne ressortent jamais comme les plus performantes, même si elles sont les plus utilisées.

Finalement, les approches couleur n'obtiennent de meilleurs scores que dans le cas des transformations de types compression, donc sans modification homographique. La compression préserve pour partie les gradients, donc les “coins”. Même si le contenu chromatique a été fortement altéré, il reste suffisamment de bords importants pour faciliter la détection. Le gain de nos détecteurs ici vient aussi de leur sélectivité entraînant peu de points d'intérêt détectés sur les zones de textures

(végétation dans cette série *ubc*). La figure 2.26c montre le peu de différence de points détectés entre la scène et la vue la plus compressée pour le détecteur Harris$_{ME}$. Au contraire, la figure 2.26d montre bien la perte de points dans les feuillages notamment avec l'augmentation du taux de compression impliquant nécessairement une diminution du nombre de points.

L'analyse comparative des deux approches couleur montre que le détecteur Harris$_{S_2}$ est toujours plus performant que le détecteur Harris$_{ME}$. La table 2.4 nous éclaire sur ce fait, l'approche Harris$_{ME}$ est trop sélective. La trop forte réduction du nombre de points possibles à apparier réduit fortement le taux de répétabilité. De l'autre coté, l'approche Harris$_{S_2}$ offre plus de souplesse,elle est moins sélective (618 points conservés face à 43 pour la dernière image de la série *wall* fig. 2.24d).

En excluant le détecteur AGAST (trop de points détectés pour une précision satisfaisante), les performances du détecteur Harris$_{S_2}$ sont comparables en termes de répétabilité aux approches DoG, Haar et Harris-Laplace. En revanche, le nombre de points extraits est bien plus faible : pas ou peu de sur-détection. L'approche Harris$_{S_2}$ est plus sélective. Ce propos ramené à des besoins d'indexation n'est pas sans intérêt. Néanmoins, il reste des améliorations à apporter pour limiter la perte de répétabilité lors des transformations homographiques importantes ; de même dans le contexte de flous d'images.

Pour conclure sur l'analyse de la répétabilité des détecteurs. Premièrement, nous observons bien l'aspect plus sélectif des deux détecteurs couleur. Cependant, une nuance est apparue dans leur usage. Pour des aspects d'appariements $3D$, une forte sélectivité associée à une quête de grande précision est attendue. Le détecteur Harris$_{ME}$ répond à ces attentes. En revanche, comme l'induit le test, pour des aspects “plus indexation”, la contrainte doit être plus lâche. Le détecteur Harris$_{S_2}$ est plus adapté. L'analyse des résultats à également mis en évidence des limites aux deux critères proposés. Ils favorisent la sur-détection sans assurer le gain de précision.

2.3.3 Comparaison point d'intérêt/point visuellement saillant

Notre propos n'est pas de montrer que les points d'intérêt détectés selon un modèle mathématique dans un espace RGB lié à un capteur (RGB ou autres) correspondent aux points saillants visuellement observés. Néanmoins, la question de la qualité de la relation entre ces deux ensembles de points est souvent sous-jacente. Il nous semble normal de l'adresser au moins une fois. Comme indiqué en entrée de ce chapitre, cette question n'a de sens que dans la phase pré-attentive de la lecture d'image par le cerveau.

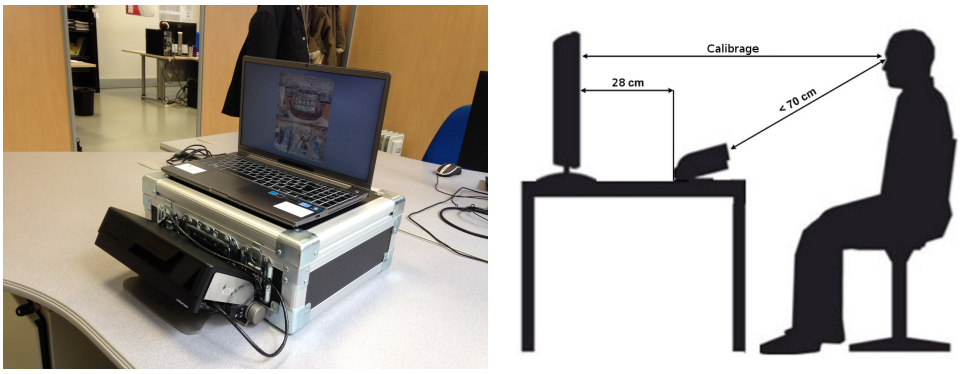
2.3.3.1 Protocole expérimental

L'outil eye-tracker

Afin de comparer le détecteur que nous proposons avec l'œil humain, nous avons mis en place une expérience psychovisuelle analysant le regard de l'utilisateur lors de la lecture d'une image. Un eye-tracker (outil de suivi du regard) et son logiciel associé ont analysé les points de fixation de chacun des utilisateurs.

L'outil eye-tracker est un matériel permettant de suivre le regard de l'utilisateur. Il peut être installé sur des lunettes pour suivre le regard dans tout l'espace ou simplement mesurer les mouvements de l'œil regardant un écran. Cet outil utilise l'infra-rouge qui sera réverbéré par les pupilles mais pas par le reste du visage.

L'eye-tracker utilisé est le *Tobii X-120*. Il permet de recueillir les deux mouvements majeurs du processus de vision. En premier lieu, il y a les saccades, puis les points de fixation. Les saccades sont des mouvements très rapides (< 50 ms de pause entre les mouvements) parcourant l'image sans



(a) Schéma de mise en œuvre (b) Mise en place expérimentale

FIGURE 2.27 – **Protocole de l'expérience** suivi du regard sur des images d'œuvre médiévale.

logique apparente. Ces saccades permettent de créer une pré-image qui indiquera les zones ou points d'intérêt qui peuvent être modélisés par les points de fixation. Ceux-ci sont des points où l'œil s'arrête suffisamment longtemps pour que le cerveau puisse analyser la zone qui l'entoure. Ce sont ces points là que nous souhaitons comparer aux points d'intérêt détectés par les détecteurs.

L'eye-tracker nécessite un calibrage pour être utilisé afin de reconnaître où regarde l'utilisateur. La figure 2.27 présente le schéma de mise en œuvre de l'expérience et une photo de la mise en place expérimentale.

L'expérience

Nous avons choisi de conserver une durée de test courte (12 à 15 min) afin de conserver l'attention des observateurs pendant toute la durée de l'expérience. Pour cela nous n'avons sélectionné que 20 images d'œuvres principalement médiévales.

Parmi celles-ci la moitié vient de la base Romane, crée et enrichie par le personnel du CESCM (Centre d'Études Supérieur de la Civilisation Médiévale). Ces images sont considérées comme connues par le panel d'utilisateurs qualifié d'experts.

Dix autres images ont été extraites d'autres bases d'œuvres médiévales. Celles-ci sont considérées comme inconnues pour les “experts”.

La plupart des images sont des peintures murales mais il y a aussi des mosaïques, des peintures sur bois… Nous avons essayé de varier les thèmes abordés même si la plupart restent religieux compte-tenu de la période. Certaines images sont bien conservées d'autres sont très dégradées et presque invisibles. La figure 2.28 présente quelques images utilisées.

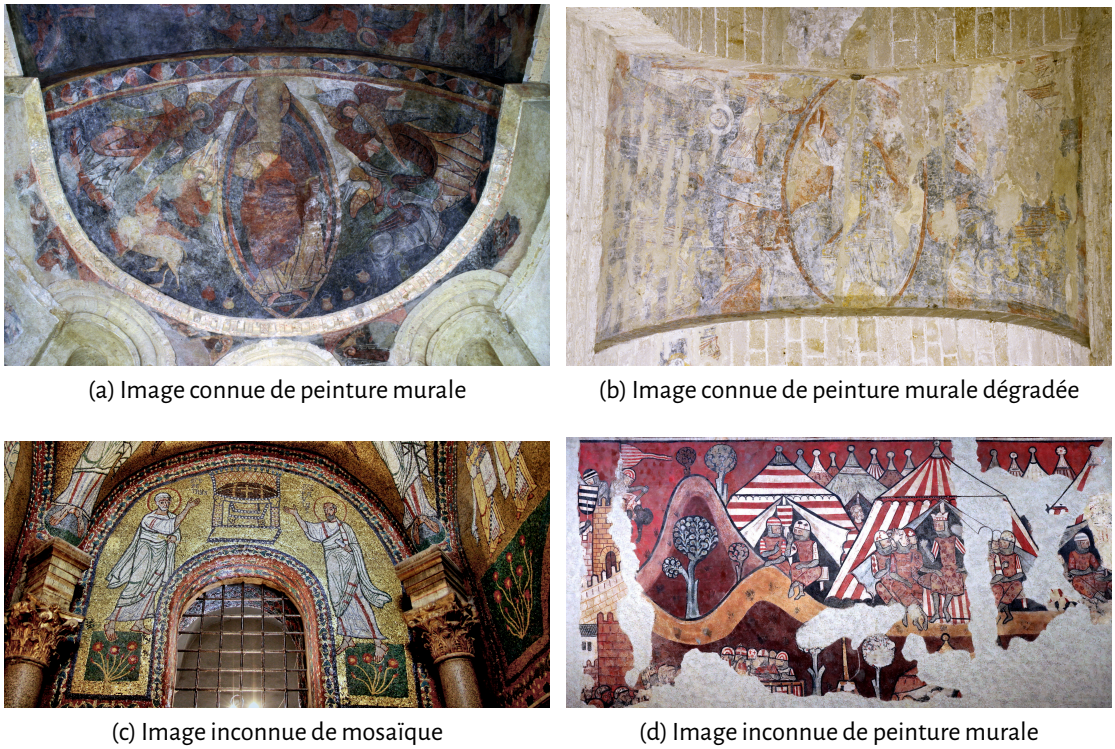
Chaque image est présentée 30 secondes entrecoupées de 5 secondes de gris neutre pour reposer les yeux. Pendant toute la durée du test, nous enregistrons la voix l'utilisateur. La tâche qui lui est confiée est de décrire l'image affichée afin, quand il le peut, d'identifier la scène observée.

Le panel d'utilisateur

L'objectif de l'expérience étant de vérifier la différence portée sur les images selon les connaissances de l'observateur, nous avons effectué l'expérience sur deux groupes.

Le premier comprend des chercheurs (étudiants, doctorants, enseignant-chercheurs…) dans le domaine de l'histoire médiévale. Tous les participants de ce groupe sont familiers avec la lecture d'une œuvre picturale médiévale. Un total de 21 participants entre dans cette catégorie.

Le second est constitué de membres du laboratoire XLIM (enseignant-chercheur, BIATS, doctorants…). Cette deuxième catégorie n'est pas familière avec les images médiévales mais certains participants sont familiers avec l'analyse d'image. 16 personnes ont réalisé l'expérience.



(a) Image connue de peinture murale

(b) Image connue de peinture murale dégradée

(c) Image inconnue de mosaïque

(d) Image inconnue de peinture murale

FIGURE 2.28 – **Exemples d'images utilisées pour l'expérience** qui étudie le suivi du regard sur des images d'œuvres médiévales.

Malheureusement certains utilisateurs n'ont pu être conservés faute de résultats fiables dus au port de lunettes ou à des mouvements répétés lors de l'acquisition du suivi du regard par exemple. De ce fait nous n'avons pu conserver que 32 observateurs.

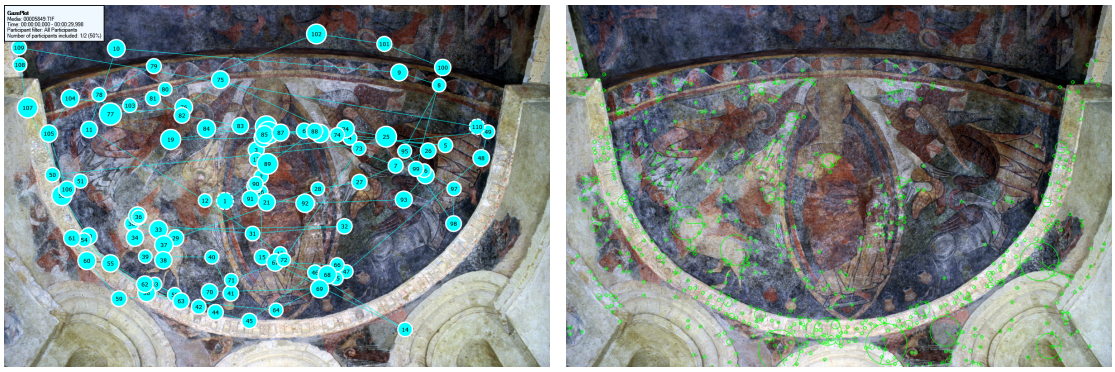
Défaut de l'expérimentation

Afin de pouvoir déplacer l'expérience, nous avons choisi d'afficher l'image sur un écran 17 pouces. L'utilisation de l'eye-tracker permettant d'extraire les saccades visuelles nécessite un recul plus grand que d'autres eye-trackers. Cette combinaison a conduit à la visualisation d'images relativement petites compte-tenu de la distance de perception induisant une plus grande difficulté de discrimination des détails.

2.3.3.2 Résultats

Nous nous intéressons ici à l'adéquation qu'il peut y avoir entre points d'intérêt et point saillants. Nous avons extrait les points d'intérêt avec le détecteur couleur sur quelques images de l'expérience. Afin de comparer au plus proche avec l'œil humain, nous avons transformé l'image dans l'espace LMS grâce à la matrice de passage donnée par le standard $CIECAM02$, et utilisé la matrice de Gram associée que nous avons présentée dans la section 1.3.3.2 du chapitre précédent.

La figure 2.29 présente les résultats de points de fixation pour un utilisateur et pour le détecteur couleur Harris$_{FV}$. Sur la première image (fig. 2.29a), les points semblent se disperser sur la même zone centrale (l'abside colorée). Notons que pour cette image, un autre utilisateur a moins regardé la frise du bas de l'image entraînant des différences plus importantes avec le détecteur. Dans la deuxième image (fig. 2.29c), les points ne se répartissent pas au même endroit, l'utilisateur écartant toutes les zones dégradés alors que le détecteur s'intéresse aux zones de changement peinture/plâtre



(a) Exemple de points de fixation d'un observateur

(b) 657 points extraits par le détecteur Harris$_{S_2}$



(c) Exemple de points de fixation d'un observateur

(d) 163 points extraits par le détecteur Harris$_{ME}$

FIGURE 2.29 – **Exemples de résultats de points extraits** par observation par un utilisateur ou par détection.

présentant des gradients forts. Ce premier constat dans les deux figures présentées (fig. 2.28a et 2.28d) montre la limite de l'analyse de la relation point d'intérêt/point saillant visuellement. Des gradients forts n'induisent pas forcément une saillance visuelle.

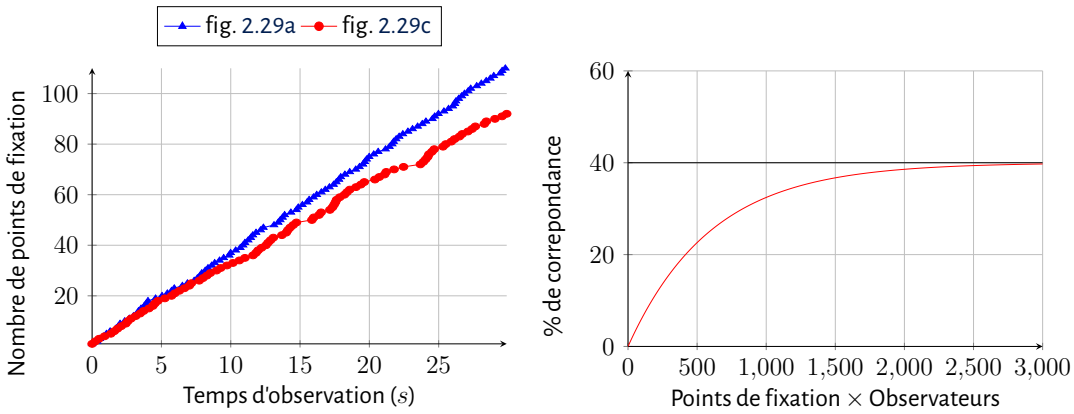
Attention, dans cette analyse nous ne prenons pas en compte le biais centré lié à l'analyse de la saillance visuelle. C'est-à-dire, le détecteur de points d'intérêt cherchera des points sur toute l'image alors que l'humain aura tendance à regarder prioritairement le centre. Nous avons estimé, que compte tenu du temps d'observation et de la taille de l'image, les observateurs auraient le temps de parcourir aussi les bords de l'image.

La table 2.5 nous donne un premier aperçu de la relation liant points d'intérêt et points de fixation. En moyenne plus de 35% des points d'intérêt détectés sont scrutés par les observateurs. Cela nous donne une première limite sur la relation, même si quatre images sont insuffisantes pour une conclusion définitive.

La figure 2.30a présente le nombre de points de fixation observés en fonction du temps. Les courbes correspondent au parcours visuel présenté figures 2.29a et 2.29c. Face à la tâche de des-

TABLE 2.5 – **Nombre de points d'intérêt détectés par nos propositions**. Le pourcentage de correspondance est le nombre de points d'intérêt ayant été regardé par un observateur divisé par le nombre de points d'intérêt total.

					
Harris$_{ME}$	Nb POI	62	164	152	163
	% Correspondance	46.7%	39.6%	32.9%	37.4%
Harris$_{S_2}$	Nb POI	657	1961	2276	1423
	% Correspondance	40.5%	42.0%	40.0%	45.9%



(a) Points de fixation en fonction de la durée

(b) Relation attendue

FIGURE 2.30 – **Résultats préliminaires**. La figure 2.30a présente la relation entre nombre de points de fixation et temps. La seconde (fig. 2.30b) présente la courbe théorique attendue pour l'adéquation entre points d'intérêt et points de fixation.

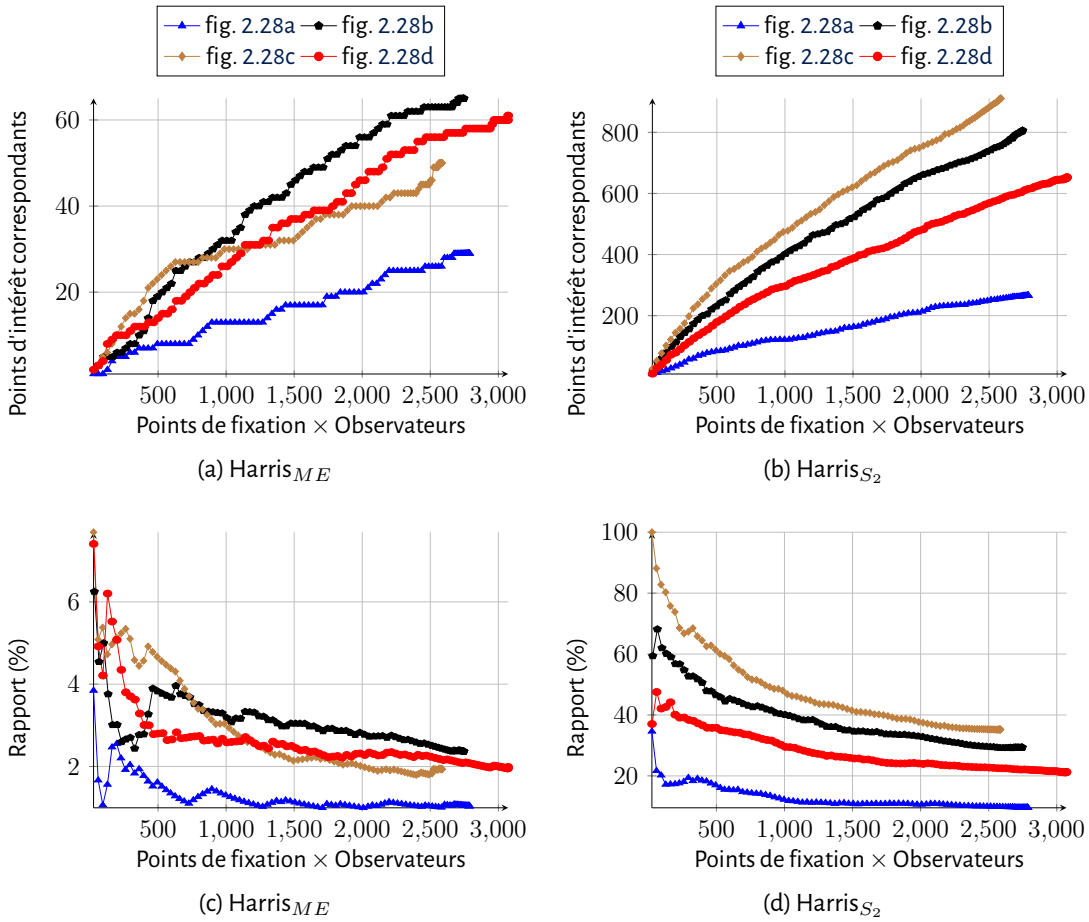
cription proposée, les observateurs ont analysé la scène sur une fréquence quasi-constante de 3 à 4 points de fixation par seconde.

Pour comparer ce parcours visuel aux points d'intérêt extraits, nous avons émis l'hypothèse que l'ordre d'observation des points pouvait varier d'un observateur à l'autre, mais qu'un même ensemble était observé au delà d'un certains temps. Ensuite, nous devons prendre en compte les différents niveaux de vision humaine, allant d'une vision pré-attentive à une vision pilotée par l'analyse sémantique du contenu. Il est donc évident que seule une partie des points d'intérêt sera observée. Nous devons donc attendre une courbe théorique du type de la figure 2.30b.

Les courbes des figures 2.31a et 2.31b indiquent que l'hypothèse initiale (liée à la courbe théorique de la figure 2.30b) semble cohérente. Cependant, les courbes obtenues ne permettent pas de définir la durée d'observation pré-attentive. Bien évidemment, le détecteur Harris$_{S_2}$, par une moindre sélectivité, offre une plus grande correspondance entre points de fixation et points d'intérêt. Les courbes des figures 2.31c et 2.31d montrent que globalement les courbes tendent à des vitesses différentes vers zéro. Donc le nombre de points d'intérêt observés tend vers un nombre fini inférieur au nombre total de points d'intérêt.

La question suivante que nous pouvons adresser est liée à la force d'attraction d'un point d'intérêt sur le regard dans les premières phases d'observation de l'image. Si cette relation n'est pas claire avec le détecteurHarris$_{ME}$ (fig. 2.32a), elle est plus nette avec le détecteur Harris$_{S_2}$ (fig. 2.32b). Notons que l'ordonnée des courbes correspond à la somme des fonctions réponses des points d'intérêt observés, ces courbes étant exprimées en fonction de leur ordre d'observation. Comme précédemment énoncé, deux observateurs peuvent ne pas visualiser les points saillants dans le même ordre. Néanmoins, l'évolution de la somme des fonctions réponses est décroissante : plus le temps passe, plus la probabilité qu'un point d'intérêt à forte valeur de fonction réponse soit observé diminue.

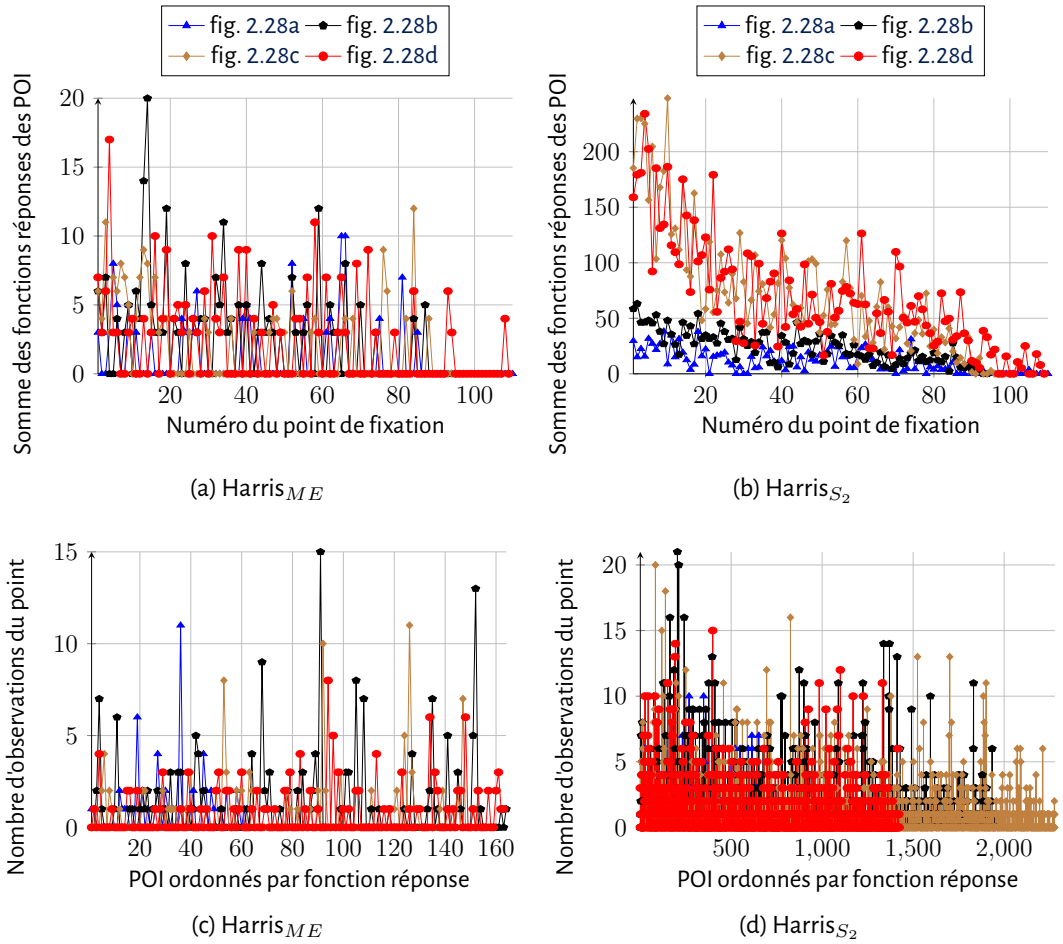
Afin de disposer d'une certaine assise statistique, nous avons mélangé les points de fixation des différents observateurs en fonction de leur ordre d'observation. Les figures 2.32c et2.32d permettent de répondre à la question de la probabilité que le regard observe plusieurs fois le même point d'intérêt en fonction de l'importance du point d'intérêt (*i.e.* de sa fonction réponse). Au final, aucune tendance ne vient confirmer cette hypothèse. Ce qui nous ramène au propos introductif de ce chapitre, mettant en garde sur les relations possibles entre points d'intérêt et points saillants. Passé une phase pré-attentive où le contenu induit l'observation, la suite de l'analyse visuelle est pilotée par le système visuel humain pour extraire un contenu sémantique qui était à oraliser dans le cadre de l'expérimentation.



(a) Harris$_{ME}$ (b) Harris$_{S_2}$

(c) Harris$_{ME}$ (d) Harris$_{S_2}$

FIGURE 2.31 – **Adéquation entre points d'intérêt et points saillants** en fonction du détecteur de points d'intérêt. Les figures 2.31a et 2.31b présentent le nombre de points d'intérêt observés en fonction du temps (nombre de points de fixation × observateur). Les figures 2.31c et 2.31d présentent le rapport entre le nombre de points d'intérêt observés et le nombre de points de fixation.



(a) Harris$_{ME}$ (b) Harris$_{S_2}$

(c) Harris$_{ME}$ (d) Harris$_{S_2}$

FIGURE 2.32 – **Adéquation entre points d'intérêt et points saillants** selon l'intensité des points. Les figures 2.32a et 2.32b présentent la somme des fonctions réponses en fonction des points de fixation ordonnés chronologiquement. Les figures 2.32c et 2.32d s'intéressent aux nombres d'observations du points d'intérêt ordonnés par fonction réponse décroissante.

2.4 Synthèse

L'objectif de ce chapitre était l'étude des détecteurs de points d'intérêt couleur. En étudiant l'état de l'art, nous avons perçu un manque de prise en compte de l'aspect vectoriel de la couleur dans les quelques approches couleur de l'outil détecteur de points d'intérêt. Au minimum, certains auteurs proposent des détecteurs couleur analysant de façon marginale la couleur.

Le manque perçu lors de l'analyse de l'état de l'art, nous a amené à proposer une méthode de détection de coins prenant en compte la couleur de façon vectorielle et générique. Pour cela nous avons repris la proposition de Harris and Stephens (1988) pour l'adapter à la couleur mais aussi au multi-échelles. Pour construire ce détecteur nous avons analysé et expliqué chaque étape de la proposition Harris and Stephens (1988). Cela nous a permis de mettre en avant l'importance et le rôle de l'utilisation d'un filtre passe-bas après la mesure de la matrice d'auto-corrélation. À partir de cette analyse à posteriori, nous avons proposé deux détecteurs couleur basés sur l'exploitation du *full-vector gradient* sur la surface associée à la taille de la zone d'intérêt. Le premier détecteur vise à extraire des points cohérents dans les échelles d'analyse. Le second relâche cette contrainte pour être moins sélectif, tout en restant précis dans la détection.

Nous avons ensuite éprouvé ce détecteur couleur face à une base de données d'images synthétiques permettant de simuler la complexité chromatique et de maîtriser celle-ci. Nos détecteurs permettent de retrouver les coins de l'étoile avec plus de justesse que les détecteurs niveaux de gris même quand les distributions couleur se chevauchent beaucoup.

Nous avons ensuite étudié les détecteurs au travers de l'expérience proposée par Mikolajczyk and Schmid (2004). Les résultats de nos propositions ne nous mettent pas en avant systématiquement. Nous avons associé l'analyse des résultats obtenus à celle du nombre de points extraits. Nous avons ainsi montré que les performances des détecteurs de la littérature étaient liées à une très forte sur-détection. Ce qui a nuit à leurs résultats dans la première expérience (image de synthèse) est le manque de point détectés. Ils semblent manquer de robustesse dans la détection.

Enfin, nous proposons une analyse de l'adéquation entre points d'intérêt et les points de saillance visuelle extraits d'une expérience psychovisuelle. Après avoir présenté l'expérience psychovisuelle menée, nous avons observé la concordance entre points d'intérêt et points de fixation. Ces quelques résultats, ne permettent pas de conclure sur une adéquation constante, mais ouvrent sur d'autres questions : quelles tâches pour l'utilisateur, quelle durée d'observation, etc.

Résumé

- Mise en évidence du manque de proposition couleur vectorielle pour les outils de type détecteurs de points d'intérêt.
- Adaptation de la proposition de Harris aux multi-échelles et à la couleur via deux approches.
- Proposition indépendante du nombre de canaux et des caractéristiques du capteur.
- Mise en place d'un protocole de validation de détection des points d'intérêt sur une base de données artificielles couleurs.
- Mise en évidence des limites des critères de comparaisons proposé par Mikolajczyk.
- Ouverture sur la concordance possible entre points d'intérêt et points saillants.

Critère de dissimilarité

Contenu

3.1 Distances et mesures de similarité 97
3.1.1 Distances et espaces vectoriels 97
3.1.2 Distances, divergences entre probabilités 97
3.1.2.1 Distances 98
3.1.2.2 Divergences 100
3.1.2.3 Le cas particulier de la divergence de Kullback-Leibler 100
3.1.3 Autres distances 101
3.1.3.1 Distance entre séquences d'objets à nombre d'états finis . . . 101
3.1.3.2 Distance topologique 101
3.1.4 Mesures intégrant une modélisation statistique 101
3.1.4.1 La divergence de KL dans le cas d'une loi normale 102
3.1.4.2 Limites de la modélisation 103
3.2 Nature des données et mesures utilisées 103
3.2.1 Séquences des bits 103
3.2.2 Les histogrammes comme densités de probabilité 105
3.2.3 Modélisation ou paramétrisation d'une pdf 106
3.2.4 Extension aux contextes couleur et spectral 106
3.3 Éléments de décision d'une mesure de similarité 108
3.3.1 Comportement avec une distribution de synthèse 108
3.3.1.1 Réponse à un écart chromatique 109
3.3.1.2 Réponse à un changement de forme 110
3.3.1.3 Réponse à un changement d'orientation 111
3.3.1.4 Comparaison d'échelle 112
3.3.2 Distributions d'images couleurs 113
3.3.2.1 Cas de textures proches 114
3.3.2.2 Cas de textures différentes 114
3.3.2.3 Test de normalité 120
3.4 Synthèse 122

Dans le chapitre précédent, nous avons proposé la prise en compte correcte et complète de la mesure couleur et spectrale pour écrire des opérateurs métrologiquement valides et donc plus performants. Avant d'aller plus loin dans la reconstruction d'une chaîne d'analyse, reconnaissance ou indexation d'images couleurs, nous abordons ici la question de la mesure de similarité entre attributs ou descripteurs. Nous établirons que le choix d'une mesure de similarité ou distance est par essence lié à la nature du descripteur et/ou de l'attribut. Deuxièmement, nous montrerons qu'un choix inadapté réduit les performances de reconnaissance/classification.

Bien qu'étudiée depuis près de 20 ans, la mesure de similarité entre deux régions d'images couleur est encore un problème ouvert. Cette question est pourtant au cœur de nombreux processus, que ce soit pour la comparaison de points d'intérêt ou celle de l'aspect de deux surfaces colorées (contrôle qualité). Au delà d'une réponse supplémentaire à ces problématiques, notre objectif est de tendre vers une mesure qui valide les propriétés métrologiques d'usage d'une mesure de différence (exactitude, fidélité...).

Replaçons notre contexte : les descripteurs de type histogrammes de gradients que nous verrons dans le chapitre suivant sont habituellement considérés comme des vecteurs de $\mathbb{R}^n$. Or l'ordre des coordonnées a une importance puisque ces descripteurs sont définis comme des fonctions de densité de probabilités. Cette séquence ordonnée ne peut donc être traitée comme une liste de coordonnées indépendantes. Nous choisissons de les considérer comme une densité de probabilités à condition de les normaliser. Si l'on travaille avec les occurrences de différences couleur, nous nous trouvons aussi face à une distribution de probabilité, mais en dimension n.

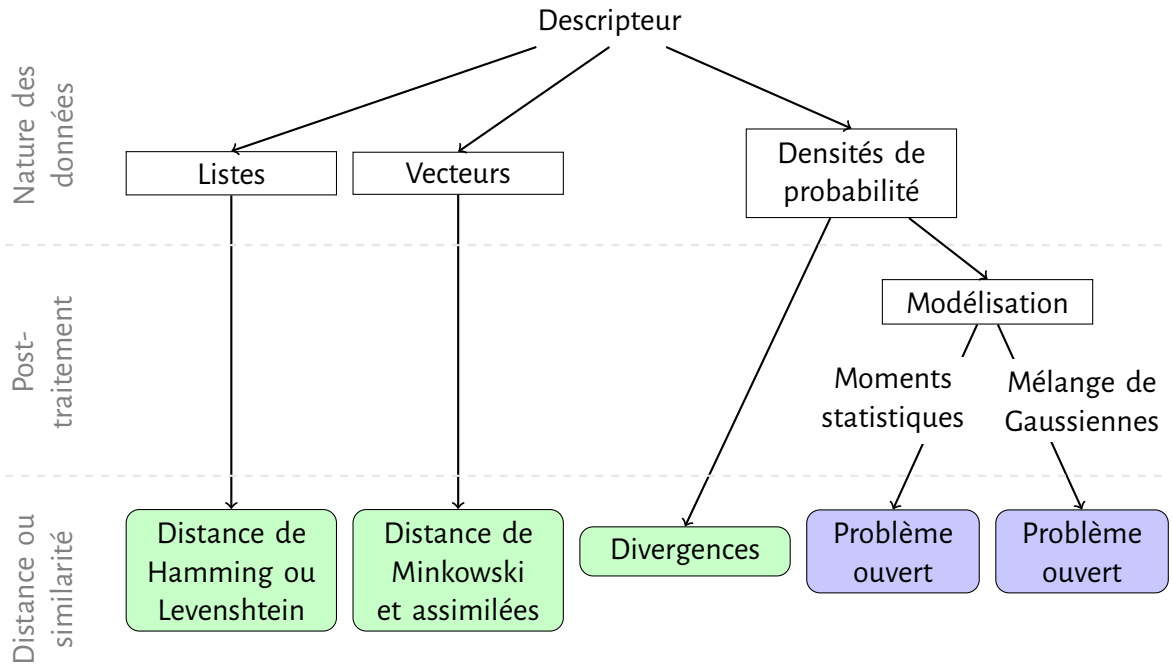


FIGURE 3.1 – **Organisation des distances à utiliser en fonction du type des données du descripteur**.

La figure 3.1 illustre les distances à mathématiquement associer aux types de données du descripteur. Le descripteur peut avoir plusieurs natures de données : le type liste, vecteur ou densité de probabilité. Nous verrons dans la section 3.2 que la nature du descripteur implique immédiatement une famille de distances associées. Ces distances seront présentées dans la section 3.1. Il peut y avoir une modélisation des données du descripteur, nous montrerons dans la section 3.2.3 que ces questions ne sont pas encore résolues.

Dans ce travail, nous montrerons la nécessité de considérer l'ensemble des couleurs issues d'une région ou image couleur comme une distribution dans un espace n dimensionnel. Nous définirons dans la section 3.1 la distance et la mesure de similarité, puis nous rappellerons les définitions

des principales distances et semi-métriques et dans quels cas elles doivent être utilisées. Dans une deuxième partie (sec. 3.2), nous étudierons les distances utilisées dans l'état de l'art par rapport aux types de données des descripteurs. Nous proposerons alors une approche basée sur la divergence de Kullback-Leibler (sec. 3.3), adaptée à l'estimation de la différence entre distributions dans des espaces n-dimensionnels (sec. 3.3.1). Ensuite, nous montrerons le comportement de cette divergence sur des distributions et des images issues de la base *VisTex* MIT (1995) (sec. 3.3.2).

3.1 Distances et mesures de similarité

La notion de distance correspond à un cadre mathématique strict. C'est d'abord une mesure et si une partie des propriétés n'est pas respectée, le terme de mesure de similarité remplace celui de distance. Il existe un grand nombre de propositions pour ces deux catégories. Toutes ces distances et divergences ont leurs atouts et points faibles. Nous allons dans cette section les présenter ainsi qu'expliquer le cadre dans lequel elles s'appliquent.

Commençons par définir mathématiquement la distance. Soit E un espace vectoriel et une application $d : E \times E \longmapsto \mathbb{R}^+$. d est une distance si elle vérifie les trois propriétés de symétrie, séparabilité et l'inégalité triangulaire :

$$\forall (a,b) \in E^2, d(a,b) = d(b,a), \tag{3.1a}$$

$$\forall (a,b) \in E^2, d(a,b) = 0 \Leftrightarrow a = b, \tag{3.1b}$$

$$\forall (a,b,c) \in E^3, d(a,c) \leq d(a,b) + d(b,c). \tag{3.1c}$$

Les mesures de similarité, respectent uniquement la symétrie, la séparabilité mais pas l'inégalité triangulaire.

3.1.1 Distances et espaces vectoriels

Si un espace vectoriel possède un produit scalaire, alors il en découle une norme associée. Nous pouvons ainsi définir une distance à partir de cette norme : c'est le résultat de la norme de la différence des deux vecteurs à comparer. La plupart des espaces vectoriels sont dits orthonormés, *i.e.* disposent d'une norme et d'une représentation orthogonale. Dans ce contexte, la norme choisie à l'origine de la fonction de distance appartient généralement à l'ensemble des normes de Minkowski. Elles se définissent par :

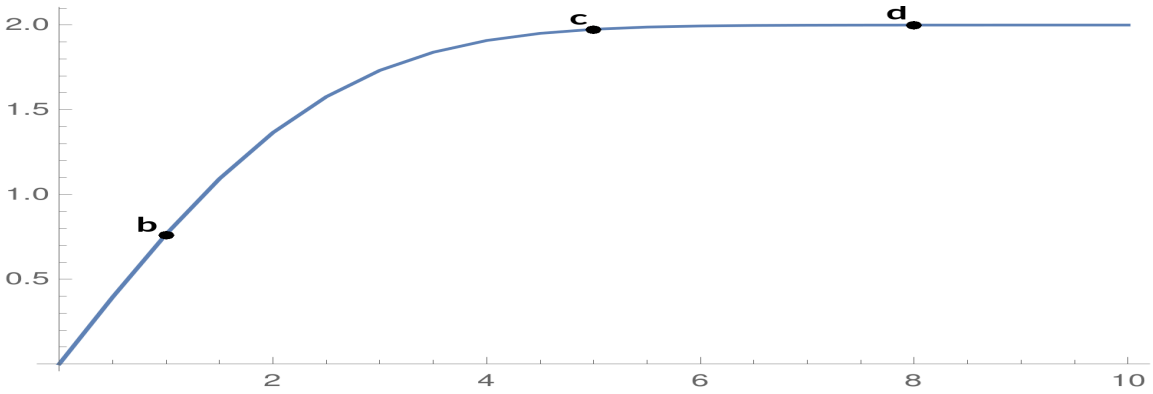
$$\forall (a,b) \in \mathbb{R}^n \, d_p(a,b) = \sqrt[p]{\sum_{i=1}^{n} (a_i - b_i)^p}, \tag{3.2}$$

où $p \in \mathbb{N}$. Pour $p = 1$ nous retrouvons la distance de Manhattan, $p = 2$ correspond à la distance Euclidienne et $p = \infty$ la distance de Tchebychev. Attention, cette écriture de la norme de Minkowski n'est valable que sur une base orthonormale et doit être adaptée si celle-ci ne l'est pas.

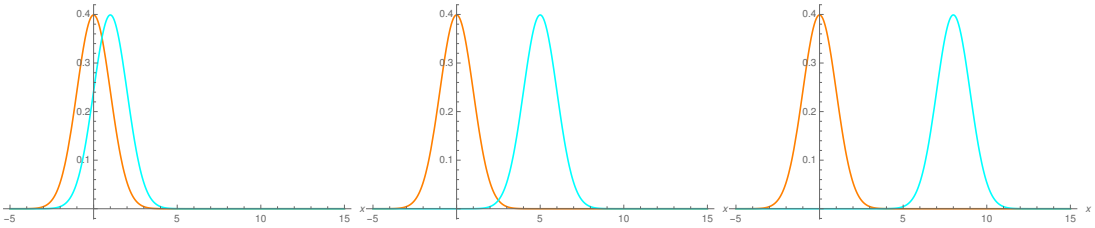
3.1.2 Distances, divergences entre probabilités

La fonction de densités de probabilités (pdf pour *probability density function*) caractérise une loi de probabilité et se définit ainsi :

$$Pr(u < A \leqslant v) = \int_u^v a(x)dx, \tag{3.3}$$



(a) Distance entre une Gaussienne centrée et une de moyenne μ : $d(\mathcal{N}(0,1),\mathcal{N}(\mu,1))$



(b) $d\left(\mathcal{N}(0,1),\mathcal{N}(1,1)\right)$ (c) $d\left(\mathcal{N}(0,1),\mathcal{N}(5,1)\right)$ (d) $d\left(\mathcal{N}(0,1),\mathcal{N}(8,1)\right)$

FIGURE 3.2 – **Schéma de saturation de la norme $\mathcal{L}_2$**. Les figures 3.2b, 3.2c et 3.2d illustrent le chevauchement et l'éloignement des lois normales $\mathcal{N}(\mu,\sigma)$.

pour une variable aléatoire A de pdf a. $Pr(u < A \leqslant v)$ correspond à la probabilité que la variable aléatoire A apparaisse sur l'intervalle $]u,v]$. Puisque nous sommes sur l'espace des probabilités, l'intégrale sur $\mathbb{R}$ de cette pdf doit être unitaire.

3.1.2.1 Distances

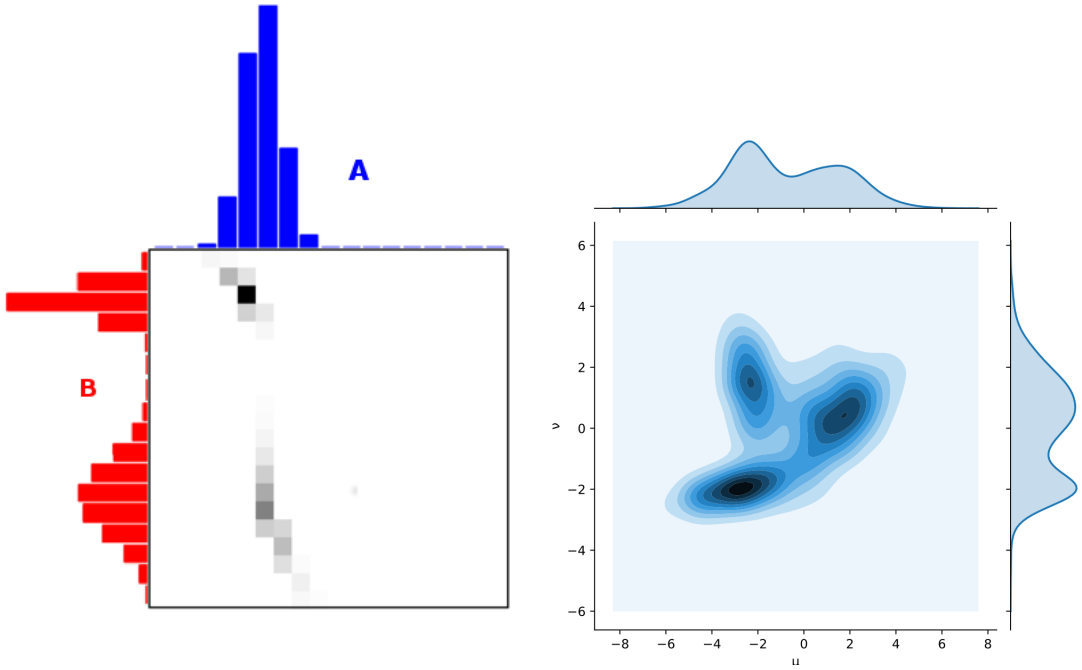
Les pdf forment un sous ensemble des fonctions intégrables sur $\mathbb{R}$, on peut donc appliquer toutes les distances de Minkowski associées à l'ensemble des fonctions intégrables :

$$d_p(a,b) = \left(\int_{\mathbb{R}} (a(x) - b(x))^p \, dx\right)^{\frac{1}{p}}. \tag{3.4}$$

Cependant, dès lors que le support des deux distributions devient disjoint, la mesure de distance sature et ce, quelle que soit l'évolution de l'écart entre les distributions (fig. 3.2). La figure 3.2a représente la distance $\mathcal{L}_2$ entre deux Gaussiennes pour le produit scalaire des fonctions Euclidiennes. La première Gaussienne est normalisée centrée en 0, la seconde est normalisée mais sa moyenne varie. Les figures 3.2b, 3.2c et 3.2d présentent différentes configurations de Gaussiennes. Nous constatons que lorsque les Gaussiennes ne se chevauchent presque plus, la distance tend vers la limite 2, ce qui correspond à la somme de l'intégrale de chaque Gaussienne établie séparément. Si nous nous plaçons dans le contexte couleur et perception humaine, nous sommes effectivement incapables de déterminer quelle couleur est la plus éloignée d'un rouge, un vert ou un bleu. En termes de longueurs d'onde ou d'orientations (dans le cas d'histogrammes de gradients orientés), cette différence existe et nous souhaitons la prendre en compte.

La distance de Hellinger s'applique à des mesures de probabilités, elle est donnée par :

$$d(A,B) = \sqrt{\frac{1}{2}\int_{\mathbb{R}} \left(\sqrt{a(x)} - \sqrt{b(x)}\right)^2 dx}, \tag{3.5}$$



(a) Espace discret (b) Espace continu

FIGURE 3.3 – **Exemple de répartition de distribution pour l'EMD**. Nous avons deux histogrammes associés à des lois de probabilités A et B. L'image en niveau de gris correspond à une possibilité de redistribution de la loi A à la loi B.

où a et b sont les densités de probabilités respectives de A et B. Proche des distances de Minkowski, cela engendrera les mêmes limites (saturation de la distance lorsque les distributions sont presque disjointes).

Une distance très appropriée aux mesures de différence entre distributions est l'*earth mover's distance* (EMD). Elle est aussi appelée la distance de Wasserstein. L'idée associée suppose que la première distribution soit constituée de tas de terre et la seconde de trous à remplir. L'EMD correspond au coût minimum pour transvaser les tas dans les trous. Plus rigoureusement, soit deux mesures de probabilités A et B et $\delta(,)$ une distance sur Ω :

$$d_p(A,B) = \left(\inf_{\gamma \in \Gamma(A,B)} \int_{\Omega \times \Omega} \delta(x,y) d\gamma(x,y) \right)^{\frac{1}{p}}, \tag{3.6}$$

où $\Gamma(A,B)$ correspond à l'ensemble de mesure de $\Omega \times \Omega$ ayant comme marginales respectives A et B. Nous pouvons voir $\Gamma(A,B)$ comme les différentes manières de déplacer le sable des tas A vers les trous B. La figure 3.3 présente deux exemples de mesure γ pour deux histogrammes $1D$, l'un en discret l'autre en continu.

En pratique, on se retrouve dans un problème d'optimisation de transport. La fonction coût a toute son importance et dépend de l'application. Nous devons ensuite résoudre le problème de transport ou problème de Monge-Kantorovich, en minimisant la fonction de coût sous certaines contraintes. C'est un problème NP-complet, sa résolution augmente donc avec la dimension. Si nous considérons les histogrammes de gradient comme des fonctions, cela limite donc le problème à une dimension. Au contraire, si nous travaillons avec dans des espaces couleur, multi- ou hyperspectral le nombre de dimensions explose rapidement.

Benamou et al. (2015) adaptent ce problème de transport pour y utiliser la divergence de Kullback-Leibler afin de projeter sur des espaces de plus petite dimension et gagner ainsi en temps de calcul. Carlier et al. (2017) donnent une preuve de convergence de cet algorithme. Malheureusement, l'optimisation au cœur de cette mesure est très sensible aux faibles variations (bruit) et ne

permet donc pas d'atteindre la robustesse attendue.

3.1.2.2 Divergences

Les divergences sont des mesures de similarité dans les espaces de probabilités, *i.e.* entre deux distributions. Il s'agit de semi-métriques, *i.e.* elles ne respectent pas l'inégalité triangulaire. Posons donc A et B deux lois de probabilité, a et b leurs pdf associées de support respectif $\mathcal{A}$ et $\mathcal{B}$. La divergence de Kullback-Leibler est la plus connue, nous la détaillerons donc dans une sous-section spécifique.

La divergence la plus générale, présentée par Bregman (1967), est une divergence associée à une fonction potentiel U strictement convexe et de classe C^1. Elle ne respecte pas non plus la propriété de symétrie. La divergence de Bregman entre a et b associée à la fonction de potentiel U est définie par :

$$D_U(a,b) = U(a) - U(b) - \langle \nabla U(a), a-b \rangle \,, \tag{3.7}$$

où ∇U est le gradient de la fonction potentiel U.

Selon la fonction potentiel choisie, nous pouvons retrouver plusieurs différences connues. La distance Euclidienne en sommant les produits croisés pour un espace vectoriel nD ou la distance de Mahalanobis en pondérant cette somme. La divergence de Kullback-Leibler s'obtient avec le potentiel :

$$U(a) = \int_{\mathcal{A}} a(x) \ln(a(x)) dx. \tag{3.8}$$

On appelle f-divergence, la mesure de différence de A par rapport à B pour la fonction convexe f :

$$D_f(A \parallel B) = \int_{\mathcal{B}} a(x) f\left(\frac{a(x)}{b(x)}\right) dx. \tag{3.9}$$

pour une fonction définie par $f(a) = a\log(a)$, nous retrouvons la divergence de Kullback-Leibler.

Une α-divergence est définie par :

$$D_\alpha(a \parallel b) = \lim_{x \to \alpha} \frac{1}{x-1} \ln\left(\int_{\mathcal{B}} \frac{a(x)}{b(x)} dx\right), \tag{3.10}$$

pour $\alpha \in \mathbb{R}^+ \bigcup \infty$. Cette famille est aussi appelée divergence de Rényi (1961). Dans le cas limite où $\alpha = 1$, nous retrouvons la formulation de la divergence de Kullback-Leibler basés sur l'entropie de Shannon.

3.1.2.3 Le cas particulier de la divergence de Kullback-Leibler

Parmi les mesures de distance/similarité existantes, la divergence de Kullback-Leibler (KL) est la plus utilisée et réputée stable (Kullback and Leibler (1951)). La mesure d'information de KL estime la mesure de dissimilarité entre deux lois de probabilités A et B. Cette mesure est donnée par :

$$D_{KL}(A/B) = \int_{\mathcal{B}} a(x) \ln\left(\frac{a(x)}{b(x)}\right) \mathrm{d}x, \tag{3.11}$$

où a et b sont les densités de probabilité de A et B respectivement, $\mathcal{B}$ correspond au support de B. Dans le contexte discret des images, la densité mesurée à partir de l'histogramme pour certaines coordonnées peut être nulle.

Afin de respecter la propriété de symétrie, Kullback et Leibler ont défini la divergence de KL comme la somme des deux mesures d'information $D_{KL}(A,B) = D_{KL}(A/B) + D_{KL}(B/A)$ (aussi

connue en tant que divergence de Jeffrey). Malheureusement, la divergence de KL n'est pas une distance (non respect de l'inégalité triangulaire).

Cette divergence fait appel à l'entropie de Shannon qui mesure la quantité d'information transmise par une source. La divergence KL s'intéresse à la similarité entre deux sources délivrant de l'information. C'est cette formulation via l'entropie qui induit des réponses en accord avec nos attentes. L'entropie implique l'écriture en $\log$ qui empêche l'inégalité triangulaire et donc l'obtention d'une distance à proprement parler.

L'estimation de cette divergence par une intégration discrète peut être coûteuse, surtout lorsque la dimension de l'espace vaut 3 ou plus. Cependant, il existe différentes méthodes pour accélérer ce processus moyennant une perte de précision comme le présentent Wang et al. (2009) ou Goldberger et al. (2003). Le second problème est lié à l'aspect discret des fonctions de densité de probabilité, induisant des probabilités nulles pour certaines coordonnées.

3.1.3 Autres distances

3.1.3.1 Distance entre séquences d'objets à nombre d'états finis

Nous pouvons aussi définir la distance de Hamming (1950) pour les ensembles de type $\mathbb{Z}/p\mathbb{Z}^n$. p représente le nombre de valeurs que peut prendre un élément : deux dans le cas binaire, 26 pour l'alphabet romain, etc. Cette distance est simplement l'addition du nombre d'éléments différents entre les vecteurs à comparer. Par exemple, pour un vecteur binaire de taille huit :

$$11010010,$$
$$10011011,$$

la distance entre ces deux ensembles est de 3. Une généralisation de cette distance est la mesure de Levenshtein qui mesure la distance entre deux mots.

Une autre estimation de différence de rang peut-être le τ de Kendall (1938). Cet opérateur estime le nombre de permutations nécessaire pour obtenir un arrangement à partir d'un autre. Cette méthode n'étant pas très rapide en temps de calcul.

3.1.3.2 Distance topologique

La distance de Haussdorf a été construite pour mesurer des écarts topologiques. En traitement d'image , elle est utilisée pour comparer ou apparier des formes d'objets. Sa définition pour deux ensembles A et B de l'espace $\mathcal{E}$ est donnée par :

$$d(A,B) = \max\left(\sup_{b\in B}\inf_{a\in A}\delta(a,b), \sup_{a\in A}\inf_{b\in B}\delta(a,b)\right), \tag{3.12}$$

où a et b sont des éléments de leurs ensembles respectifs et $\delta(,)$ correspond à la distance associée à l'espace $\mathcal{E}$.

Cette distance est notamment utilisée pour la morphologie mathématique.

3.1.4 Mesures intégrant une modélisation statistique

Nous nous plaçons dans le cas de mesure entre pdf. Les données sont modélisées, le plus souvent, par une loi normale. Ce choix est souvent justifié par le gain en temps de calcul, et en occupation

mémoire (remplacer une distribution complète par quelques scalaires). Dans le cas des réalisations discrètes, cela résout les problèmes de parcimonie dans des espaces continus.

Dans ce cas, pour mesurer la distance d'un point a à une probabilité P de matrice de variance-covariance Σ et de moyenne μ, la méthode la plus appropriée est la distance de Mahalanobis (1936) :

$$d_\Sigma(a) = \sqrt{(a-\mu)^T \Sigma^{-1} (a-\mu)}. \tag{3.13}$$

La distance du χ^2 est associée à celle de Mahalanobis et est utilisée pour estimer l'accord entre une hypothèse effectuée sur les données et les données elles-mêmes.

Montesinos et al. (1998); Ojala et al. (2002); Mikolajczyk and Schmid (2005) utilisent la distance de Mahalanobis pour résoudre le problème de saturation puisque la mesure se calcule face à une distribution. Cependant, nous sommes en présence de deux distributions et non d'un point face à une distribution, le problème sera donc le même. Cette limite de construction impacte également des mesures dérivées de la norme de Minkowski ou du χ^2 van de Sande et al. (2010); Mandal et al. (2015) qui sembleraient à priori plus adaptées.

Le calcul de la divergence de Kullback-Leibler peut s'avérer complexe, nous proposons dans la sous-section suivante de modéliser la loi de probabilité pour simplifier le calcul de mesure de similarité.

3.1.4.1 La divergence de KL dans le cas d'une loi normale

Pour diminuer le coût calcul et éviter les supports de distribution discrets qui impliquent souvent des probabilités nulles en certains points, nous proposons de modéliser la distribution A dans un espace de dimension n par une loi multinormale $\tilde{A}$. Celle-ci est caractérisée par une coordonnée moyenne $\mu_P \in \mathbb{R}^n$ et une matrice de variance-covariance Σ_A de taille $n \times n$:

$$\tilde{A} = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_A|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu_A)^t \Sigma_A^{-1}(x-\mu_A)\right). \tag{3.14}$$

L'intérêt d'une approximation de la distribution couleur A en une ellipsoïde tridimensionnelle $\tilde{A}$ réside dans la solution directe au calcul de la mesure de l'information de Kullback-Leibler (eq. 3.15) et, par là-même, à la mesure de la divergence associée :

$$\begin{aligned} D_{KL}(A/B) &\approx D_{KL}(\tilde{A}/\tilde{B}) \\ &\approx \frac{1}{2}\Bigg(\ln\left(\frac{|\Sigma_B|}{|\Sigma_A|}\right) + \operatorname{tr}\left(\Sigma_B^{-1}\Sigma_A\right) + \\ &\qquad (\mu_B-\mu_A)^t \Sigma_B^{-1}(\mu_B-\mu_A) - n\Bigg). \end{aligned} \tag{3.15}$$

Avec la symétrisation nous obtenons :

$$\begin{aligned} D_{KL}(A,B) &\approx D_{KL}(\tilde{A},\tilde{B}) \\ &\approx \frac{1}{2}\Bigg(\operatorname{tr}\left(\Sigma_B^{-1}\Sigma_A\right) + (\mu_B-\mu_A)^t \Sigma_B^{-1}(\mu_B-\mu_A) + \\ &\qquad \operatorname{tr}\left(\Sigma_A^{-1}\Sigma_B\right) + (\mu_A-\mu_B)^t \Sigma_A^{-1}(\mu_A-\mu_B) - 2n\Bigg). \end{aligned} \tag{3.16}$$

3.1.4.2 Limites de la modélisation

Afin de vérifier la modélisation choisie, un test d'hypothèses doit être réalisé. Il permet de valider le choix du modèle multinormal pour approximer la distribution tridimensionnelle couleur. Dans le cas où le test indique une inadéquation au modèle, il est toujours possible de revenir vers un calcul direct de la divergence de Kullback-Leibler, plus coûteuse en temps calcul et requérant une gestion des valeurs nulles de la distribution.

3.2 Nature des données et mesures utilisées

Les descripteurs de points d'intérêt que nous présenterons au prochain chapitre, sont des ensembles de mesure. Selon leur nature, ils peuvent être considérés comme des listes, des vecteurs indépendants, des distributions de probabilités, etc. En fonction de la nature de notre ensemble, telle distance ou telle divergence sera plus adaptée. Dans cette section, nous allons présenter les types de données utilisés par les descripteurs et les distances qui leurs sont attribuées dans la littérature.

La métrologie requiert l'établissement de mesures, idéalement de distance. Malheureusement, la plupart des mesures entre distributions sont des mesures de divergence qui ne respectent pas l'inégalité triangulaire. Nous utiliserons donc le terme générique de différence, recommandé par la CIE et englobant les deux possibilités dans le contexte d'écarts entre couleurs.

La question de la distance à utiliser n'est pas nouvelle, Mandal et al. (2015) et Mikolajczyk and Schmid (2005) ont sélectionné diverses distances et différents descripteurs afin de comparer les résultats des paires distance/descripteur.

La figure 3.4 présente les distances utilisées en fonction du type de données pour quelques descripteurs couleurs. Tout d'abord nous présenterons les descripteurs de nature séquences de bit ou listes (sec. 3.2.1). Un type de liste régulièrement trouvé dans toutes les familles de descripteurs est la concaténation. Il s'agit d'accoler des données de nature identiques ou différentes. Cependant deux histogrammes mis bout à bout ne font pas un histogramme. Il n'y a pas à notre connaissance de descripteur de nature vectorielle. La majorité des descripteurs sont de type histogramme donc des densités de probabilités (sec. 3.2.2). La plupart des auteurs les traitent comme des vecteurs de $\mathbb{R}^n$, ce qui donne lieu à l'emploi de la norme de Minkowski. Dans un premier temps, nous verrons quelles sont les limites de ces normes face aux données à traiter. Puis nous les analyserons comme données de probabilités. Leur nature suggère l'emploi de mesure de type divergences ou distance du cantonnier. Les données de type histogramme sont gourmandes en ressource et souvent une modélisation est employée pour réduire leur dimension, la section suivante 3.2.3 présentera les approches et leur distance associée. Enfin, la question des distributions multidimensionnelles et notamment couleur sera étudiée.

3.2.1 Séquences des bits

Les descripteurs binaires, tel que BRIEF (Calonder et al. (2010)) ou BRISK (Leutenegger et al. (2011)), sont composés de résultats binaires et se définissent comme un ensemble d'éléments de type booléen :

$$b = \begin{cases} 1 & \text{si } I(x) < I(y) \\ 0 & \text{sinon .} \end{cases} \tag{3.17}$$

On obtient donc une signature Sig de la région d'intérêt ROI (*Region Of Interest*) représentant par un mot, binaire dans les exemples cités :

$$Sig_{BRIEF}(ROI) = \{b_1, b_2, .., b_i, .., b_n\}, \tag{3.18}$$

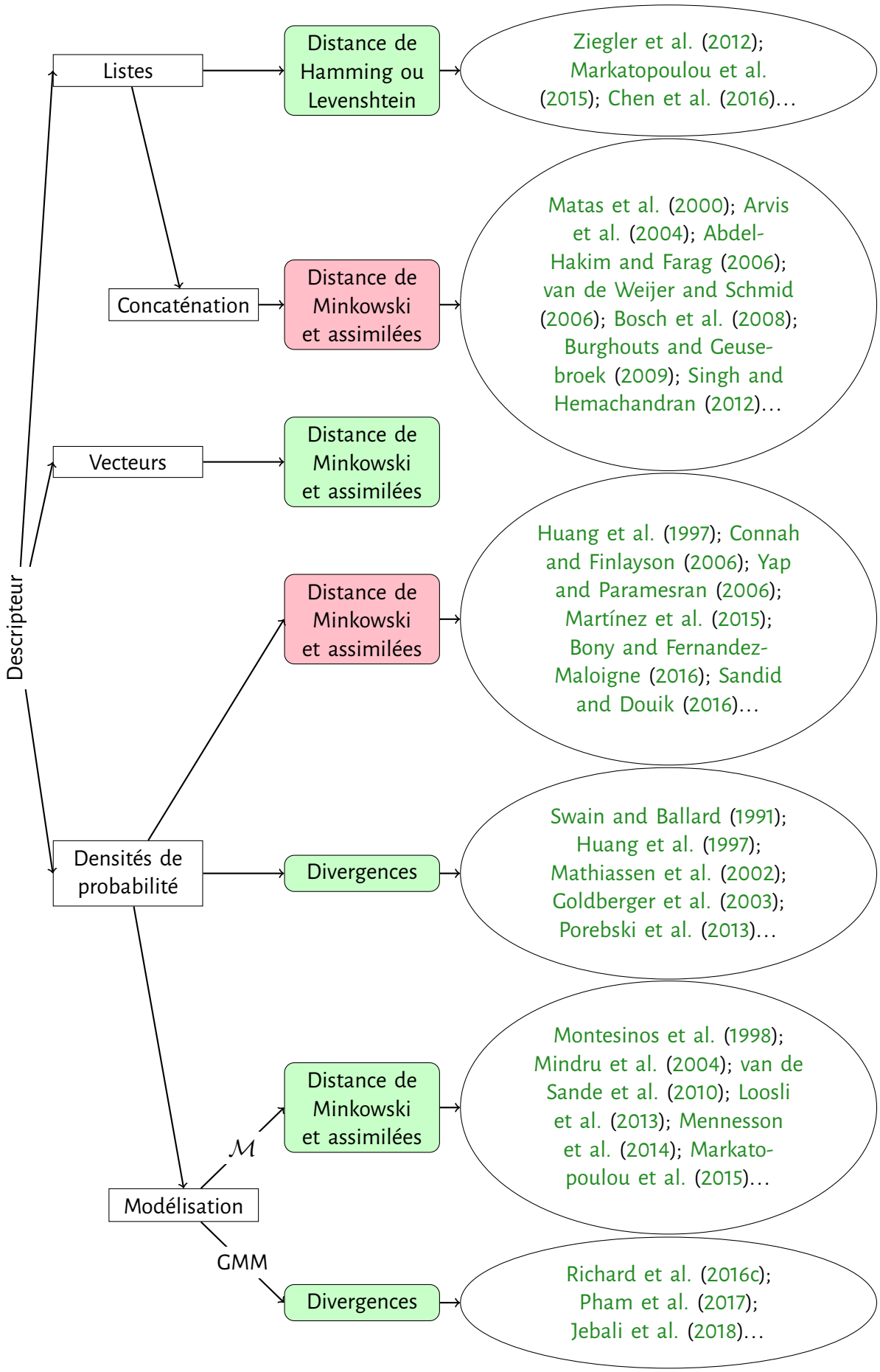


FIGURE 3.4 – **Organisation des distances utilisées en fonction du type des données du descripteur dans la pratique**. Contrairement à la figure 3.1, les distances utilisées ne sont pas toujours en accord avec le type de données.

où n est la longueur du mot. La suite d'éléments ainsi obtenus peut être considérée comme un code (mot binaire de longueur fini) autorisant l'utilisation de la distance de Hamming.

Ces descripteurs sont définis comme un vecteur de $\mathbb{Z}/2\mathbb{Z}$. Nous estimons qu'il s'agit d'un vecteur car l'ordre dans lequel les comparaisons sont effectuées n'a pas d'importance, contrairement aux descripteurs basés gradients.

Avec l'arrivée des descripteurs binaires, de nouvelles distances devraient être utilisées pour être en cohérence avec le modèle des données. Le coût combinatoire est souvent une justification à la recherche de solutions alternatives plus rapide. Ziegler et al. (2012) présentent les distances de corrélations d'ordre et choisissent l'emploi de la distance de Hamming. Calonder et al. (2010); Leutenegger et al. (2011); Rublee et al. (2011); Alahi et al. (2012) sélectionnent la distance de Hamming pour évaluer les différences entre descripteurs.

Ojala et al. (1994) définissent des textons par comparaison binaire et créent les *Local Binary Pattern* (LBP). Le descripteur LBP se compose d'un histogramme de textons :

$$\begin{aligned} texton &= \{b_1, b_2, .., b_i, .., b_n\}, \\ Sig_{LBP}(ROI) &= hist(texton), \end{aligned} \tag{3.19}$$

où n est généralement égal à 8 et $hist$ correspond à l'histogramme des différents textons estimés autour de la zone d'intérêt. Cette construction permettant ainsi d'utiliser la divergence de Kullback-Leibler comme mesure de similarité comme le fait Ojala et al. (1994). Lors de l'adaptation des LBP au multi-échelles, Ojala et al. (2002) utilisent la distance de Mahalanobis.

Bien que les distances soient adaptées aux natures des données, la mise en œuvre abandonne ici toute une partie de l'information, induisant une perte d'exactitude non maîtrisée.

3.2.2 Les histogrammes comme densités de probabilité

Un grand nombre de descripteurs utilisent la forme d'histogramme comme signature. Rubner et al. (1998) prennent d'ailleurs le temps de définir la différence entre histogramme et signature.

Parmi ceux-ci, Freeman and Roth (1995); Lowe (2004), par exemple, sont constitués d'histogrammes de gradient. Freeman and Roth (1995) mesurent l'histogramme de l'orientation du gradient sur toute la zone d'intérêt :

$$Sig_{HOG}(ROI) = hist_\theta(\nabla I_{ROI}(x, y)), \tag{3.20}$$

où I_{ROI} est l'image restreinte à la région d'intérêt et $hist_\theta$ correspond à l'histogramme sur les orientations.

Lowe (2004), avec l'incontournable SIFT, reprend l'idée des histogrammes de gradients orientés, mais combine plusieurs histogrammes pour garder une partie de l'information spatiale :

$$Sig_{SIFT}(ROI) = \cup_{P \in ROI} \left(hist_\theta(\nabla I_P(x, y))\right), \tag{3.21}$$

où P est une partition de la zone d'intérêt.

La famille à laquelle appartiennent les descripteurs cités précédemment se base sur des histogrammes de gradients. Nous ne pouvons donc pas dire qu'il s'agit d'un vecteur de $\mathbb{R}^n$. L'ordre des données a une importance et les données ne peuvent être permutées sans modifier l'information contenue. Nous sommes donc face à des données probabilistes.

Nous devons donc considérer ces histogrammes comme des densités de probabilités. Il faut donc utiliser des distances appropriés à ce type de données telles que l'*earth mover's distance* ou la divergence de Kullback-Leibler.

Certains emploient des distances appropriées au type de donneés. Rubner et al. (1998) utilisent l'EMD sur des histogrammes de réponses fréquentielles. L'EMD est aussi utilisée par Lazebnik et al. (2003) sur des histogrammes d'occurrences d'intensité.

Swain and Ballard (1991) proposent une mesure de similitude entre histogrammes. Le calcul proposé ne respecte pas la propriété de symétrie, Smith and Chang (1996) adaptent l'écriture pour proposer une distance.

D'autres, au contraire, n'adaptent pas la distance aux données. Par exemple, Lowe (2004); Freeman and Roth (1995) utilisent la mesure de distance la plus largement répandue : la norme de Minkowski d'ordre 2.

3.2.3 Modélisation ou paramétrisation d'une pdf

Beaucoup de descripteurs proposent d'extraire des caractéristiques des distributions (histogrammes) extraites de la zone d'intérêt. Bay et al. (2008), proposent le descripteur SURF et somment les dérivées partielles sur la zone d'intérêt :

$$Sig_{LBP}(ROI) = \{\sum dx; \sum dy; \sum |dx|; \sum |dy|\}, \tag{3.22}$$

où dx et dy correspondent aux dérivées partielles selon les axes. Ke and Sukthankar (2004) utilisent une analyse en composante principale sur le descripteur SIFT pour accélérer le temps d'exécution.

Un autre type de données exprimé pour la création d'un descripteur est le moment statistique. Les moments sont des caractéristiques extraites d'une loi de probabilité. Haralick (1979) calcule les moments à partir de la distribution des occurrences d'intensité de la zone d'intérêt. van Gool et al. (1996) emploient des invariants affines et photométriques pour construire un descripteur. Nous pouvons modéliser ce type de données ainsi :

$$Sig_{Moments}(ROI) = \{\mathcal{M}^i, i \in [1, n]\}, \tag{3.23}$$

Face à ce type de données quelle distance utiliser ? Les moments ne sont pas orthogonaux entre eux, il est donc délicat d'employer les distances de type Minkowski. de plus, un ordre possible correspond à celui de l'ordre des moments, mais comment ordonner les moments $2D$ ou plus (Haralick (1979); Connah and Finlayson (2006)). Les considérer comme des sacs de mots est envisageable mais l'information grandeur de la différence est alors perdue. Or celle-ci n'est pas anodine sur des moments par exemple. L'utilisation en composante principale crée effectivement une donnée de type vecteur, et Ke and Sukthankar (2004) utilisent la distance Euclidienne. Cependant l'analyse se fait sur un ensemble d'images définies. La notion d'apprentissage implique nécessairement une adaptation à l'application et perd sa généricité.

Ce sont des idées intéressantes mais le problème est mal posé car il y a dissociation entre la nature des données et la distance. Or, avec la classification ou l'indexation en tête, nous savons que l'application finale est une comparaison entre distances pour obtenir des plus proches voisins ou un étiquetage.

3.2.4 Extension aux contextes couleur et spectral

En se plaçant dans le cas spécifique où le descripteur extrait une distribution couleur, la question de la distance à utiliser se pose aussi. Nous retrouvons les deux types précédents de signatures, c'est à dire des histogrammes et une extraction de paramètres, majoritairement des moments associés aux distributions étudiées.

Différentes paramétrisations des données existent. van de Sande et al. (2010) calculent les moments couleur de l'image. Montesinos et al. (1998) extraient, quant à eux, des moments différentiels invariants sur les canaux couleur. Ces données peuvent être modélisées avec l'équation (3.23).

D'une autre manière, Arvis et al. (2004) combinent des moments texture extraits d'un filtrage de Gabor avec des moments couleur dans l'espace HSI. La modélisation d'un tel descripteur peut s'écrire de cette manière :

$$Sig_{Moments}(ROI) = \{\mathcal{M}^i(text), i \in [1,n]\} \cup \{\mathcal{M}^j(coul), j \in [1,m]\}, \tag{3.24}$$

où n est le nombre de caractéristiques texture et m le nombre de caractéristiques couleur.

Nous pouvons nous poser la même question que précédemment sur la distance à utiliser. Il sera d'autant plus difficile d'y répondre avec une signature de type donné par l'équation (3.24) où la nature des données pour extraire les moments change dans la signature. Les combinaisons de signature amplifient le problème initial.

Montesinos et al. (1998) emploient la distance de Mahalanobis sur les signatures. van de Sande et al. (2010) utilisent la distance du χ^2.

Lorsque la nature des données s'apparente à des pdf, d'autres distances sont utilisées. Parmi celles-ci, Liu et al. (2013); Maliani et al. (2014) proposent une distance géodésique entre distributions $3D$, prenant en compte une partie réduite du problème. Une alternative duale à cette construction est Cao et al. (2011) qui exploitent une distance de chanfrein entre deux ensembles. Malheureusement, la mesure de la différence entre deux distributions ne se limite pas à l'écart entre les supports trichromatiques, mais doit intégrer également les variations de densité.

L'usage d'une divergence de Kullback-Leibler (KL) n'est pas nouveau pour estimer la différence entre deux distributions couleur. Goldberger et al. (2003) mélangent ainsi coordonnées couleur dans $CIELAB$ et positions spatiales pour calculer une approximation de la divergence de Kullback-Leibler. Mathiassen et al. (2002) considèrent, quant à eux, une caractéristique de texture basée sur les filtres de Gabor puis évaluent la similarité entre les distributions avec la divergence de KL. Plus récemment, ce type de divergence a été utilisé en combinaison avec des classifieurs de types KNN (Wang et al. (2009), ou Liu and Shum (2003)). La divergence de KL prend naturellement en compte l'aspect distribution de valeurs et de variations dans une forme cohérente.

Le contexte couleur impose de discuter du choix de l'espace couleur de travail. Comme souvent, il faut distinguer le cas perceptuel du cas physique.

Une application n'impliquant pas la comparaison du résultat à une analyse humaine, peut s'affranchir de la contrainte perceptuelle et envisager de travailler dans l'espace d'acquisition RGB. En restant dans cet espace, la matrice de Gram introduite dans le chapitre 1, nous permet de mesurer des différences couleur de façon métrologique. Cependant, cette comparaison se fait par paires de couleurs. Et il faudrait ici étendre cette différence aux distributions.

Pour une application nécessitant une corrélation avec le système visuel humain, travailler directement dans l'espace RGB d'acquisition n'est plus possible. En effet, nous savons que les canaux sont fortement corrélés dans cet espace. De plus les différences de couleurs calculées dans RGB ne sont pas perceptuellement uniformes. En conséquence, il serait impossible de relier les résultats d'une expérimentation psychophysique aux différences calculées (ce qui n'exclut pas que la corrélation soit valide pour quelques cas particuliers).

Depuis quelques années, les résultats se succèdent pour démontrer l'intérêt d'un traitement dans des espaces où les métriques de différence ont été validées. Typiquement, Bello-Cerezo et al. (2016) montrent que les résultats de classification sont presque toujours meilleurs dans l'espace $CIELAB$, notamment lorsque les informations d'étalonnage et/ou de calibration sont disponibles. L'espace $CIELAB$ est recommandé par la CIE pour l'évaluation des différences perceptuelles de couleur.

Dans ce travail, nous mesurerons les distances dans les deux espaces RGB et $CIELAB$. Notre objectif est seulement de mettre en avant la nécessité d'adapter la mesure au type de données et

non spécifiquement une application perceptuelle ou physique. D'autres espaces préconisés par la CIE auraient pu être considérés comme le $CIELUV$, ou d'autres versions plus récentes. Néanmoins la dynamique des différences couleur considérées est susceptible d'être importante et très supérieure à la plage de validation de ces distances plus récentes (inférieure à 5 unités).

Nous nous sommes limités ici aux espaces couleur, mais lorsque nous traitons des images multi ou hyper-spectrale le nombre de dimensions augmente. Comme nous souhaitons garder le lien avec la physique de la scène acquise, l'utilisation des espaces décorrélés tel qu'obtenus avec l'analyse en composante principale ou la transformation de Karhunen et Loeve ne sont pas appropriés puisque dépendantes de l'information contenue dans l'image. En plus grande dimension, la question de l'espace couleur à utiliser ne se pose plus, cependant, la métrique à utiliser reste une question ouverte.

3.3 Éléments de décision d'une mesure de similarité

Les données que nous traitons sont majoritairement des histogrammes d'une ou plusieurs dimensions. C'est pourquoi nous optons pour un type de mesure permettant de comparer des densités de probabilité. La distance de Hellinger créée pour les espaces de probabilités, possède les mêmes limitations que la norme $\mathcal{L}_2$, elle sature. Parmi les fonctions potentielles classiquement utilisées, l'entropie est celle qui rend le mieux compte de la disparité des données. C'est pour cela que nous avons choisi la divergence de Kullback-Leibler.

Pour valider cette mesure de divergence, nous allons effectuer deux tests. Le premier vise à valider le comportement de la mesure en fonction des variations entre les distributions. Le second test applique la mesure de divergence entre différentes images pour en comprendre la dynamique en condition d'usage.

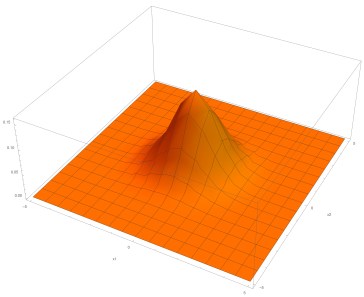
3.3.1 Comportement avec une distribution de synthèse

Si la divergence de Kulback-Leibler est bien connue pour la comparaison de fonctions de densités de probabilité unidimensionnelle, elle est peu utilisée dans des dimensions supérieures. Avant de l'exploiter entre des histogrammes trichromatiques, nous proposons d'en analyser le comportement dans le cadre de fonctions de densités de probabilité bidimensionnelle pour des facilités d'affichage.

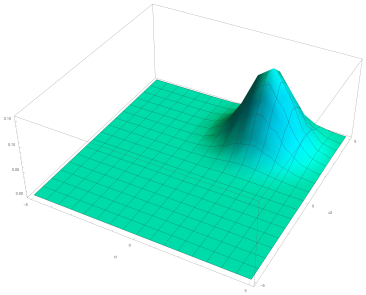
Afin d'utiliser la divergence de Kullback-Leibler nous nous restreignons aux densités de probabilité, autrement dit les histogrammes couleur seront normalisées. **La contrainte d'usage que nous fixons est que la distance ou mesure de similarité entre une pdf de référence et une pdf transformée à partir de celle-ci soit une fonction variant avec l'amplitude de la transformation**. Nous nous comparerons à la distance $\mathcal{L}_2$ des fonctions intégrables sur $\mathbb{R}^n$ en utilisant la pdf de la loi normale utilisée.

Nous allons considérer deux fonctions de type loi normale bidimensionnelle, définies par une position moyenne dans l'espace de représentation et une matrice 2×2 de variance-covariance. La première sera la loi de référence et nous ferons varier les paramètres de la seconde pour analyser le comportement de la mesure de similarité. Nous allons faire varier successivement l'écart entre les distributions (variation de la position moyenne), l'écart-type puis l'orientation des deux lois. En effet, une loi normale en deux dimensions dépend de trois paramètres : sa moyenne, son écart-type et son orientation. Ces paramètres se résument par la moyenne et la matrice de variance-covariance de la loi normale :

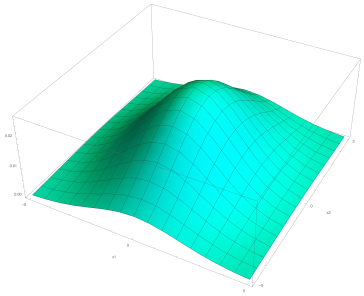
$$\begin{aligned} \mu &= (\mu_1, \mu_2), \\ \Sigma &= \begin{pmatrix} \sigma_1 & \rho \\ \rho & \sigma_2 \end{pmatrix}, \end{aligned} \tag{3.25}$$



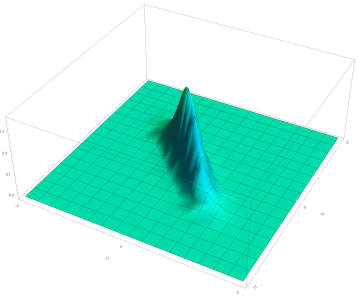
(a) Loi normale centrée réduite



(b) Variation sur la moyenne



(c) Variation sur l'écart-type



(d) Variation sur l'orientation

FIGURE 3.5 – **Illustration des variations d'une loi normale** $2D$. Nous mesurerons la différence entre la loi normale centrée réduite (fig. 3.5a) et les variations de la loi normale (fig. 3.5b, 3.5c, 3.5d).

où σ_i correspond à l'écart-type sur la direction i et ρ représente la covariance c'est à dire la corrélation entre les deux directions. En pratique, pour le cas $2D$, ρ oriente la Gaussienne sur un axe différent du repère normal. La figure 3.5 illustre les variations de chacun des paramètres indépendamment.

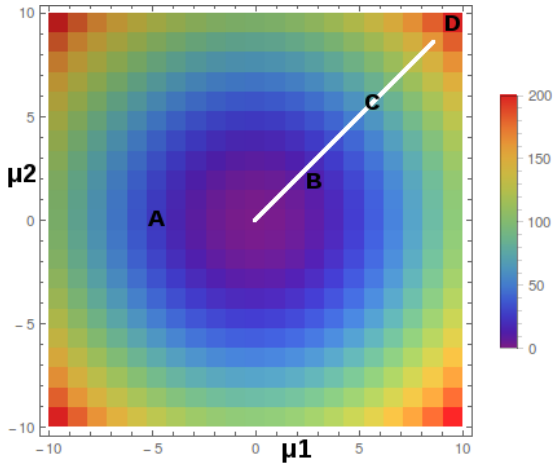
Pour la suite de cette section, les illustrations de Gaussiennes orange correspondent aux lois normales de référence, celles cyan ou bleu sont celles à qui la référence est comparée. Elles seront illustrées sur un même graphique car il est ainsi plus facile, à l'œil, d'en saisir les différences.

3.3.1.1 Réponse à un écart chromatique

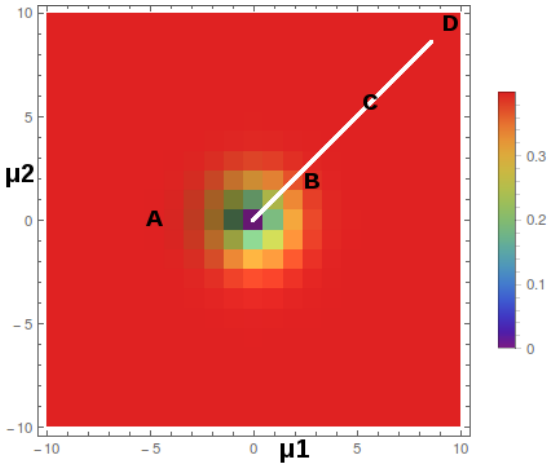
Dans ce premier test, nous considérons deux lois normales de même variance unitaire et de covariance nulle. Nous allons faire varier l'écart chromatique entre les deux lois, *i.e.* la différence entre les moyennes des deux lois.

La figure 3.6a présente les résultats mesurant la divergence de KL entre la loi normale centrée réduite $R = \mathcal{N}((0,0), I)$ et une loi normale réduite de moyenne variable $P = \mathcal{N}((\mu_1, \mu_2), I)$ où I est la matrice identité. De même la figure 3.6b montre la réponse de la distance $\mathcal{L}_2$ entre les deux distributions. Le point O correspond au cas où les deux distributions se superposent, la divergence est donc nulle. Le point A correspond à la divergence de KL entre les deux distributions de la figure 3.6c, la Gaussienne orange étant associée à la loi normale R et la cyan à la loi $\mathcal{N}((-5,0), I)$. La figure 3.6d présente le cas B où les deux distributions semblent plus proches. Cela est vérifié par la valeur de la divergence qui est plus faible pour la figure 3.6d ($D_KL(R,B) = 13$) que pour la figure 3.6c ($D_KL(R,A) = 25$). Comme dans le cas $1D$ présenté sur la figure 3.2, la distance $\mathcal{L}_2$ sature rapidement empêchant de distinguer les distributions C et D de la figure 3.6e (zone rouge de la figure 3.6b).

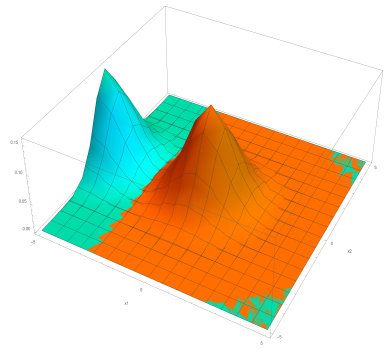
La figure 3.6e présente deux distributions (cyan-C et bleu-D) comparées à la distribution de référence (orange). Les supports des distributions sont presque disjoints, impliquant une distance $\mathcal{L}_2$ quasiment identique pour les deux mesures (0.399). L'évolution de la divergence de KL est quant à elle strictement croissante, comme attendue ($D_{KL}(R,C) = 72$, $D_{KL}(R,D) = 200$).



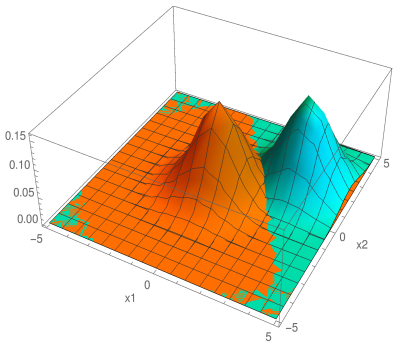
(a) Évolution de la divergence de KL en fonction de la moyenne μ_P : $D_{KL}(R, \mathcal{N}((\mu_1, \mu_2), I))$



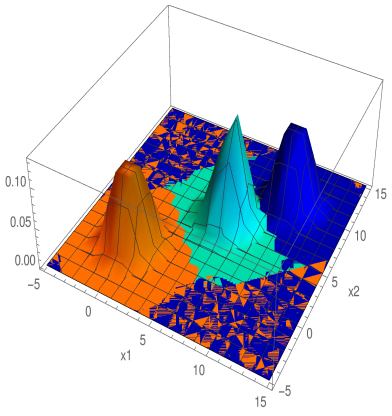
(b) Évolution de la norme $\mathcal{L}_2$ en fonction de la moyenne μ_P : $D_{\mathcal{L}_2}(R, \mathcal{N}((\mu_1, \mu_2), I))$



(c) $\boldsymbol{\mu_A} = (-5, 0)$:
$D_{KL}(R, A) = 25.000$,
$D_{\mathcal{L}_2}(R, A) = 0.398$.



(d) $\boldsymbol{\mu_B} = (3, 2)$:
$D_{KL}(R, B) = 13.000$,
$D_{\mathcal{L}_2}(R, B) = 0.391$.



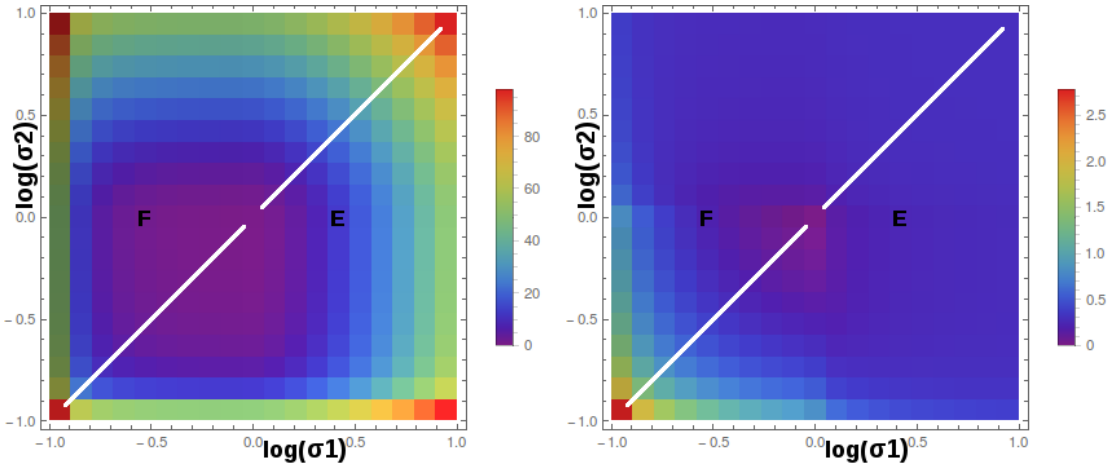
(e) $\boldsymbol{\mu_C} = (6, 6)$:
$D_{KL}(R, C) = 72.000$,
$D_{\mathcal{L}_2}(R, C) = 0.399$;
$\boldsymbol{\mu_D} = (10, 10)$:
$D_{KL}(R, D) = 200$,
$D_{\mathcal{L}_2}(R, D) = 0.399$.

FIGURE 3.6 – **Réponse de la divergence de KL à l'éloignement** de deux lois normales réduites. La référence R est centrée en $(0, 0)$, la seconde P varie selon $\mu_P = (\mu_1, \mu_2)$. Les figures 3.6c, 3.6d, 3.6e illustrent les Gaussiennes pour mieux se rendre compte des différences avec la référence.

Notons que lorsque les deux lois ont la même matrice de corrélation , nous retrouvons la formulation de la distance de Mahalanobis. Dans le cas où $\Sigma = I$, la divergence vaut la distance Euclidienne entre les moyennes des deux distributions : $D_KL(N, P) = \|\mu_N - \mu_P\|^2$.

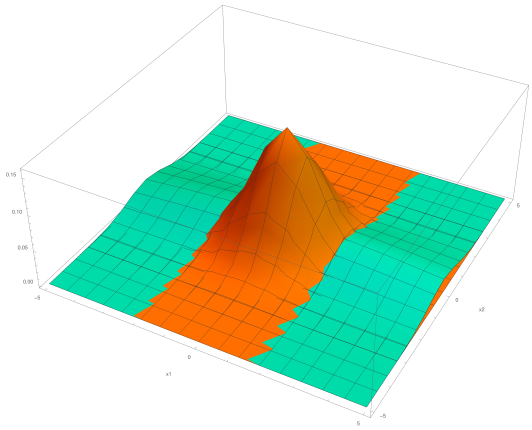
3.3.1.2 Réponse à un changement de forme

Dans ce second test, nous considérons deux lois multinormales centrées à l'origine et de covariance nulle. La loi de référence est toujours la loi normale centrée réduite, *i.e.* sa variance est unitaire ($\sigma_R = (1, 1)$). La loi à laquelle elle est comparée, explore des variances comprises entre 1 et 10. Les figures 3.7c et 3.7d présentent deux exemples de configuration des deux densités de probabilité, un avec une variance supérieure à 1 et l'autre une variance inférieure à 1. Les figures 3.7a et 3.7b présentent la réponse de la mesure de divergence de KL et de la distance $\mathcal{L}_2$ respectivement aux variations σ_1 et σ_2 de la seconde loi. L'échelle est logarithmique permettant de voir les variations de 0.1 à

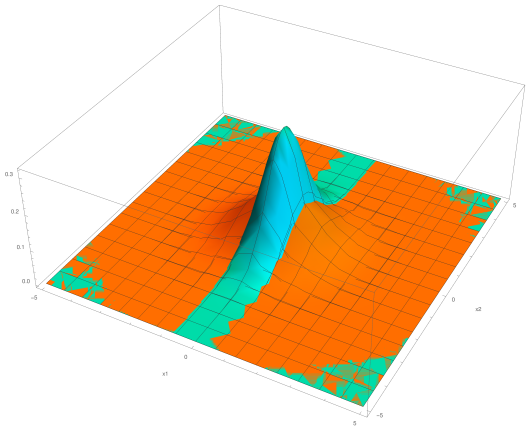


(a) Évolution de la divergence de KL en fonction de σ_P :

(b) Évolution de la norme $\mathcal{L}_2$ à en fonction de σ_P :



(c) $\boldsymbol{\sigma_E} = (5, 1)$:
$D_{KL}(R, E) = 11.520$,
$D_{\mathcal{L}_2}(R, E) = 0.227$.



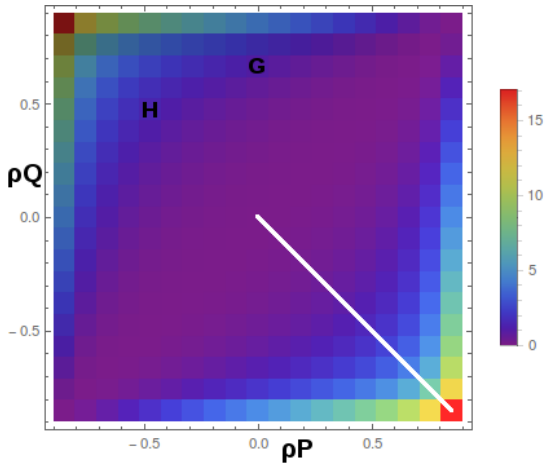
(d) $\boldsymbol{\sigma_F} = (0.5, 1)$:
$D_{KL}(R, F) = 1.125$,
$D_{\mathcal{L}_2}(R, F) = 0.193$.

FIGURE 3.7 – **Réponse de la divergence de KL à un changement de variance**. La loi de référence R est la loi normale centrée, réduite. La seconde P est centrée sans corrélation entre les axes, seule sa variance varie $\sigma_P = (\sigma_1, \sigma_2)$.

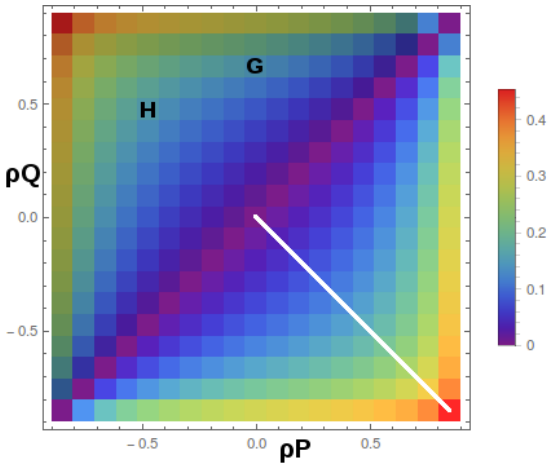
1 par pas de 0.1 puis de 1 à 10 par pas de 1. Nous observons une évolution strictement croissante de la mesure de similarité pour une direction définie. Cependant, cette évolution n'est pas proportionnelle aux différences entre les variances des deux lois. Sur la figure 3.7b, nous percevons une cassure lors du changement de pas, cette rupture ne se perçoit pas avec la divergence de KL (cela s'explique par l'emploi du $\log$ dans la formulation de Kullback-Leibler). Pour la divergence, la différence entre nos deux lois est faible (entre 0.5 et 2), puis augmente rapidement. Pour la distance $\mathcal{L}_2$, dès que l'un des coefficients de la seconde variance est supérieure à 1, la distance varie entre 0 et 0.5, alors que pour la loi normale de variance $\sigma_P = (0.1, 0.1)$ la distance $\mathcal{L}_2$ vaut 2.5. Puisqu'il s'agit de pdf, il y a une normalisation qui augmente le pic des lois de variance inférieure à 1 ce qui explique la disparité des mesures pour la distance Euclidienne.

3.3.1.3 Réponse à un changement d'orientation

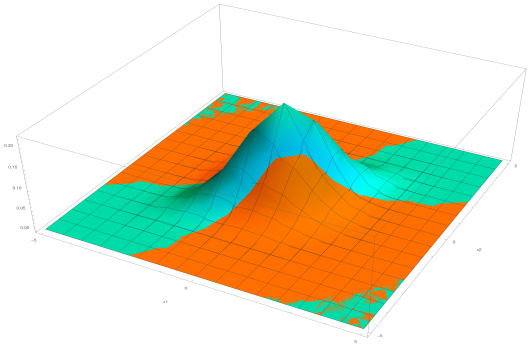
Dans le troisième test, nous faisons varier les covariances ρ_q et ρ_p des deux lois, soit la corrélation. Ce second paramètre a pour impact de transformer les iso-niveaux dans la densité de probabilité de cercle en ellipses orientées selon la covariance (fig. 3.8c). Les cartes de divergence (fig. 3.8a) et de



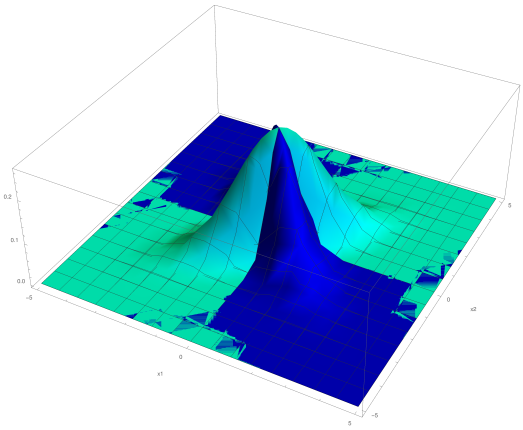
(a) Évolution de la divergence de KL selon ρ_P et ρ_Q



(b) Évolution de la norme $\mathcal{L}_2$selon ρ_q et ρ_p



(c) $\boldsymbol{\rho_P} = 0, \boldsymbol{\rho_Q} = 0.7$:
$D_{KL}(G) = 0.961$,
$D_{\mathcal{L}_2}(G) = 0.145$.



(d) $\boldsymbol{\rho_P} = -0.5, \boldsymbol{\rho_Q} = 0.5$:
$D_{KL}(H) = 1.333$,
$D_{\mathcal{L}_2}(H) = 0.157$.

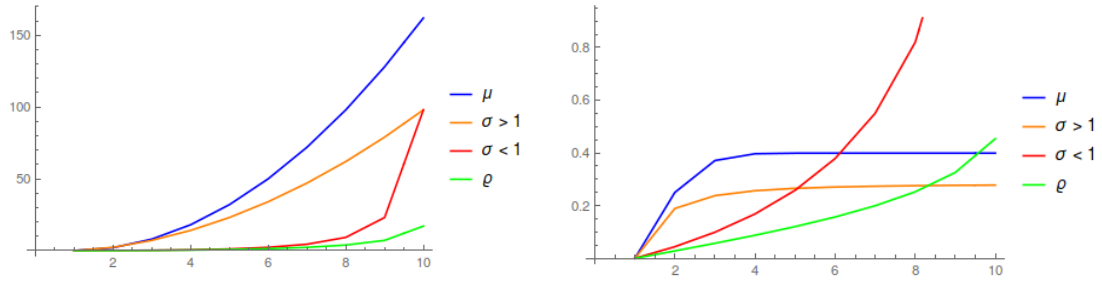
FIGURE 3.8 – **Réponse de la divergence de KL à une variation de covariance** ρ. L'axe $\rho_P = \rho_Q$ est l'axe de mesure de divergence nulle, la divergence croit symétriquement au delà de cet axe.

distance (fig. 3.8b) sont différentes des précédentes. En effet, comme les deux distributions varient, nous avons une droite nulle sur l'axe $\rho_p = \rho_q$ ce qui correspond à l'axe où les deux distributions sont identiques. En dehors de cet axe, la divergence augmente de façon strictement croissante. La figure 3.8c représente les densités de probabilité lorsque P présente une covariance nulle et $\rho_Q = 0.7$. La figure 3.8d représente les densités de probabilité lorsque $\rho_P = -0.5$ et $\rho_Q = 0.5$. La distance $\mathcal{L}_2$ de même est nulle sur la diagonale et semble augmenter plus rapidement avec l'éloignement à la droite identité. Il faut cependant faire attention à l'échelle puisque celle de la divergence varie de 0 à 15 alors que la distance varie de 0 à 0.4.

Avant de conclure sur cette analyse théorique avec des Gaussiennes $2D$, reprenons chacune de ces variations pour nous intéresser à leurs échelles respectives.

3.3.1.4 Comparaison d'échelle

Il faut noter la différence d'échelle entre les trois figures 3.6a, 3.7a et 3.8a et leurs figures équivalentes pour la distance $\mathcal{L}_2$. Nous avons choisi une droite tracée en blanc sur chacune des figures. Les valeurs de ces droites sont reportées sur une figure pour comparer les échelles en fonctions du paramètres variant. La figure 3.9a compare les différences d'échelles pour la divergence de KL, la



(a) Comparaison des échelles pour la divergence de KL (b) Comparaison des échelles pour la norme $\mathcal{L}_2$

FIGURE 3.9 – **Comparaison d'échelle des réponses pour les différentes mesures**. Les droites bleues sont extraites des figures 3.6a et 3.6b, les oranges et rouges des figures 3.7a et 3.7b et les vertes de 3.8a et 3.8b suivant l'axe noté en blanc sur les figures.

figure 3.9b celles de la distance Euclidienne. En bleu, la variation de moyenne, en vert la variation d'orientation. Nous avons deux courbes pour les écarts sur la variance selon que celle-ci diminue (orange) ou augmente (rouge) à partir de 1.

Nous souhaitons que le changement de moyenne soit plus impactant qu'un écart de variance lui-même plus fort qu'une variation d'orientation. En identifiant les variations de moyenne, d'écart-type ou d'orientation à des changements de couleur, l'effet psychovisuel attendu suit cette logique. Un changement de moyenne implique un changement de couleur qui sera plus perceptible qu'un autre. Une différence de variance implique une modification de la disparité des couleurs alors qu'une différence d'orientation maintient la couleur moyenne et la disparité, le changement sera donc moins visible par l'humain (variations sur des statistiques d'ordre supérieur à deux).

Nous retrouvons cet ordre sur les différences d'échelle de la divergence de KL. Modifier la moyenne a un effet prédondérant sur un changement des variances puisque les fortes valeurs des deux distributions ne sont pas confondues. De même, la modification de variance induit une divergence plus forte qu'un changement de covariance pour des variances fixes car la zone de valeurs similaires est plus grande dans le second cas. Bien que les écarts sur les variances ne soient pas identiques pour les variances inférieures ou supérieures à 1, pour une même différence logarithmique, nous atteignons le même seuil dans le cas de la divergence de KL.

La distance $\mathcal{L}_2$ présente une saturation pour les écarts de moyenne et de variance supérieure à 1. Au contraire, pour les deux autres courbes, il ne semble pas y avoir de saturation. Cela implique qu'à moyenne et variance équivalente, il est possible d'obtenir des distances Euclidiennes plus grandes que pour deux Gaussiennes de moyennes très éloignées. Cela ne convient pas à notre attente pour une réponse couleur.

Nous venons de vérifier que les propriétés attendues (proportionnalité de la réponse aux variations des paramètres) sont respectées pour la divergence de Kullback-Leibler. Ce n'est pas le cas de la distance Euclidienne. Il y a une valeur limite pour un changement de moyenne ou pour une variance supérieure à 1. Dans la section suivante, nous allons analyser le comportement de cette mesure sur des images de texture couleur.

3.3.2 Distributions d'images couleurs

Après une première partie théorique sur les mesures de similarité entre distributions, nous étudions la réponse de la fonction de divergence à la mesure de similarité entre textures couleur issues de scènes réelles. Leurs distributions sont donc de plus grande complexité. Ce test porte sur des images de la base *VisTex*. Nous mesurons la différence entre deux images dans l'espace $CIELAB$ qui est perceptuellement uniforme et dans l'espace RGB très utilisé dans la littérature.

Nous considérons l'ensemble des points couleur d'une image comme une distribution. Nous l'approchons par la loi normale obtenue en calculant la moyenne et la matrice de variance covariance associée à la distribution des points couleur. Nous mesurerons donc la divergence de KL entre ces modélisations avec la simplification d'écriture de l'équation 3.16. Nous comparons ces résultats à deux distances $\mathcal{L}_2$.

Tout d'abord, nous utilisons celle associée aux fonctions intégrables (eq. (3.4)). Elles est appliquées aux modélisations Gaussiennes des distributions. Nous nommons cette distance comme précédemment : $D_{\mathcal{L}_2}$. Nous avons choisi de la présenter car c'est l'écriture exacte de la distance $\mathcal{L}_2$ associée aux hypothèses que nous avons faites (espace des probabilités et modélisation par une loi normale). Or la distance de Minkowski d'ordre 2 est majoritairement utilisée pour calculer la similarité entre descripteurs.

La seconde nécessite de calculer un histogramme $3D$. Ce dernier est échantillonné à 10 classes/bins par dimension de l'image couleur. En considérant l'histogramme obtenu comme un vecteur de $\mathbb{R}^{1000}$, nous pouvons utiliser la distance Euclidienne (eq. (3.2)) pour mesurer la différence entre les deux images. Cette distance sera notée D_E. Nous avons choisi empiriquement 10 classes par dimension pour obtenir un histogramme de taille raisonnable tout en gardant la sélectivité nécessaire pour représenter l'image.

3.3.2.1 Cas de textures proches

Dans une première expérimentation, nous choisissons des textures similaires en contenu avec des variations de couleur, d'échelle de prise de vue ou de type (fig. 3.10a,3.10g, 3.10d et 3.10j). Le calcul de la mesure de similarité s'établit selon l'équation (3.16) à partir de l'approximation des distributions de chaque image par une loi multinormale. Celle-ci est représentée par l'ellipsoïde tridimensionnelle, dessinée en rouge sur les figures, englobant 99.7% de la loi de probabilité (3σ).

Les figures 3.10b, 3.10h, 3.10e et 3.10k présentent les histogrammes dans l'espace RGB, leurs équivalents dans l'espace $CIELAB$ sont représentés sur les figures 3.10c, 3.10i, 3.10f et 3.10l. La table 3.1 présente les résultats de différences ainsi que le rapport entre ces différences pour comparer les dynamiques relatives entre les mesures. Nous observons d'abord la similarité qui transparaît visuellement entre les trois premiers histogrammes dans leurs espaces respectifs. C'est donc sans surprise que nous constatons que les mesures de divergence de KL rapportées dans chacun des espaces présentent des variations relatives comparables (tab. 3.1 colonnes *Rapports*). Pour la distance entre histogrammes sous forme de vecteur (D_E), l'ordre de grandeur est aussi similaire. Si, pour la distance $\mathcal{L}_2$ entre les modélisations, les variations relatives sont comparables, les ordres de grandeur ne le sont pas (tab. 3.1 colonnes *Distances*). En effet, l'espace RGB autorise des variations entre 0 et 1 alors que l'espace $CIELAB$ est défini sur $[0;100] \times [-100;100] \times [-100;100]$. Cela induit un changement notable sur les échelles de variance et nous avons vu dans la sous-section précédente qu'une variance inférieure à 1 ou très supérieure à 1 impliquait une réponse très différente de la part de la distance $D_{\mathcal{L}_2}$ en fonction de l'espace utilisé. Par contre, les variations relatives sur la distance D_E ne sont pas équivalentes d'un espace à l'autre.

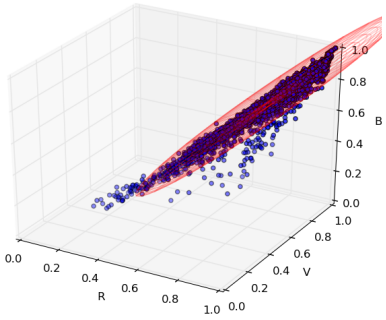
Les valeurs obtenues dans les deux espaces pour la divergence de KL permettent sans aucun doute la différenciation entre le changement d'échelle, de couleur ou d'espèce. Nous ne pouvons en dire autant pour les deux autres distances.

3.3.2.2 Cas de textures différentes

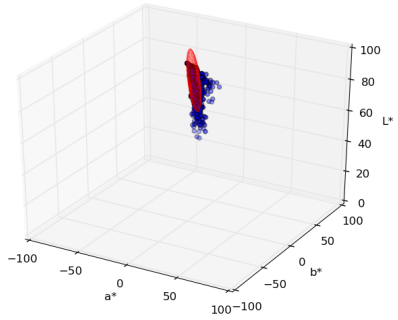
Le second test porte sur une série d'images de distributions couleur très différentes (fig. 3.11 et 3.12). La première figure présente des sensations colorées similaires, alors que la seconde introduit des images plus variées. La différence couleur entre les moyennes des distributions pour les deux



(a) Image de référence



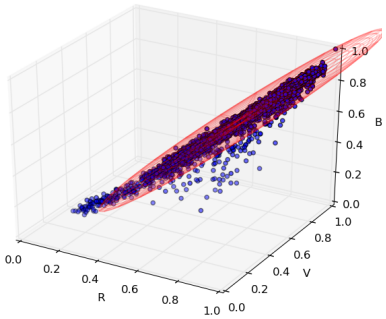
(b) Histogramme dans RGB



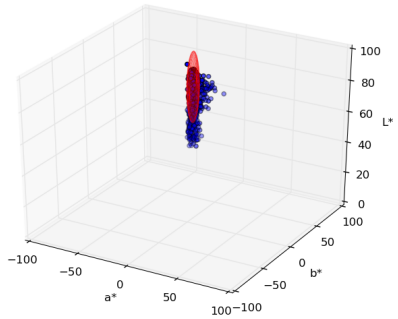
(c) Histogramme dans $CIELAB$



(d) Changement d'échelle



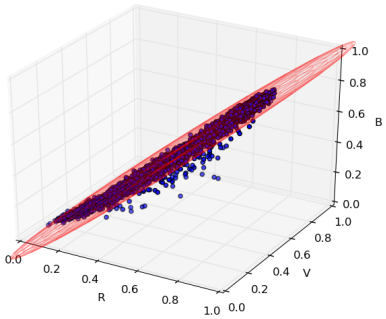
(e) $D_{KL}(3.10a, 3.10d) = 1.175$
$D_{\mathcal{L}_2}(3.10a, 3.10d) = 9.309$
$D_E(3.10a, 3.10d) = 0.163$



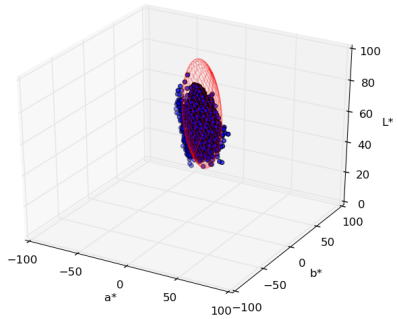
(f) $D_{KL}(3.10a, 3.10d) = 1.170$
$D_{\mathcal{L}_2}(3.10a, 3.10d) = 0.006$
$D_E(3.10a, 3.10d) = 0.162$



(g) Changement de teinte



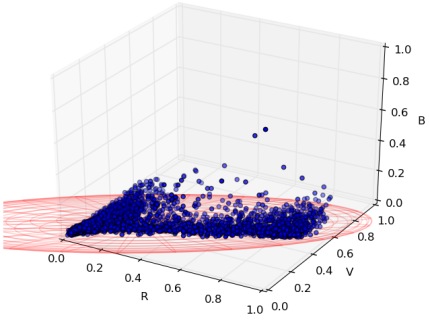
(h) $D_{KL}(3.10a, 3.10g) = 45.391$
$D_{\mathcal{L}_2}(3.10a, 3.10g) = 16.635$
$D_E(3.10a, 3.10g) = 0.364$



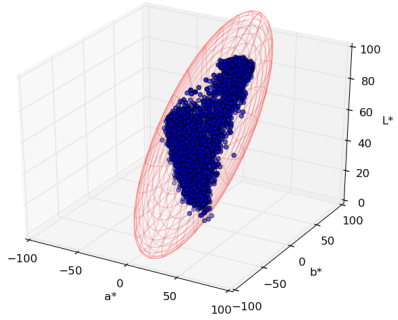
(i) $D_{KL}(3.10a, 3.10g) = 45.226$
$D_{\mathcal{L}_2}(3.10a, 3.10g) = 0.012$
$D_E(3.10a, 3.10g) = 0.524$



(j) Changement de fleur



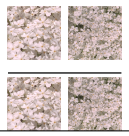
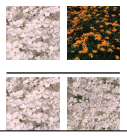
(k) $D_{KL}(3.10a, 3.10j) = 191.914$
$D_{\mathcal{L}_2}(3.10a, 3.10j) = 15.172$
$D_E(3.10a, 3.10j) = 0.487$



(l) $D_{KL}(3.10a, 3.10j) = 217.291$
$D_{\mathcal{L}_2}(3.10a, 3.10j) = 0.011$
$D_E(3.10a, 3.10j) = 0.530$

FIGURE 3.10 – **Quatre textures de même nature avec leurs histogrammes** et l'ellipsoïde correspondant au modèle multinormal associé dans RGB et $CIELAB$. L'image 3.10a correspond à la référence pour les mesures de distances indiquées sous les histogrammes.

TABLE 3.1 – **Table des mesures entre textures proches** et rapports entre ces mesures (dynamiques relatives) pour chacune des distances et divergence calculées.

		Distances			Rapports		
							
D_{KL}	RGB	1.175	45.391	191.914	1	38.6	163.3
	$CIELAB$	1.170	45.226	217.291	1	38.7	185.7
$D_{\mathcal{L}_2}$	RGB	9.309	16.635	15.172	1	1.8	1.6
	$CIELAB$	0.006	0.012	0.011	1	2.0	1.8
D_E	RGB	0.163	0.364	0.487	1	2.3	3.0
	$CIELAB$	0.162	0.524	0.530	1	3.2	3.3

espaces couleur est fournie table 3.2. L'idée est d'établir un ordonnancement des images à partir de cette similarité entre leurs moyennes. Cela justifie la séparation en deux cas : images de variances faibles (fig. 3.11) et fortes (fig. 3.12). L'ordre est très différent dans les deux espaces. Or les axes de représentation couleur de l'espace RGB ne sont pas adaptés puisque non orthogonaux, comme nous l'avons vu dans le chapitre 1. Effectuer une moyenne dans cet espace n'a donc pas de sens. Nous choisissons donc de classer les images par proximité croissante à la moyenne de l'image de référence (*Water* : fig. 3.11a et 3.12a) dans l'espace $CIELAB$. Nous utilisons donc la différence couleur ΔE comme référence. Compte-tenu des dynamiques des différentes tables (entre 0 et plus de 1000), nous avons volontairement réduits les mesures à un chiffre après la virgule.

TABLE 3.2 – **Table des distances entre images à partir de la différence couleur entre leurs moyennes**.

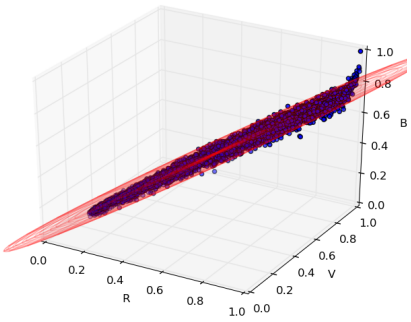
		Différence entre couleurs moyennes					
							
RGB	*Water*	0	0.1	0.4	0.2	0.1	0.3
	Stone		0	0.3	0.1	0.1	0.3
	Paintings			0	0.2	0.4	0.3
	Food				0	0.2	0.2
	Water'					0	0.4
	Fabric						0
$CIELAB$	*Water*	0	6.7	25.5	15.1	22.9	48.5
	Stone		0	20.0	14.9	23.0	51.0
	Paintings			0	18.6	38.3	49.9
	Food				0	36.3	38.8
	Water'					0	64.4
	Fabric						0

L'image de référence (fig. 3.11a) est plutôt achromatique (voir histogrammes dans RGB et $CIELAB$, fig. 3.11b et 3.11c). Il en est de même pour la seconde image figure 3.11d. La similarité attendue est donc très faible. L'image suivante présente un contenu plus chromatique, ses histogrammes se distribuent autour de l'axe achromatique de façon plus diffuse.

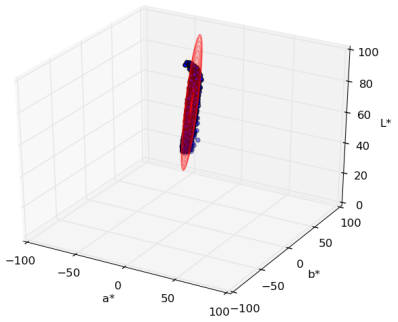
L'image *Food* (fig. 3.12d), comme l'image *Paintings* (fig. 3.11g), est centrée sur l'axe achromatique mais se distribue sur tout le spectre couleur. Les histogrammes des images *Fabric* (fig. 3.12j) et *Water'* (fig. 3.12d) se positionnent sur un axe décalé de l'axe achromatique, de forme très différente de celle de la référence pour l'image *Fabric*.



(a) Image de référence



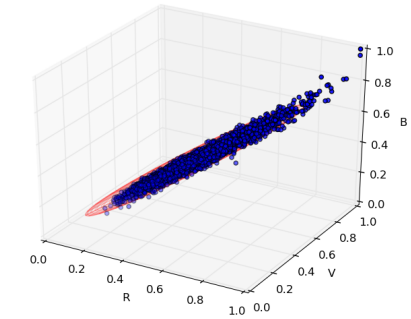
(b) Histogramme dans RGB



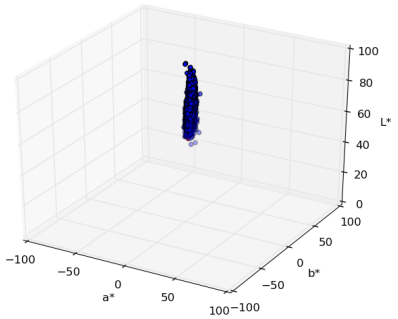
(c) Histogramme dans $CIELAB$



(d) *Stone*



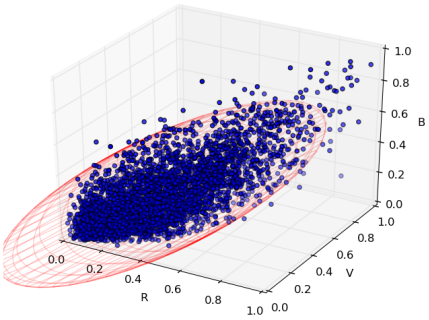
(e) $D_{KL}(3.11a, 3.11d) = 4.7$
$D_{\mathcal{L}_2}(3.11a, 3.11d) = 28.7$
$D_E(3.11a, 3.11d) = 0.2$



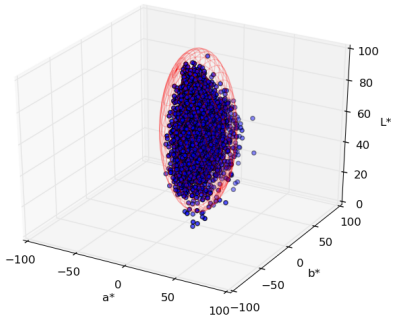
(f) $D_{KL}(3.11a, 3.11d) = 4.1$
$D_{\mathcal{L}_2}(3.11a, 3.11d) = 0.0$
$D_E(3.11a, 3.11d) = 0.3$



(g) *Paintings*



(h) $D_{KL}(3.11a, 3.11g) = 22.3$
$D_{\mathcal{L}_2}(3.11a, 3.11g) = 14.9$
$D_E(3.11a, 3.11g) = 0.2$

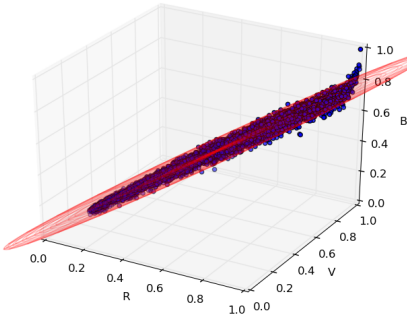


(i) $D_{KL}(3.11a, 3.11g) = 29.3$
$D_{\mathcal{L}_2}(3.11a, 3.11g) = 0.0$
$D_E(3.11a, 3.11g) = 0.3$

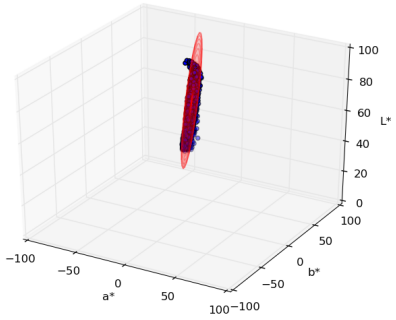
FIGURE 3.11 – **Trois textures neutres et leurs histogrammes** dans les espaces RGB et $CIELAB$. L'image 3.11a correspond à la référence pour les mesures de distances indiquées sous les histogrammes.



(a) Image de référence



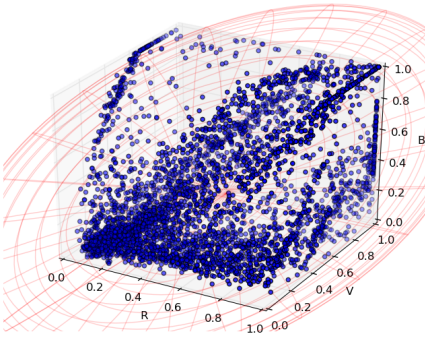
(b) Histogramme dans RGB



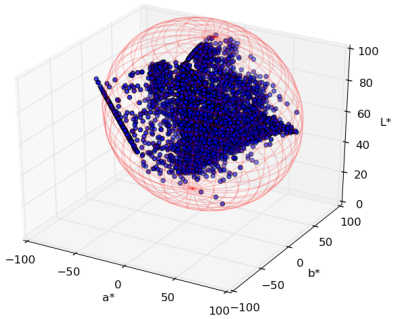
(c) Histogramme dans $CIELAB$



(d) *Food*



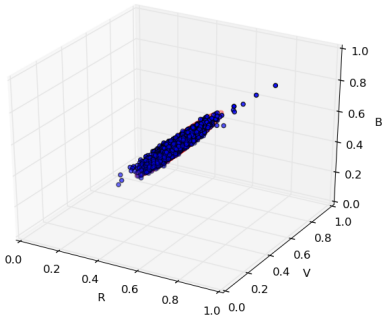
(e) $D_{KL}(3.12a, 3.12d) = 167.0$
$D_{\mathcal{L}_2}(3.12a, 3.12d) = 15.3$
$D_E(3.12a, 3.12d) = 0.2$



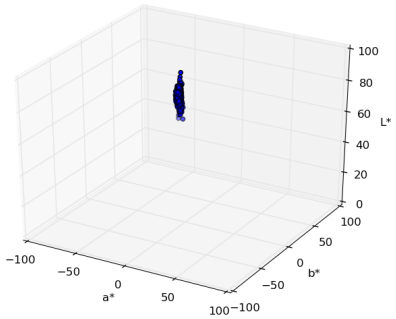
(f) $D_{KL}(3.12a, 3.12d) = 130.5$
$D_{\mathcal{L}_2}(3.12a, 3.12d) = 0.0$
$D_E(3.12a, 3.12d) = 0.2$



(g) *Water'*



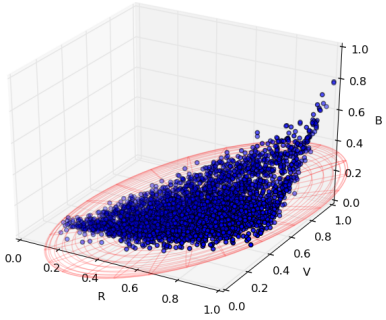
(h) $D_{KL}(3.12a, 3.12g) = 179.9$
$D_{\mathcal{L}_2}(3.12a, 3.12g) = 48.7$
$D_E(3.12a, 3.12g) = 0.5$



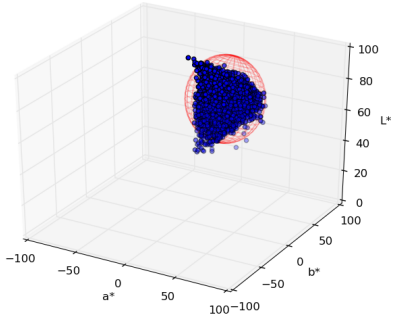
(i) $D_{KL}(3.12a, 3.12g) = 170.2$
$D_{\mathcal{L}_2}(3.12a, 3.12g) = 0.0$
$D_E(3.12a, 3.12g) = 0.7$



(j) *Fabric*



(k) $D_{KL}(3.12a, 3.12j) = 125.1$
$D_{\mathcal{L}_2}(3.12a, 3.12j) = 16.6$
$D_E(3.12a, 3.12j) = 0.2$

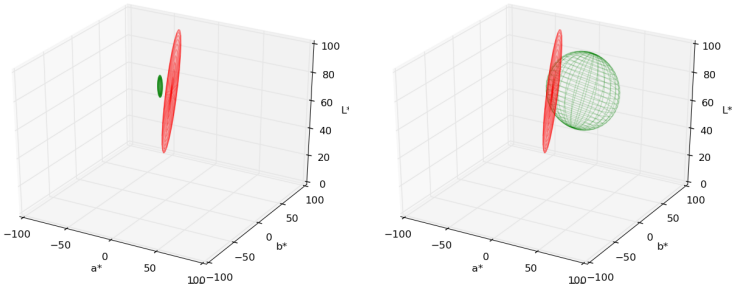


(l) $D_{KL}(3.12a, 3.12j) = 148.3$
$D_{\mathcal{L}_2}(3.12a, 3.12j) = 0.0$
$D_E(3.12a, 3.12j) = 0.4$

FIGURE 3.12 – **Trois textures colorées et leurs histogrammes** dans les espaces RGB et $CIELAB$. L'image 3.12a correspond à la référence pour les mesures de distances indiquées sous les histogrammes.

L'image *Food* (fig. 3.12d) a une moyenne plus proche de l'image de référence *Water* (fig. 3.12a) que ne l'est celle de l'image *Fabric* (fig. 3.12j). La mesure de différence associée est donc plus faible pour l'image *Food* que celle de *Fabric*. Nous constatons l'inverse dans l'espace RGB.

On s'attendrait à obtenir une mesure plus faible pour l'image *Water'* (fig. 3.12g) que pour *Fabric* (fig. 3.12j) puisqu'elle représente le même objet que l'image de référence. Or sur le plan purement chromatique, les ellipsoïdes caractéristiques des distributions des images *Water'* et *Water* (fig. 3.13) sont disjointes, induisant une différence plus importante que ne le laisserait penser l'interprétation sémantique.



(a) *Water* et *Water'* (b) *Water* et *Paintings*

FIGURE 3.13 – **Comparaison de chevauchement de distributions**. Les ellipses des figures 3.12g et 3.12j (en vert) avec celle de l'image de référence (Figure 3.11a en rouge). Les représentations sont dans l'espace $CIELAB$.

TABLE 3.3 – **Divergence de KL entre les six images**.

		D_{KL}					
							
RGB	*Water*	0	4.4	22.3	167.0	179.9	125.1
	Stone		0	77.0	659.8	231.9	180.4
	Paintings			0	11.9	179.9	15.4
	Food				0	822.2	33.1
	Water'					0	1096.8
	Fabric						0
$CIELAB$	*Water*	0	4.1	29.3	130.5	170.2	148.3
	Stone		0	89.9	433.5	236.5	176.9
	Paintings			0	6.1	242.1	21.4
	Food				0	727.5	43.3
	Water'					0	1316.1
	Fabric						0

La table 3.3 présente les mesures de divergence obtenues pour les six images considérées, dans l'espace $CIELAB$ et RGB. Comme attendu les distributions organisées selon l'axe achromatique sont les plus proches de la référence. L'image *Food* est naturellement plus proche de l'image *Paintings* et *Fabric*, de par sa complexité chromatique, que de l'image *Stone* ou *Water'*. Les images *Food* et *Paintings* ont des moyennes similaires, leur différence étant liée plutôt à leurs matrices de covariance. La moyenne de l'image *Fabric* diffère des deux précédentes. Un écart de moyenne entraîne une divergence plus grande qu'un écart de covariance. Cette différence est mieux représentée dans les mesures effectuées dans l'espace $CIELAB$ que dans l'espace RGB.

La table 3.4 récapitule les mesures $D_{\mathcal{L}_2}$ entre les six images pour les deux espaces couleurs. Nous observons tout d'abord qu'à un chiffre significatif après la virgule, les distances mesurées dans l'es-

TABLE 3.4 – **Distance $\mathcal{L}_2$ entre les six images**.

		$D_{\mathcal{L}_2}$					
							
RGB	*Water*	0	28.7	14.9	15.3	48.7	16.6
	Stone		0	35.6	36.0	58.6	37.6
	Paintings			0	3.9	46.3	7.2
	Food				0	46.1	6.0
	Water'					0	46.7
	Fabric						0
$CIELAB$	*Water*	0	0.0	0.0	0.0	0.0	0.0
	Stone		0	0.0	0.0	0.0	0.0
	Paintings			0	0.0	0.0	0.0
	Food				0	0.0	0.0
	Water'					0	0.0
	Fabric						0

pace $CIELAB$ sont toutes nulles, ce qui n'est pas le cas du RGB. Comme expliqué précédemment (fig. 3.7b) cela vient de l'échelle des variances qui varie d'un espace à l'autre. Pour effectuer ce calcul, nous avons utilisé la pdf associée à la loi normale approximant nos données. La normalisation de la Gaussienne utilisée est donc automatique. Les valeurs obtenues sans cette normalisation seraient très différentes, mais alors nous ne rentrerions plus dans le cas des mesures entre probabilités.

En regardant les résultats dans l'espace RGB, la première chose à noter est l'inversion d'ordre entre les images *Stone* et *Paintings*. En comparant leurs histogrammes à ceux de l'image *Water* (fig. 3.11b et 3.11c), ceux de l'image *Stone* (fig. 3.11e et 3.11f) semblent plus proches de la référence que ceux de *Paintings* (fig. 3.11h et 3.11i). Pourtant $D_{\mathcal{L}_2}$ est supérieure entre *Water* et *Stone* qu'entre *Water* et *Paintings*.

Dans la table 3.2 les ΔE entre la moyenne de l'image *Stone* et les autres varient de 6.7 à 51.0, or les distances $\mathcal{L}_2$ calculées ne montrent pas autant de variation. Il en est de même pour les distances à l'image *Water'*. Ces deux images ont de petites valeurs de variance. L'écart avec les autres se fait majoritairement sur la différence à la moyenne. Or la section 3.1.2.1 a montré la saturation qu'impliquait l'utilisation d'une mesure de Minkowski. Les résultats obtenus pour ces deux images illustrent cette limite généralement oubliée.

Les résultats de la distance Euclidienne entre les histogrammes couleur sont présentés table 3.5. Les différences entre l'image *Water* et les images *Stone*, *Paintings* et *Food* sont indiscernables dans les deux espaces pour cette mesure. Les histogrammes de l'image *Food* (fig. 3.12e et 3.12f) suggèrent pourtant des histogrammes à 10 classes différents des histogrammes de l'image *Stone* (fig. 3.11e et 3.11f). La perte de précision due à la quantification obscurcit les résultats et implique une incapacité à séparer les images.

Il faut rappeler que ces mesures ont été effectuées après une quantification de 10 classes par canal. Un autre choix de nombre de classes implique des résultats différents. Se pose alors la question du nombre de classes à utiliser et de la pertinence d'une telle mesure si sujette aux paramètres d'entrée…

3.3.2.3 Test de normalité

En modélisant nos images par des lois normales, nous faisons une hypothèse forte sur l'allure des données. Dans ce paragraphe nous éprouvons cette hypothèse à l'aide de test de normalité.

TABLE 3.5 – **Distance D_E entre les six images**.

		D_E 					
RGB	*Water*	0	0.2	0.2	0.2	0.5	0.3
	Stone		0	0.3	0.3	0.6	0.3
	Paintings			0	0.1	0.5	0.2
	Food				0	0.5	0.2
	Water'					0	0.5
	Fabric						0
$CIELAB$	*Water*	0	0.3	0.3	0.3	0.7	0.4
	Stone		0	0.5	0.5	0.8	0.5
	Paintings			0	0.1	0.7	0.3
	Food				0	0.7	0.2
	Water'					0	0.7
	Fabric						0

Nous avons supposé que les données d'entrée suivent une loi normale. Un test de normalité permet d'établir la probabilité que l'hypothèse effectuée soit cohérente. Ces tests permettent de rejeter l'hypothèse de normalité mais pas de confirmer que celle-ci est valide. Il existe différents types de tests.

Parmi ceux-ci nous trouvons les tests graphiques qui estiment la concordance graphique entre les données et la loi supposée suivie. C'est le cas des ellipsoïdes rouges présentes sur les histogrammes des figures 3.10, 3.11 et 3.12. Chaque ellipse représente l'enveloppe englobante à 3Σ de la loi normale associée. Si certaines ellipses semblent relativement bien coïncider avec sa distribution associée (fig. 3.11e et 3.11f), ce n'est pas le cas de toutes. Les histogrammes associés à *Food* (fig. 3.12e et 3.12f) ou *Paintings* (fig. 3.11h et 3.11i) s'intègrent mal dans l'ellipsoïde associée.

D'autres tests estiment la qualité de l'ajustement du modèle statistique supposé avec les données. C'est le cas des tests du χ^2 de Pearson (1900) et celui de Cramér-von Mises (Cramer (1928)). Le test du χ^2 de Pearson vérifie que l'écart entre les histogrammes des distributions suit une loi du χ^2. Le test de Cramér-von Mises compare la fonction de répartition avec celle des données.

D'autres tests mesurent l'adéquation du kurtosis et du coefficient d'asymétrie avec le modèle testé. C'est le cas des tests Jarque and Bera (1987) et de Mardia (1970). Le test de Mardia a été construit pour des distributions multivaluées. Il existe de nombreux autres tests. Nous n'avons cité que ceux que nous avons utilisé.

Soit l'hypothèse H_0 : les données I suivent une loi normale. Nous calculons la classique p-valeur qui correspond à $P(I|H_0)$ pour une confiance de 95%. Autrement dit, si $p < 0.05$ alors l'hypothèse de normalité est rejetée. Tous les tests de toutes les images dans tous les espaces rejettent l'hypothèse de normalité. Pour l'image *Stone* dont l'ellipsoïde semblait correspondre à l'enveloppe des données, nous obtenons les résultats donnés table 3.6. La table présente aussi les résultats de l'image *Food* dont la matrice de variance-covariance est très différente de celle de l'image *Stone*. Toutes les valeurs sont effectivement inférieures à 0.05. Les autres images présentent des résultats similaires. Nos données sont composées de 512×512 échantillons. On ne peut donc supposer que c'est le manque d'échantillons qui implique l'échec aux tests.

Nous prenons une image couleur représentant une texture naturelle ou manufacturée. Nous abandonnons les informations spatiales de cette image pour établir une distribution statistique que nous supposons normale. Cependant, les textures analysées ne sont-elles pas déterministes ? Cela explique le rejet des tests de normalité.

Cet échec ne remet pas en cause les résultats de la divergence de KL mise en avant dans ce cha-

TABLE 3.6 – **Tests d'hypothèse de normalité**.

		RGB	$CIELAB$
	χ^2 de Pearson	2.3×10^{-333261}	$8.5 \times 10^{-14929517}$
	Cramér-von Mises	3.0×10^{-43}	9.3×10^{-38}
	Jarque-Bera	7.1×10^{-10621}	1.6×10^{-22195}
	Mardia kurtosis	3.3×10^{-9587}	2.2×10^{-300238}
	Mardia asymétrie	1.8×10^{-12636}	1.8×10^{-147035}
	χ^2 de Pearson	3.9×10^{-244048}	$1.8 \times 10^{-23249170}$
	Cramér-von Mises	3.7×10^{-41}	3.0×10^{-38}
	Jarque-Bera	1.6×10^{-12649}	1.6×10^{-22195}
	Mardia kurtosis	4.2×10^{-22}	2.2×10^{-300238}
	Mardia asymétrie	6.4×10^{-26560}	1.8×10^{-147035}

pitre. En effet, nous avons fait cette hypothèse pour améliorer le temps de calcul et éviter le problème de probabilité nulle. Il est toujours possible de calculer de manière directe la divergence de KL entre deux distributions. Ces résultats induisent qu'un modèle plus précis (mélange de Gaussiennes par exemple) associé avec une métrique de distance ou similarité adaptée ne pourra encore qu'améliorer les résultats obtenus ; d'autant que ceux-ci sont déjà meilleurs que ceux existants et de plus théoriquement valides.

3.4 Synthèse

Afin de compléter notre association détecteur/descripteur, nous avons abordé dans ce chapitre la question de mesure de similarité. Dans une première partie, nous avons présenté les principales mesures utilisées dans l'état de l'art et expliqué pourquoi elles n'étaient pas adaptées, notamment la distance $\mathcal{L}_2$, dans le cas des densités de probabilités.

Nous avons ensuite proposé d'utiliser la divergence de Kullback-Leibler. Même s'il s'agit seulement d'une divergence, puisqu'elle ne respecte pas l'inégalité triangulaire. En revanche, elle permet, grâce à sa forme, d'intégrer la différence d'entropie entre deux densités de probabilités, d'où son intérêt. Afin de simplifier les calculs, nous proposons d'approcher les distributions nD à estimer par des lois normales nD, notant quand même l'importance d'effectuer un test de normalité sur les distributions utilisées.

Premièrement, nous avons validé le concept avec des distributions Gaussiennes $2D$ de synthèse. L'écart de moyenne correspond à un écart de teinte sur des images couleurs. Un écart de variance-covariance correspond à une différence d'étalement couleur (par exemple, une aplat vert et la photo d'une canopée). L'éloignement des moyennes induit une augmentation dans les mesures de différence plus importante.

Preuve de concept théorique réalisée, nous avons testé cette idée sur l'estimation de la similarité entre images couleur à partir de leur distribution normale associée. Nous avons montré que les résultats sont cohérents et que cette mesure de similarité ne sature pas comme d'autres ($\mathcal{L}_2$ typiquement) et permet de distinguer les textures entre elles. Deux limites sont à noter pour cette étude sur le choix d'une mesure de similarité entre pdf dans le contexte des images couleur. Tout d'abord, les tests de normalité ne confirment pas notre hypothèse initiale, mais surtout nous perdons totalement l'information spatiale qui est indispensable pour un attribut de texture.

Nous avons présenté une mesure de similarité entre distributions de couleur basée sur la divergence de Kullback-Leibler établie dans l'espace $CIELAB$. En utilisant une approximation de la distribution couleur par une loi multi-normale, l'estimation est très rapide, tout en préservant la structure vectorielle des distributions dans l'espace couleur. Après avoir validé le comportement de cette

mesure sur des lois normales bidimensionnelles de synthèse, nous avons montré qu'elle était cohérente avec notre perception pour la comparaison de textures naturelles couleur.

Suite à cette étude préliminaire, d'autres tests sont bien évidement à considérer : impact sur la classification de textures, gains liés à l'exploitation du modèle complet de mesure de la divergence (sans l'estimation par une loi normale tridimensionnelle), comparaison étendue à d'autres distances plus simples ou même à une distance de Mahalanobis, basée sur une métrique de Minkowski limitée dans ce contexte. Mais surtout, dans un second temps, il est nécessaire de mener des expériences psychovisuelles afin de valider l'adéquation de la mesure avec la perception.

Nous nous sommes interrogés sur le choix des fonctions de distance/similarité à associer à chaque descripteur. Cette interrogation va de pair avec le choix d'une définition mathématique adaptée à chaque descripteur. Cette définition induit par elle-même un choix de familles de mesures de similarité/distances et un usage non systématique de la $\mathcal{L}_2$.

Nous avons montré dans ce chapitre que la paire $\{$descripteur, distance de similarité$\}$ était indissociable et au cœur des performances. Cela devient désormais notre contrainte pour la création de descripteurs.

Résumé

- Reformulation mathématique de la distance ou mesure de similarité entre descripteurs.
- Création d'une taxonomie des fonctions de distance ou mesure de similarité entre théorie et pratique.
- Mise en place d'un protocole de comparaison des mesures de distance/divergences entre histogramme nD.
- Preuve du bien-fondé d'une approche mathématiquement adaptée dans le cas de la recherche d'images à partir de leurs histogrammes couleurs (base *VisTex*).

Descripteurs de points d'intérêt

Contenu

4.1 Décrire le contenu spatio-chromatique local 126
4.1.1 Valeurs de pixels 130
4.1.2 Décomposition fréquentielle 130
4.1.3 Gradients 131
4.1.4 Différences de valeurs 133
4.1.5 Motifs 134
4.2 Invariance des descripteurs 136
4.2.1 Protocole 136
4.2.1.1 Images et transformations utilisées 136
4.2.1.2 Détecteur, descripteurs et critère de comparaison 139
4.2.2 Résultats 140
4.2.2.1 Invariance à l'échelle 140
4.2.2.2 Invariance à la rotation 141
4.2.2.3 Invariance au cisaillement 141
4.2.2.4 Invariance à la luminosité 141
4.2.2.5 Invariance à la réflexion 143
4.2.2.6 Impact de la taille de la zone d'intérêt 143
4.2.2.7 Influence de l'orientation principale de la zone d'intérêt . . . 145
4.3 Un descripteur couleur déduit des conjectures de Julesz 146
4.3.1 Première conjecture : cooccurrence 146
4.3.2 Deuxième conjecture : ordres supérieurs à deux 148
4.3.3 Troisième conjecture : texton 149
4.3.4 Relocated colour contrast occurrence 150
4.4 Preuve de l'approche 152
4.4.1 Base de données artificielles 153
4.4.1.1 Protocole des expérimentations 154
4.4.1.2 Discrimination selon la couleur moyenne 155
4.4.1.3 Discrimination selon la matrice de covariance 156
4.4.1.4 Discrimination selon le coefficient de Hurst 157
4.4.1.5 Classification sur toutes les images 159
4.4.2 Classification de textures 160
4.4.2.1 Protocole de l'expérimentation 160
4.4.2.2 Performance de classification 162
4.5 Synthèse 164

Nous allons traiter ici des descripteurs associés aux points d'intérêt. La description porte sur la partie d'image entourant le point d'intérêt. Celle-ci peut être plus ou moins étendue en fonction de la nature du point d'intérêt ou de son mode d'extraction (dense ou non).

Comme nous l'avons vu dans le chapitre précédent, un point d'intérêt peut être détecté à une échelle donnée. C'est l'échelle qui définit la taille du voisinage exploité. Plus la taille du filtre de détection est grande, plus la zone d'étude est grande et donc plus le descripteur perd en spécificité liée à sa localisation. Au contraire, une échelle trop petite induit un voisinage trop restreint pour extraire des statistiques suffisamment pertinentes.

La recherche d'images ou de parties d'images similaires à une référence induit le fait que l'objet à retrouver diffère sur plusieurs éléments non-signifiants sémantiquement. Au cœur de ces différences se trouvent les variations de point de vue d'observation ou celles de l'éclairement. Les descripteurs recherchés doivent donc être invariants à certaines transformations.

L'objectif final d'une tâche de reconnaissance/classification est de pouvoir apparier des points ayant des contenus environnants similaires. Cet objectif induit le recours à un descripteur et une distance de similarité associée. Les mesures que nous utiliserons seront issues du chapitre précédent. Ce chapitre se focalise donc sur les descripteurs.

Nous commencerons par présenter l'état de l'art au regard de notre problématique couleur et métrologique. Certains descripteurs seront présentés plus en détail puisqu'ils seront l'objet d'une analyse d'invariance dans la deuxième partie. Cette notion d'invariance est très présente dans l'utilisation des descripteurs et nous avons voulu vérifier leurs propriétés. La troisième partie amènera notre contribution aux descripteurs basés texture et introduira le $\mathrm{RC_2O}$, basé sur la cooccurrence couleur et nécessitant l'utilisation de la divergence de Kullback-Leibler pour une utilisation optimale. La dernière partie vise à valider les propositions, tout d'abord sur une base de données artificielles d'images couleur fractales. Enfin, les résultats de classification obtenus avec les occurrences de différences couleur présentées au chapitre précédent associées à la divergence de Kullback-Liebler (sec. 4.4.2) seront présentés.

4.1 Décrire le contenu spatio-chromatique local

Avant toute chose, il semble important de préciser le vocabulaire que nous allons utiliser. L'ensemble des termes employés appartient autant au vocabulaire courant que technique. D'autant que dans ce dernier cas, selon les disciplines ou même les domaines d'applications, le sens des mots diffère.

Au premier plan, nous avons la **caractéristique** [1] **physique** que nous cherchons à analyser, comparer, indexer, reconnaître… Celle-ci peut être **composite** et de très haut niveau sémantique, en témoigne l'activité en cours sur l'apparence des matériaux (projet ApPREARS : http ://www.appears-itn.eu). L'objectif de l'outil numérique est de traduire sous des formes mathématiques adaptées la caractéristique physique à analyser.

L'aspect composite de la caractéristique induit une décomposition en deux niveaux. Le premier niveau contient les **mesures** des caractéristiques de bas niveau (couleur, forme, fréquence spatiale…). Ce niveau est défini comme produisant des **attributs** [2] d'une région ou d'une image. C'est dans un second temps qu'apparaissent les mesures de l'aspect composite des caractéristiques physiques et visuelles (liées à la perception de celle-ci). Dans la suite, nous qualifierons ce niveau de **des-**

1. Larousse 2011 : ce qui constitue la particularité, le caractère distinctif de quelqu'un ou quelque chose. Encyclopédie Universalis : qui caractérise, qui fournit un caractère distinctif.

2. Larousse 2011 : ce qui appartient, ce qui est inhérent à quelqu'un ou quelque chose. Encyclopédie Universalis : ce qui est propre, caractéristique de.

cripteur[3]. Dans l'état de l'art (Trémeau et al. (2004), ce niveau est associé à la construction de vecteurs, d'ensembles ou de concaténations d'attributs. En termes d'indexation, il correspond à la notion de **signature**. Le terme de descripteur est ici générique et permet d'englober toutes les formes mathématiques d'assemblage d'attributs, sans les limites de l'une ou de l 'autre. Ici encore, nous attachons une grande importance au choix d'une définition mathématique, à laquelle une mesure de similarité efficace puisse être associée.

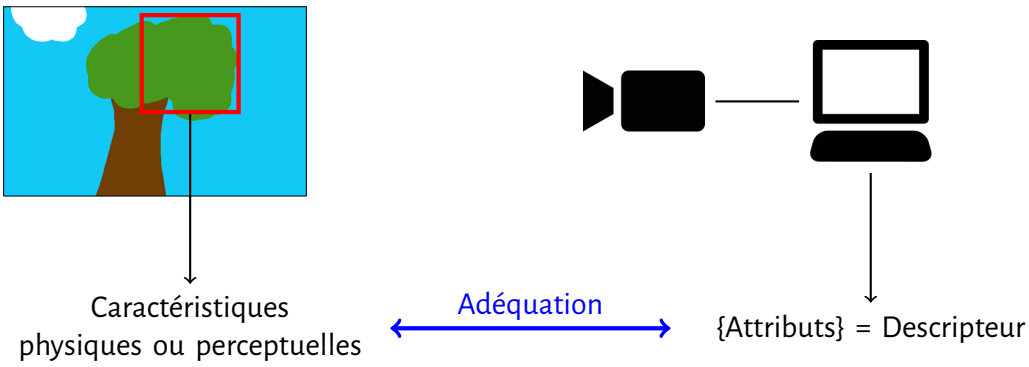


FIGURE 4.1 – **Illustration du positionnement sémantique des termes utilisés**.

Au final, la problématique **implicite** est celle de l'adéquation de la mesure d'une caractéristique physique et/ou de son impact perceptuel avec le descripteur calculé à partir d'une image prise (fig. 4.1). Cette problématique complexe est souvent réduite à la capacité à associer une même sémantique (étiquette, classe…) à la caractéristique et au descripteur.

Au sein de ce chapitre, nous nous intéressons à une caractéristique de l'apparence des matériaux ou surfaces : l'aspect coloré et non-uniforme[4]. La mesure de cette caractéristique induit d'associer dans un même descripteur les aspects liés à la distribution des couleurs de la scène/surface **et** la distribution spatiale de ces couleurs. Plusieurs attributs ont été utilisés/testés à cette fin (fig. 4.2). Ils ont la particularité de tous être basés sur des histogrammes. Cependant, les auteurs se posent la question

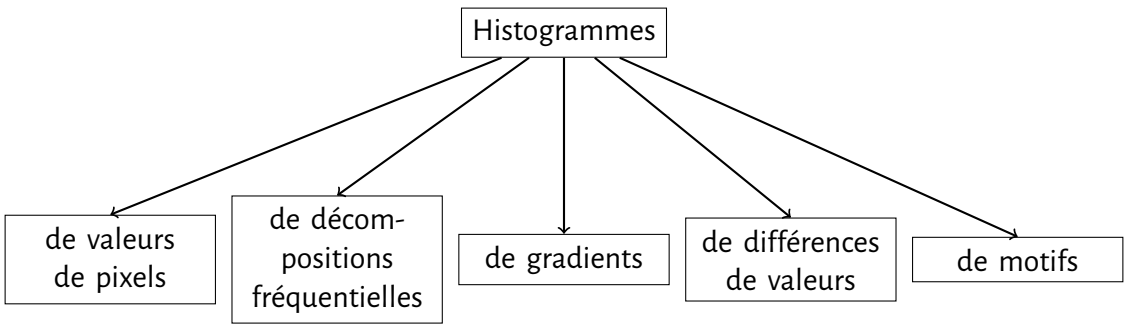


FIGURE 4.2 – **Organisation des attributs pour la caractérisation de l'aspect non-uniforme**.

de l'évaluation de la mesure de similarité entre ces histogrammes. Dès lors une première dichotomie apparaît dans l'état de l'art (fig. 4.3) suivant que la mesure ait été effectuée directement ou après une phase de modélisation de l'histogramme (suivant une loi normale, un mélange de Gaussiennes…). Au delà du gain mémoire, une partie de la perte de précision de la discrimination se joue à ce niveau.

Naturellement, la non-uniformité a été évaluée d'abord sous l'angle de la disparité des valeurs de pixels (fig. 4.2 : valeurs de pixels). Lazebnik et al. (2003) utilisent un histogramme $2D$ stockant

3. Larousse 2011 : ensemble de signes donnant une description d'un fichier, d'une variable, etc. Encyclopédie Universalis : ensemble de caractéristiques d'un fichier.

4. **meeting CIE div. 8, Prague, sept 2016** : le terme texture est trop générique et peut prêter à confusion entre les différentes communautés. Il est demandé de le remplacer par le terme **aspect non-uniforme**.

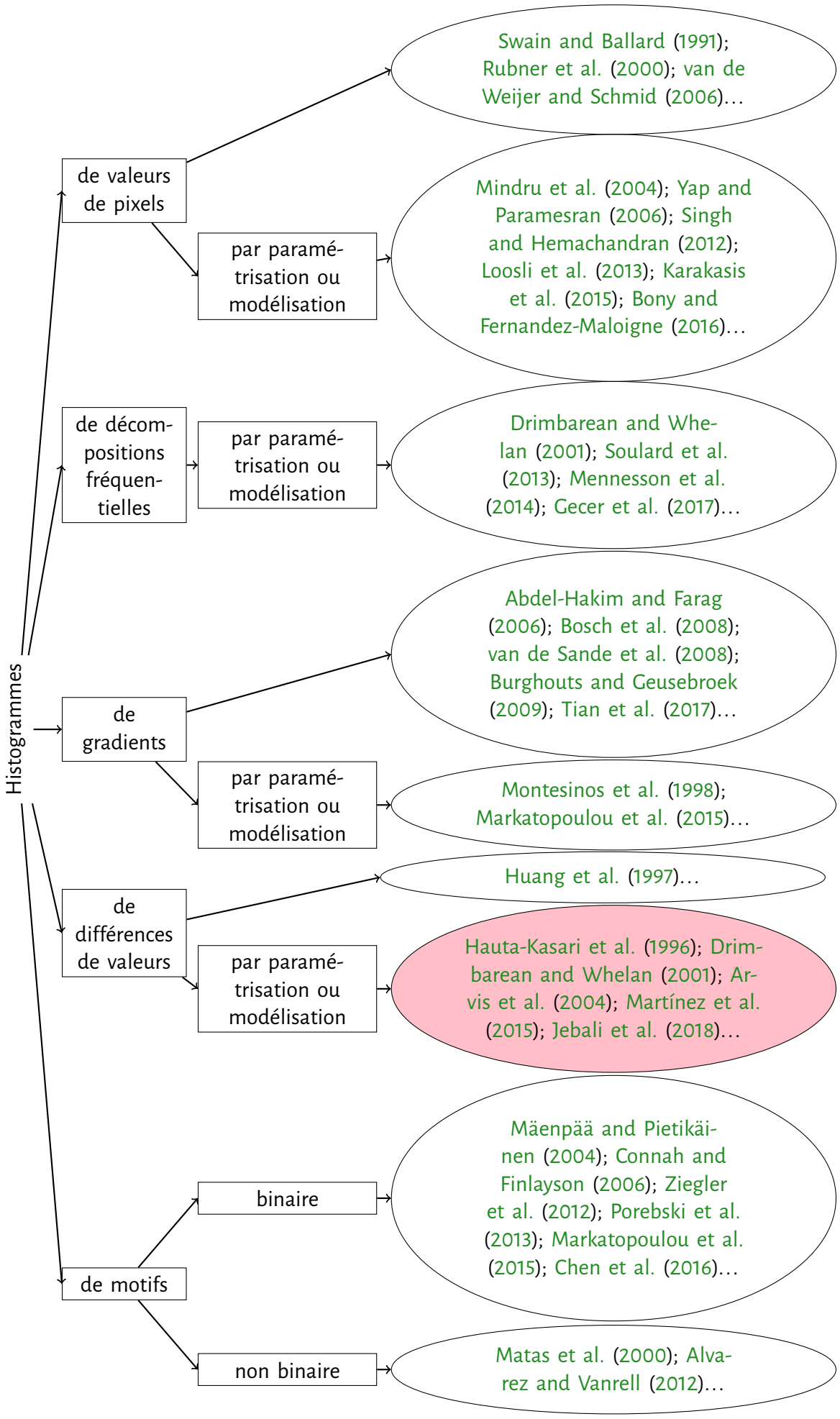


FIGURE 4.3 – **État de l'art des descripteurs couleur** organisé en fonction du type des attributs du descripteur.

l'information intensité et distance au centre du point d'intérêt. Les premières adaptations couleurs ont été faites dans cette famille comme Swain and Ballard (1991). Néanmoins cette famille d'attributs n'inclut pas la caractérisation de l'organisation spatiale des valeurs.

En terme de traitement du signal, l'aspect non-uniforme s'évalue naturellement au travers des représentations fréquentielles (fig. 4.2 : décompositions fréquentielles). Compte tenu de la complexité du propos, les auteurs se sont tous appuyés sur une modélisation des distributions de fréquences pour évaluer les similarités entre contenus d'images. Nous pouvons citer par exemple, Drimbarean and Whelan (2001) qui utilisent les filtres de Gabor avant d'extraire des moments statistiques. Soulard et al. (2013) utilisent les quaternions pour incorporer la couleur dans les ondelettes.

Une autre façon de caractériser la non-uniformité d'une surface/scène est de considérer cette question sous l'angle des gradients (fig. 4.2 : gradients). Une surface uniforme présentera en tout point des gradients de norme nulle, par opposition à une surface non-uniforme. Cette famille met en avant les gradients des images générés autour du point d'intérêt. Un des descripteurs les plus célèbres est SIFT développé par Lowe (1999, 2004). Il est basé sur des histogrammes de gradients orientés. Malgré son efficacité, il est relativement lent à calculer. Pour l'améliorer, Ke and Sukthankar (2004) proposent d'utiliser l'analyse en composante principale (ACP) pour ne garder que les plus fortes composantes. Une autre proposition s'éloignant plus nettement de l'expression de Lowe, est celle faite par Bay et al. (2006) qui bien que légèrement moins efficace est beaucoup plus rapide.

La calcul des gradients est relativement coûteux, même s'il a les bonnes propriétés de relation directe avec la notion de contraste local, global… Dans les années 70, Haralick et al. (1973) ont réduit cette question à celle d'une simple différence (fig. 4.2 : différences de valeurs). Cette réduction est cohérente avec les premières conjectures de Julesz (1962). Et elle est à l'origine de la famille la plus étendue (Unser (1986); Galloway (1974) ; etc.).

La dernière famille crée des motifs autour du point d'intérêt(fig. 4.2 : motifs). Ces motifs ou textons sont les attributs de cette famille. Le texton définit un élément primitif d'une texture distribuée de façon aléatoire ou déterministe permettant de la caractériser. En anglais, le terme *pattern* contient une certaine notion de répétabilité que l'on retrouve de manière plus diffuse ici. En effet, une texture naturelle ne se répète pas *stricto sensu*, mais suppose une distribution de probabilité similaire impliquant une variabilité du texton lui-même.

La première branche de cette famille n'utilise pas de données binaires pour la création du descripteur (fig. 4.3). Florindo et al. (2016a) emploient les histogrammes de pixels connectés. Cela n'est pas sans rappeler les longueurs de plages présentées par Chu et al. (1990). La deuxième branche de cette famille comprend les *Local Binary Pattern* (LBP) introduits par Ojala et al. (1994). Ils décrivent les relations entre un pixels et ses voisins par comparaison binaire, puis créent un histogramme pour décrire toute la zone d'intérêt.

Évidement, certains auteurs composent avec les différentes branches, par exemple, Nguyen and Caplier (2015) calculent des histogrammes de motifs extraient des gradients de l'image initiale .

Chacune des ces familles propose des descripteurs dit couleur. Ces approches couleur se scindent en quatre catégories :

— GLACI (*Grey-level approaches with colour information*) ;

— CMA (*Combined marginal approaches*) ;

— CCMA (*Cross-channel marginal approaches*) ;

— vectorielle.

GLACI combine une approche texture en niveaux de gris avec une approche couleur séparément comme le présentent Singh and Hemachandran (2012). CMA mesure les attributs d'aspect non-uniforme sur chacun des canaux de façon marginale (Karakasis et al. (2015)). La catégorie CCMA est une approche marginale qui prend aussi en compte la corrélation entre les canaux (Mindru et al. (2004)). Les approches vectorielles considèrent la couleur et l'aspect non uniforme en un tout et les

mesurent de façon vectorielle (Jebali et al. (2018)). Nous retrouverons ces approches dans les descripteurs des cinq familles, les CCMA et approches vectorielles étant moins usitées.

4.1.1 Valeurs de pixels

Cette famille s'intéresse seulement aux valeurs des pixels sans analyse plus complexe du contenu de l'image.

Contenu uniquement chromatique

Une tendance naturelle est de considérer les variations spatio-chromatiques uniquement sous leurs aspects chromatiques, indépendamment de l'agencement spatial des couleurs.

Swain and Ballard (1991) proposent de conserver l'information couleur sous forme d'histogramme. Ils utilisent ensuite différentes comparaisons possibles entre les histogrammes. L'utilisation d'histogrammes permet de gagner l'invariance à la rotation et à la taille de l'objet à condition de normaliser la distribution. Le calcul s'effectue dans l'espace des couleurs opposées.

van de Weijer and Schmid (2006) s'interrogent sur l'intégration de la couleur dans un descripteur. Ils proposent d'utiliser les invariants colorimétriques regroupés sous forme d'histogramme.

Yap and Paramesran (2006) mesurent l'histogramme des couleurs de l'image dans l'espace des opposés. Ils le décomposent ensuite en le projetant dans l'espace des polynômes de Legendre pour ne garder que les coefficients comme attributs du descripteur.

Contenu spatio-chromatique

Les moments couleur généralisés, introduits par Mindru et al. (2004), tentent d'extraire les deux en même temps. Ils prennent en compte la localité et la couleur du pixel dans un même moment. Loosli et al. (2013) appliquent ces moments couleur à la segmentation supervisée. Les moments sont calculés dans RGB et prennent une partie de la corrélation en compte grâce aux moments inter-canaux.

Rubner et al. (2000) calculent l'histogramme couleur et spatial de l'image dans l'espace $CIELAB$. Karakasis et al. (2015) reprennent les moments affines invariants et les appliquent en marginal sur chacun des canaux RGB. Bony and Fernandez-Maloigne (2016) adaptent les matrices de longueur de plage à la couleur dans l'espace $CIELAB$ pour la classification.

Singh and Hemachandran (2012) utilisent l'union d'un descripteur d'aspect non-uniforme avec un descripteur couleur. Les trois premiers moments couleur sont extraits sur chacun des canaux de l'espace HSV. Leur hypothèse est que l'information "d'homogénéité d'une région" peut être obtenue directement à partir du plan intensité. Ils emploient les filtres de Gabor pour extraire ces caractéristiques de texture. Au delà de la validité de l'hypothèse utilisée, ce travail de 2012 revient encore et toujours sur la question de l'analyse d'une texture couleur comme un tout. L'emploi d'une transformation par famille de filtres nous amène directement à la famille de descripteurs suivante.

4.1.2 Décomposition fréquentielle

Cette famille considère l'analyse de la non-uniformité sous l'angle d'une décomposition fréquentielle, ou d'une décomposition du contenu en bandes fréquentielles (Gabor typiquement).

Contenu uniquement spatial

Cette décomposition s'appuie classiquement sur une transformée dans l'espace de Fourier, ou sur un ensemble de filtres comme c'est le cas d'une décomposition en ondelettes ou de Gabor.

Au delà d'une simple décomposition, Ojansivu and Heikkilä (2008) proposent un descripteur invariant au flou à l'aide de la transformée de Fourier discrète. Schulz-Mirbach (1994) utilise la moyenne d'un patch comme information caractéristique de la zone d'intérêt et propose de l'utiliser après une transformée dans l'espace de Hilbert.

Koenderink and van Doorn (1987) s'intéressent aux différents niveaux de dérivations pour créer un ensemble de filtres permettant la création d'un descripteur. Freeman and Adelson (1991) décomposent l'information contenue dans la zone d'intérêt à l'aide de filtres orientables sur différentes échelles. Schaffalitzky and Zisserman (2002) utilisent une famille de filtres permettant une invariance par rotation.

Contenu spatio-chromatique

Historiquement, l'information de texture ou d'aspect non uniforme était celle qui décrivait le plus le contenu avant l'arrivée des images couleur. Il a fallu attendre la montée des tâches d'indexation pour que cela redevienne le cas.

Drimbarean and Whelan (2001) proposent une adaptation couleur de différents descripteurs d'aspect non-uniforme, notamment l'analyse par transformée cosinus discrète et par les filtres de Gabor. Deux approches sont présentées : une union entre un descripteur texture et un descripteur couleur et une approche marginale des descripteurs texture. Les résultats présentés par les auteurs confortent l'avantage d'une approche liant texture et couleur dès le début.

Mennesson et al. (2014) se servent de la transformée de Fourier dans l'algèbre de Clifford pour estimer des moments couleur à l'aide des quaternions. Denis et al. (2007) analysent l'information contenue dans la décomposition fréquentielle basée sur les quaternions. Soulard et al. (2013) proposent une extension couleur d'ondelettes se focalisant sur les phases induites par la transformation de Riesz. L'apport conséquent de ces approches est d'enfin proposer une écriture vectorielle de la transformée de Fourier. Le problème posé est alors celui de l'écriture de l'attribut.

Gecer et al. (2017) proposent l'apprentissage de mélange de filtres couleur pour la reconnaissance d'objet. Les auteurs utilisent l'espace $CIELAB$.

Cette famille pourrait être plus amplement détaillée mais ce n'est pas la direction que nous avons choisie. Les approches couleur dans des espaces fréquentiels multiplient les corrélations de phases inter-canaux compliquant l'extraction d'information compacte et discriminante.

4.1.3 Gradients

Cette famille détaille l'aspect non-uniforme au travers des gradients autour du point d'intérêt.

Contenu uniquement spatial

Un des premiers descripteurs de cette catégorie est l'histogramme de gradients orientés (HOG pour *histogram of oriented gradients*) proposé par Freeman and Roth (1995). Appliqué à toute l'image, l'idée est reprise par Dalal and Triggs (2005) mais sur une zone locale. Cette notion de description par gradient orienté est suffisamment robuste pour pouvoir apparier des photos, des dessins et des peintures comme le montrent Aubry et al. (2016). C'est ce principe qui est repris dans la plupart des descripteurs présentés dans cette sous-section.

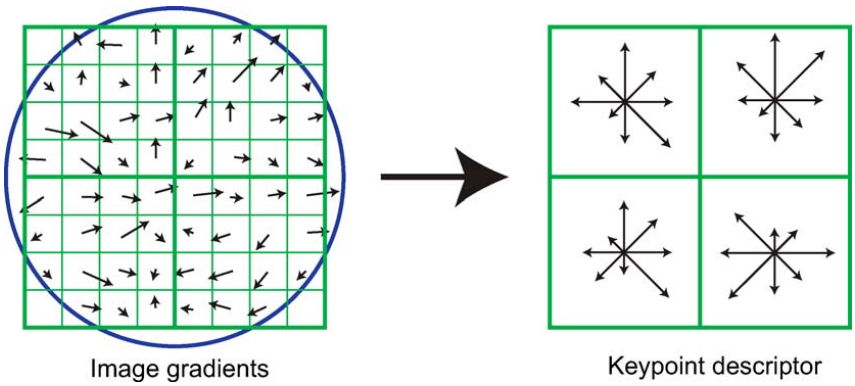


FIGURE 4.4 – **Schéma de conception de SIFT**. Image extraite de l'article de Lowe (2004).

Lowe (1999) propose un nouveau couple détecteur/descripteur. Le détecteur est basé sur une différence de Gaussiennes comme nous l'avons vu au chapitre précédent qui permet de déterminer l'angle et la taille du point d'intérêt. Le descripteur est donc basé sur les histogrammes de gradients orientés. La partie descripteur est encore très utilisée aujourd'hui de par son efficacité. L'idée principale est que l'orientation et les gradients sont plus importants à la vision humaine qu'une petite délocalisation du point d'intérêt. Cette méthode calcule les gradients autour de la région du point d'intérêt divisé en 4×4 sous régions. Pour chaque sous région l'histogramme de gradients orientés est estimé. La figure 4.4 présente l'idée générale de regroupement des histogrammes de gradients orientés. L'invariance à la luminosité est obtenue par une normalisation $\mathcal{L}_2$. L'invariance à la rotation et à l'échelle est due à la différence de Gaussienne. Le descripteur SIFT, bien que très efficace, est lent à calculer et sa licence n'autorise pas une utilisation libre de son implémentation.

Un autre descripteur, SURF, beaucoup plus rapide à estimer mais un peu moins efficace est introduit par Bay et al. (2008). Le détecteur associé doit donner une estimation de l'échelle du point d'intérêt. SURF se décompose en deux parties, tout d'abord l'estimation de l'orientation du point d'intérêt, puis les descriptions du point d'intérêt à proprement parlé. SURF se base sur les ondelettes de Haar pour estimer une approximation des gradients dx et dy sur une image lissée. Comme pour SIFT, la région du point d'intérêt est subdivisée. Pour chacune des sous régions, le descripteur évalue $\sum dx$, $\sum dy$, $\sum |dx|$ et $\sum |dy|$. Le descripteur est enfin normalisé avec la norme $\mathcal{L}_2$.

À notre connaissance, DAISY, créé par Tola et al. (2010), est le seul descripteur conçu pour une utilisation dense. Des mesures de gradient sont effectuées autour du point d'intérêt à différents angles, rayons et tailles de zones à comparer. Une orientation est estimée de ces mesures. Chaque zone est convoluée avec un noyau Gaussien de taille proportionnelle à la distance au centre du point d'intérêt. Ce noyau correspond à un pétale. Le descripteur est constitué de nombreux pétales qui forment une fleur d'où le nom DAISY. Une série d'histogrammes de gradients qui sont normalisés indépendamment les uns des autres est ainsi obtenue.

Contenu spatio-chromatique

De nombreuses adaptations de SIFT à la couleur ont été publiées. Abdel-Hakim and Farag (2006) calculent le descripteur sur les trois canaux des invariants couleur proposés par Geusebroek et al. (2001). Burghouts and Geusebroek (2009) emploient ces invariants couleurs afin de rendre leur proposition de SIFT robuste au changement d'illuminant. Une autre adaptation dans l'espace HSV proposée par Bosch et al. (2008) utilise une grille dense avec cette transposition de SIFT. van de Weijer and Schmid (2006) calculent SIFT classiquement et ajoutent un histogramme de dérivées couleur dans l'espace des couleurs opposées. van de Sande et al. (2008) comparent les performances de SIFT appliqué sur différents canaux de différents espaces couleur. Tian et al. (2017) s'inspirent du fonctionnement par opposé de l'œil humain pour créer un espace de doubles opposés où sera adapté SIFT.

Une modification de SURF est proposée par Markatopoulou et al. (2015) pour prendre en compte la couleur. Les auteurs calculent sur chaque canal couleur indépendamment le descripteur puis réduisent sa dimension à l'aide d'une analyse en composante principale. Ils proposent aussi cette approche dans l'espace des couleurs opposées.

Montesinos et al. (1998) emploient les invariants différentiels du premier ordre sur les canaux RGB pour créer des histogrammes.

Les descripteurs de cette famille sont majoritairement utilisés avec la norme $\mathcal{L}_2$ malgré leur forme d'histogramme. Le point noir de ces descripteurs, par ailleurs très efficaces est le temps de calcul.

4.1.4 Différences de valeurs

Cette famille utilise des différences de pixels comme attribut. Nous retrouvons les notions de dérivation et de gradient utilisées dans une sous-section précédente dans cette notion de différence de pixels. Les méthodes présentées ici s'affranchissent de la limite imposée par la dérivation, et du lissage induit pour la limitation du bruit. Comme les différences se font sur toute la zone d'intérêt, l'histogramme lisse le bruit.

Contenu uniquement spatial

Haralick et al. (1973) introduisent les cooccurrences. Ils calculent la matrice des occurrences du pixel de valeur c_j partant d'un pixel de valeur c_i (fig.4.5a). Dans la décomposition proposée, positionner les cooccurrences dans la section différences de valeurs s'explique par la nature même de ce que cherchait Haralick. Comme nous y reviendrons plus tard (sec. 4.3), Haralick s'est inspiré des travaux de Julesz pour définir son attribut. Et compte tenu des moyens techniques et des images disponibles dans les années 70, il a adapté les propos de Julesz. Mais clairement, au cœur de la probabilité de trouver une valeur c_i pour un pixel et une valeur c_j pour un autre pixel plus loin, se retrouve l'analyse de la probabilité de trouver $(c_j - c_i)$. Haralick extrait ensuite quatorze moments associés à ces matrices comme attributs du descripteur.

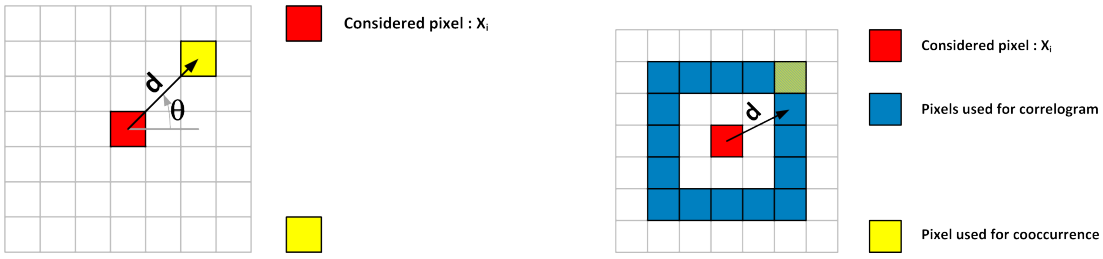
Dans cette même section au cœur des approches existantes, nous devons également considérer les travaux de Unser (1986) qui montrent que les matrices de sommes et de différences offrent une bonne approximation au sens de l'ACP de la matrice de cooccurrences.

Nous devons aussi associer à ce niveau le corrélogramme (fig. 4.5b) comme un attribut rendant invariante par rotation la cooccurrence. Il est intéressant de noter la proximité entre la proposition des corrélogrammes et celle des LBP. Même si les attributs construits diffèrent à la fin, le point de départ : la caractéristique physique à considérer est la même.

Si au début du système d'indexation et de recherche d'image, des descripteurs d'aspect non-uniforme niveaux de gris étaient considérés, la problématique depuis plus de deux décennies est celle de la description spatio-chromatique de l'aspect non-uniforme.

Contenu spatio-chromatique

Hauta-Kasari et al. (1996) adaptent les cooccurrences aux contextes couleur, puis multi-spectral. Les cooccurrences sont mesurées en spatial puis en spectral. L'espace des couleurs spectrales est quantifié pour la mesure des cooccurrences. Le descripteur obtenu est l'union des différents moments de chaque cooccurrence. La troisième proposition de Drimbarean and Whelan (2001) se base sur les moments associés aux cooccurrences. Une version concatène les informations texture et couleur et l'autre applique la cooccurrence en marginal sur un espace RGB. Arvis et al. (2004) proposent



(a) Cooccurrence (b) Corrélogramme

FIGURE 4.5 – **Illustrations des approches par différence de valeurs**. Images extraites du cours texture de Noël Richard 2019.

trois adaptations couleur. La première estime les cooccurrences sur les trois canaux RGB et les intercorrélations entre canaux. La seconde se place dans l'espace HSV et extrait des moments purement couleur puis intensité. La troisième quantifie l'espace couleur pour obtenir un espace peu parcimonieux lors du calcul des cooccurrences. Le chapitre sur les mesures de similarité nous a permis de montrer les difficultés rencontrées face à une union lors du choix de la distance à employer. L'approche marginale bien que préférable est toujours limitée dans un espace de type RGB qui n'est pas orthogonal.

Martínez et al. (2015) proposent les *Colour Contrast Occurrences* ($\mathrm{C_2O}$). Ils s'affranchissent de la valeur du pixel initial et seules les valeurs des différences créent l'histogramme $3D$. Afin d'obtenir une différence perceptuelle, ils effectuent le calcul dans l'espace $CIELAB$. Ils extraient un vecteur $1D$ de l'histogramme $3D$ correspondant à un éloignement hélicoïdal du centre de la distribution. Richard et al. (2016a) proposent d'employer le volume de l'histogramme $3D$ obtenu à plusieurs échelles pour discriminer la complexité d'une texture fractale synthétique.

4.1.5 Motifs

Cette famille s'appuie sur la notion de texton pour décrire le contenu de la zone d'intérêt. Le texton correspond à un motif variant légèrement d'une répétition à l'autre.

Contenu uniquement spatial

Ce sont les *local binary pattern* (LBP) qui initient cette famille. Ojala et al. (1994) proposent une méthode très orientée informatique pour décrire l'organisation des pixels autour d'un pixel central. L'ensemble de ces organisations crée un histogramme qui constitue le descripteur. La méthode consiste à comparer les pixels du tour au central et de ne conserver que la réponse binaire de la comparaison.

Les LBP ont servi d'inspiration lors de la création des descripteurs de la suite de cette section. En effet, utiliser uniquement le résultat booléen d'une comparaison comme attribut d'un descripteur permet d'employer la distance de Hamming (1950) et gagner en temps de calcul. La signature obtenue dépend cependant des choix de pixels à comparer et c'est dans ce sens que nous les qualifions de motifs.

BRIEF (*Binary Robust Independent Elementary Features*) introduit par Calonder et al. (2010) est un des premiers descripteurs binaires. Il est basé sur des différences entre pixels choisis aléatoirement dans le voisinage du point d'intérêt. Tout d'abord l'image est lissée, puis les paires de pixels (x, y) sont choisies autour du point d'intérêt. La sélection suit une loi normale centrée autour du point d'intérêt. Chaque paire est comparée avec :

$$b = \begin{cases} 1 & \text{si } I(x) < I(y) \\ 0 & \text{sinon .} \end{cases} \tag{4.1}$$

La valeur de la différence entre les pixels n'a pas d'influence sur l'estimation du descripteur. Cela assure l'invariance au contraste et à la luminosité. Cependant, le moindre changement d'orientation impacte sur les paires choisies et donc sur l'appariement possible entre deux points. Pour des paires de pixels choisies aléatoirement, ce descripteur donne des résultats très corrects. Cependant, sa sensibilité à la rotation a induit de nouvelles propositions.

Pour pallier la sensibilité à la rotation, Rublee et al. (2011) introduisent le couple détecteur/descripteur ORB (*Oriented fast and Rotated Brief*). Ils utilisent le même principe que BRIEF mais prennent en compte l'orientation. Pour chaque orientation procurée par le détecteur FAST, la sélection de paires à comparer est adaptée à l'orientation estimée par le détecteur. Ils proposent aussi l'apprentissage des paires pour améliorer l'efficacité du descripteur. Ces deux descripteurs n'utilisent pas symétriquement l'information autour du point d'intérêt.

Leutenegger et al. (2011) proposent BRISK (*Binary Robust Invariant Scalable Key points*). Ce descripteur demande une information sur l'échelle du point en entrée. Contrairement à BRIEF, toutes les paires sont uniformément réparties autour du point d'intérêt. Le lissage du point à comparer dépend de la distance du point au centre du point d'intérêt. Les paires sont séparées en deux ensembles, $\mathcal{L}$ et $\mathcal{S}$, correspondant respectivement aux petites et grandes distances entre les éléments d'une paire. L'ensemble des grandes distances permet d'évaluer l'orientation du point d'intérêt :

$$\begin{pmatrix} g_x \\ g_y \end{pmatrix} = \frac{1}{\#\mathcal{L}} \sum_{(p_i,p_j)\in\mathcal{L}} \frac{I(p_i,\sigma_i) - I(p_j,\sigma_j)}{\|p_i - p_j\|^2} \begin{pmatrix} x_i - x_j \\ y_i - y_j \end{pmatrix}, \tag{4.2}$$

où p est un pixel, (x, y) ses coordonnées et g le vecteur résultant de l'équation. À partir du gradient g, l'orientation est extraite. Les éléments de l'ensemble $\mathcal{S}$ sont modifiés en fonction de l'angle estimé. Les nouvelles paires sont comparées pour obtenir le descripteur.

Similairement, FREAK (*fast retina rey point*) présenté par Alahi et al. (2012), s'inspire de la perception humaine. La rétine humaine est composée de trois zones : fovéa, para-fovéal et péri-fovéal. Ce descripteur est basé sur ce principe. Différents points autour du point d'intérêt sont sélectionnés. Plus, ils sont éloignés du point d'intérêt, plus le filtre Gaussien qui leur est appliqué est grand. En effet, le nombre de cônes diminue avec l'éloignement de la fovéa, la zone est donc moins détaillée. La mesure de l'orientation est la même que pour BRISK mais le nombre de paires pour l'estimation est plus petit et choisi symétriquement. Cette analyse du grossier au plus fin, est censée reprendre le système de vision humaine.

Un des derniers descripteurs proposés est LATCH (*Learned Arrangements of Three Patch Codes*) introduit par Levi and Hassner (2015). L'usage d'un filtrage Gaussien entraîne un perte d'information pour la plupart des descripteurs binaires. Pour résoudre ce problème, les auteurs utilisent des patchs 3×3 autour du point d'intérêt. Les patchs sont comparés deux à deux grâce à la distance de Frobenius. La comparaison de deux patchs à un troisième dit ancre permet d'obtenir l'élément binaire. Le choix des patchs a été appris, puis sélectionné pour éviter une trop grande corrélation. Ce descripteur est à la frange de la famille binaire puisque même si la distance de Hamming est utilisée pour l'appariement, la distance de Frobenius est employée pour la comparaison des patchs. Cela augmente le temps de calcul, mais surtout cela complexifie le système de comparaison pour revenir à des descripteurs type texture qui vont être présentés dans la prochaine section.

Lazebnik et al. (2003) calculent l'histogramme des intensités de pixels en fonction de la distance au cente de la région d'intérêt.

Contenu spatio-chromatique

Les adaptations des descripteurs utilisant une signature binaire, se contentent d'utiliser les approches niveaux de gris de façon marginale sur un espace RGB. Ziegler et al. (2012) étendent le descripteur BRIEF à la couleur. *Locally Uniform Comparison Image Descriptor* (LUCID) utilise BRIEF de

façon marginale sur RGB en triplant ainsi sa longueur. Markatopoulou et al. (2015) adaptent ORB à la couleur de la même manière qu'ils ont adapté SURF de façon marginale en RGB ou sur l'espace des couleurs opposées. Chen et al. (2016) proposent une adaptation couleur du descripteur FREAK. Les comparaisons se font marginalement sur les canaux RGB sur les pixels correspondant à la zone fovéale de la vision.

Les approches visant à adapter les LBP à la couleur sont un peu plus diversifiées. Mäenpää and Pietikäinen (2004) effectuent une quantification de l'espace couleur pour pouvoir créer un histogramme non parcimonieux des LBP mesurés. Ils comparent les résultats obtenus dans différents espaces couleur ($CIELAB$, RGB, HSV, etc.). Connah and Finlayson (2006) mesurent les LBP mais en créant un histogramme $3D$ dans RGB pour prendre en compte la couleur en même temps que l'information spatiale. Porebski et al. (2013) mesurent les différences entre le centre et les pixels extérieurs sur des canaux différents en RGB (approche inter-canaux). Richard et al. (2016b) s'inspirent de LBP pour créer les CLP (*Colour Local Pattern*). Les CLP se calculent dans l'espace $CIELAB$. Leurs signatures se composent des différences couleur entre le pixel central et les pixels distants de d.

Parmi les approches de type motifs, les binaires sont les plus célèbres, il en existe d'autres prenant en compte toute l'information autour du point d'intérêt sans binarisation. Matas et al. (2000) sélectionnent les zones multicolores uniquement. Elles sont ensuite décrites par leurs statistiques dans un espace RGB. Alvarez and Vanrell (2012) emploient les laplaciens de Gaussiennes pour déterminer la forme de la zone d'intérêt puis utilisent l'espace HSV pour décrire la couleur de la zone. Cela leur permet ensuite de créer un histogramme contenant les formes et les couleurs décrivant la zone. Ils créent ainsi des sacs de mots visuels à partir de la théorie du texton. Nous pouvons effectivement ajouter les approches de type sacs de mots visuels à cette famille de type motifs puisqu'il s'agit de décrire l'image ou la zone à l'aide de mots visuels qui sont une forme de texton. Ce type d'approche étant généralement utilisé pour des applications de plus haut niveau, nous ne nous attarderons pas dessus dans ce chapitre.

Chacun des descripteurs présentés se targue de certaines invariances dues à sa construction. Nous nous sommes demandés dans quelle mesure cette construction permettait les invariances proposées. C'est l'objet de la prochaine section.

4.2 Invariance des descripteurs

Notre objectif est d'étudier si les hypothèses d'invariance faites par les auteurs sont vérifiées en pratique. Dans ce but, nous avons mis les descripteurs dans des conditions le plus idéales possibles : tous les points d'intérêt présents dans l'image initiale le seront dans l'image transformée, l'appariement ne dépend pas d'une vérité-terrain humaine, pas de saturation sur la transformation luminosité. Nous espérons donc des résultats “parfaits”. La pratique est moins clémente.

Nous présenterons dans un premier temps le protocole de l'expérience, détecteur et descripteurs étudiés, images et transformations utilisées, critère d'analyse. Puis nous étudierons les résultats.

4.2.1 Protocole

4.2.1.1 Images et transformations utilisées

Nous utilisons la base de données d'images *photo CV KODAK* constituée de 24 images couleur de taille 768×512. **Les descripteurs comparés travaillent uniquement sur la couche luminance de l'image**. Les images ont donc été transformées en niveau de gris pour l'expérience.

Une base de données images avec appariement des points par vérité-terrain dépend nécessairement de l'utilisateur qui a fixé cette vérité. Or l'erreur est humaine. Nous avons donc choisi d'ef-

fectuer une transformation automatique des images nous permettant de calculer où se trouvent les nouveaux points d'intérêt.

Les images sont ensuite déformées par cinq types de transformations : changement d'échelle, rotation, cisaillement, variation de luminosité et réflexion. Afin de ne pas créer d'occlusion lors de la rotation, chaque image est redimensionnée par ajout de zéro sur les bords. La figure 4.7b montre un exemple d'image redimensionnée. Un récapitulatif des transformations effectuées est donné table 4.1.

TABLE 4.1 – **Résumé de transformations créées**.

Transformation	Nombre total	De	À	Pas
Échelle	8	$/1.5$	$/5$	0.5
Rotation	35	$10°$	$350°$	$10°$
Cisaillement	16	$10°$	$80°$	$10°$
Luminosité	15	0.2	3	0.2
Réflexion	2			

Tout d'abord nous transformons l'image par changement d'échelle. L'image est divisée par un facteur allant de 1.5 à 5 par pas de 0.5. Nous n'augmentons pas la taille de l'image car cela demanderait une interpolation pour certaines valeurs de pixels et donc conduirait à une addition d'information. La taille de la zone d'intérêt est adaptée en conséquence pour les images transformées. Un exemple de changement d'échelle est donné figure 4.6b.

La deuxième famille de modifications appliquées à l'image est la rotation. Pour ces transformations, nous appliquons pour chaque angle compris entre $10°$ et $350°$ avec un pas de $10°$, la matrice suivante :

$$\begin{pmatrix} x_\theta \\ y_\theta \end{pmatrix} = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \tag{4.3}$$

où (x_θ, y_θ) et (x, y) sont les coordonnées du pixel et θ est l'angle de rotation. Le centre de l'image est l'origine de la rotation. La figure 4.6c présente un exemple de transformation par rotation.

Pour analyser l'impact de l'information angle sur les descripteurs, l'expérience est menée une fois sans information à priori de l'angle et une fois avec.

La troisième famille de transformations correspond à un changement de point de vue. Dans cette étude, le cisaillement projette l'image sur le plan (x, y) comme si l'image effectuait une rotation autour d'un axe horizontal (bas de l'image fixe) ou vertical (coté gauche fixe). L'observateur permettant la projection dans le plan est posé au centre de l'images. Les angles de rotations choisis vont de $10°$ à $80°$ avec un pas de $10°$. La matrice de rotation pour le cisaillement vertical est :

$$\begin{pmatrix} x_\theta \\ y_\theta \\ p_\theta \end{pmatrix} = \begin{pmatrix} D\cos(\theta) & 0 & \frac{DCol(\cos(\theta)-1)}{2} \\ 0 & D & 0 \\ \sin(\theta) & 0 & D + \frac{Col\sin(\theta)}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{4.4}$$

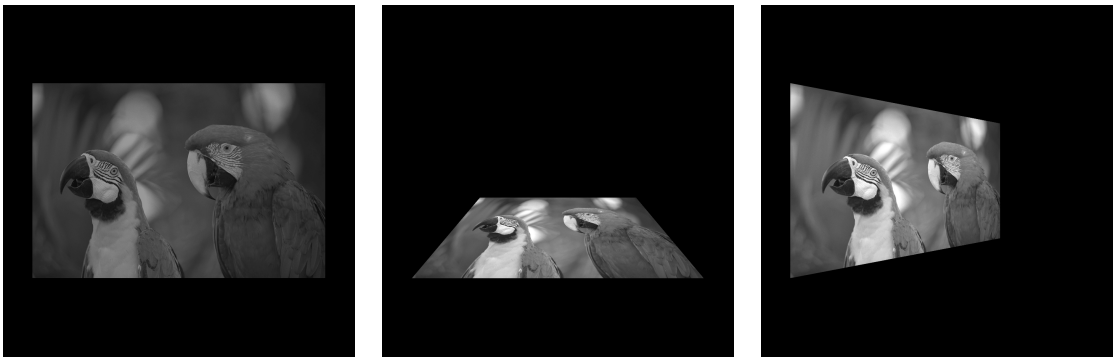
et l'horizontal :

$$\begin{pmatrix} x_\theta \\ y_\theta \\ p_\theta \end{pmatrix} = \begin{pmatrix} D & 0 & 0 \\ 0 & D\cos(\theta) & \frac{DRow(\cos(\theta)-1)}{2} \\ \sin(\theta) & 0 & D + \frac{Row\sin(\theta)}{2} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{4.5}$$

où p_θ est la profondeur du pixel résultant. Les coordonnées $3D$ sont ensuite projetées pour obtenir une image $2D$. θ correspond à l'angle de rotation, Col et Row sont respectivement le nombre de colonnes et lignes de l'image originale (768×512). Enfin, D est la distance de l'observateur. Ici, nous avons choisi D pour que la largeur de l'image corresponde au champ de vision humain habituel ($80°$).



(a) Image initiale (b) Échelle divisée par 3.5 (c) Rotation : $\theta = 40°$



(d) Luminosité multipliée par 0.6 (e) Cisaillement horizontal : $\theta = 70°$ (f) Cisaillement vertical : $\theta = 30°$

FIGURE 4.6 – **Exemple de transformations appliquées à une image**.

Il faut noter que les matrices des équations (4.3), (4.4) et (4.5) considèrent le centre de l'image comme origine. Dans la pratique, il faudra faire une translation de coordonnées.

La taille du point d'intérêt est adaptée pour ne pas avoir de chevauchement. Il y a une légère rotation des points d'intérêt, mais celle-ci n'est pas uniforme sur l'image, nous ne l'avons donc pas prise en compte.

Le quatrième type de modifications vise à mesurer l'invariance au changement de luminosité. Nous modifions donc l'intensité du pixel et non sa localisation. Nous choisissons d'appliquer un changement linéaire de luminosité avec saturation. Cela est censé représenter le comportement d'un appareil photo réglé avec un mauvais temps d'exposition. Le coefficient de luminosité varie de 0.2 à 3 avec un pas de 0.2. Les nouvelles valeurs d'intensité sont données par :

$$G = \begin{cases} \alpha G & \text{si } \alpha G < 1 \\ 1 & \text{sinon,} \end{cases} \tag{4.6}$$

où G est le niveau de gris et α le coefficient de changement de luminosité.

Les dernières transformations sont deux transformations par symétrie, une verticale puis une horizontale. Il est important de noter la différence induite par ces transformations. Dans les autres cas, nous espérons que les descripteurs permettront l'appariement correct des points d'intérêt. Ici, nous ne souhaitons pas d'appariement juste. En effet, deux objets symétriques par rapport à un axe ne représentent à priori pas la même chose, prenez le 2 et le 5 ci-dessous par exemple.



TABLE 4.2 – **Invariance des descripteurs selon leurs auteurs**.

	Échelle	Rotation	Cisaillement	Luminosité
SIFT Lowe (2004)	x	x	x	x
SURF Bay et al. (2008)	x	x	x	x
DAISY Tola et al. (2010)			x	x
BRIEF Calonder et al. (2010)				x
ORB Rublee et al. (2011)		x		x
BRISK Leutenegger et al. (2011)	x	x		x
FREAK Alahi et al. (2012)	x	x	x	x
LATCH Levi and Hassner (2015)		x		x

Nous obtenons un total de 76 images transformées par cinq types d'altération. Nous avons choisi de ne pas mesurer l'impact du floutage ou de la compression *jpeg*. Dans notre domaine d'application avec l'amélioration de l'archive-digitale, nous supposons que les images que nous utiliserons seront nécessairement nettes et non-compressées ou compressées sans perte.

4.2.1.2 Détecteur, descripteurs et critère de comparaison

Huit descripteurs monochromatiques ont été choisis pour cette étude. La liste des descripteurs utilisés ainsi que les invariances dont ils sont supposés faire preuve est donnée dans la table 4.2. Nous fixons pour chaque point d'intérêt la localisation de ceux-ci pour toutes les transformations d'images. Pour les descripteurs le permettant nous fixons aussi la taille du point d'intérêt.

Nous avons choisi d'étudier les descripteurs suite à un détecteur dense. En effet, la puissance de calcul augmentant rapidement, la multiplication des opérations n'est plus un problème majeur. C'est aussi une pratique qui semble offrir des résultats intéressants comme le montrent Tola et al. (2010). Hietanen et al. (2015) comparent plusieurs descripteurs associés à différents détecteurs dont une grille dense. Ce détecteur donne de bons résultats. Nous souhaitons vérifier l'invariance des descripteurs, et non la capacité d'un détecteur à percevoir les mêmes coins dans deux images avec points de vues différents.

Nous estimons la grille dense de points d'intérêt puis nous appliquons la même transformation à l'image et à la localisation du point d'intérêt ainsi nous étudierons le même point et estimerons l'invariance théorique du descripteur. La figure 4.7 montre les différences depuis l'image originale jusqu'à l'image avec les points d'intérêt sélectionnés.

Pour mesurer l'invariance des descripteurs, nous utilisons l'algorithme du plus proche voisin. Puisque, nous avons choisi une grille dense et interdit les occlusions, le nombre de correspondances est le même que le nombre de points d'intérêt. Le critère de comparaison *précision* se simplifie ici, nous l'appelons taux d'appariement :

$$ap = \frac{|\mathscr{T}|}{|\mathscr{C}|}, \tag{4.7}$$

où $\mathscr{T}$ correspond au nombre d'appariements corrects et $\mathscr{C}$ à l'ensemble de correspondances possibles.

La base de données d'image a été présentée ainsi que le détecteur et les descripteurs. Le critère de comparaison est connu, nous pouvons donc présenter les résultats.



(a) Image initiale (b) Image redimensionnée (c) Points d'intérêt sélectionnés

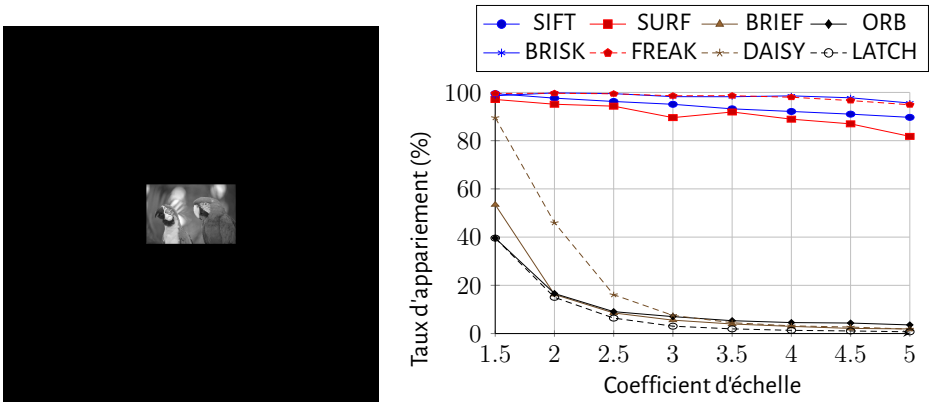
FIGURE 4.7 – **Exemple d'une image de la base *photo CVKODAK***. L'image 4.7a représente l'image telle que dans la base. L'image est ensuite passée en niveaux de gris et redimensionnée 4.7b. 4.7c présente un exemple de sélection dense de points d'intérêt.

4.2.2 Résultats

Les résultats de chacune des invariances sont présentés dans une sous-section, ensuite nous commentons des résultats préliminaires sur la nature de la zone d'étude autour du point d'intérêt. Nous nous intéressons à la taille associée au point d'intérêt, puis à l'importance de l'information angle qui peut être fournie par certains détecteurs.

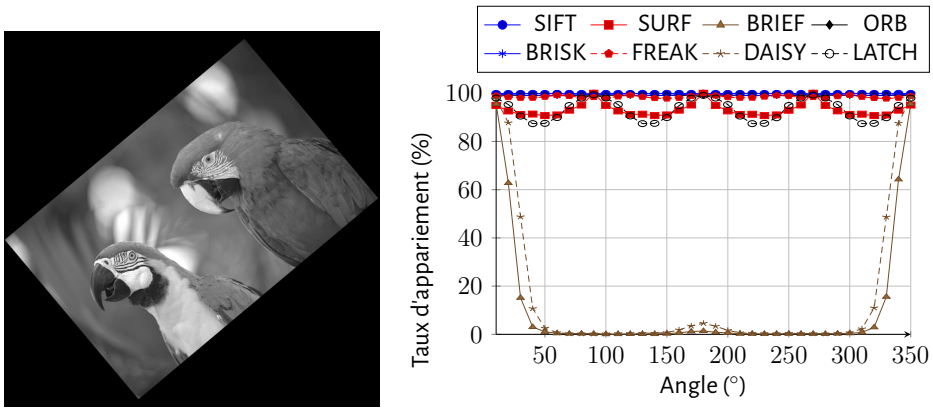
4.2.2.1 Invariance à l'échelle

L'invariance aux changements d'échelle est présentée figure 4.8. Nous pouvons confirmer les invariances de SIFT, SURF, BRISK et FREAK. Les deux derniers produisant les meilleurs résultats. Les autres sont fortement impactés par le changement d'échelle. Ces résultats confirment le tableau des invariances de la table 4.2.



(a) Échelle divisée par 3.5 (b) Échelle

FIGURE 4.8 – **Invariance à l'échelle**. La première figure rappelle l'effet visuel de la transformation mesurée 4.8a, la seconde présente les résultats d'invariance à l'échelle 4.8b.



(a) Rotation : $\theta = 40°$

(b) Rotation avec à priori

FIGURE 4.9 – **Invariance à la rotation**. La première figure rappelle l'effet visuel de la transformation mesurée 4.9a, la seconde présente les résultats d'invariance à la rotation 4.9b.

4.2.2.2 Invariance à la rotation

La figure 4.9 présente l'invariance des descripteurs à la rotation. BRIEF et DAISY ne retrouvent aucun point d'intérêt entre les angles $50°$ et $300°$. Ce qui est cohérent avec la non invariance à cette transformation de ces descripteurs. Pour tous les autres descripteurs, nous observons une baisse du taux d'appariement sur certains angles. Ces angles permettent d'orienter des gradients ou des ondelettes de Haar pour les descripteurs SIFT et SURF. Par construction, ils sont donc invariants par rotation d'un angle multiple de $90°$. Pour ORB, il s'agit d'utiliser les paires de pixels préalablement orientées. La discrétisation due aux pixels implique la même invariance tous les quarts de tour. Le taux d'appariement diminue donc dès l'éloignement des multiples de $90°$. Nous observons une symétrie dans les résultats de $0°$ à $90°$. Le plus grand taux d'erreur est donc atteint pour les angles de $45° + k90°$, k entier relatif.

4.2.2.3 Invariance au cisaillement

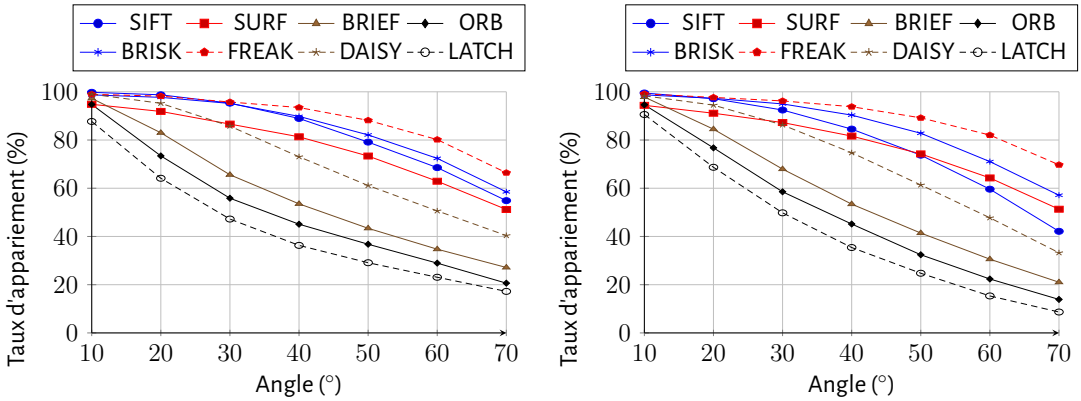
Les résultats du cisaillement horizontal et vertical sont résumés figure 4.10 et sont similaires. Les meilleurs résultats sont obtenus par FREAK qui conserve un taux d'appariement au dessus de $80\,\%$ jusqu'à $60°$. SIFT, BRISK et SURF sont légèrement moins bons et restent au dessus de $60\,\%$ pour la plupart des angles. Alors que le descripteur BRISK ne prétendait pas d'être invariant au cisaillement, il donne de bons résultats. Le taux des quatre autres descripteurs décroît rapidement démontrant la non invariance au cisaillement de ces descripteurs.

4.2.2.4 Invariance à la luminosité

L'influence de la luminosité sur les descripteurs est donnée figure 4.11. A l'exception des DAISY, tous les descripteurs montrent un taux d'appariement qui décroît avec l'éloignement de la luminosité initiale. Notre méthode pour modifier l'illumination implique une saturation lors d'un éclairage trop fort. Cela explique la forte diminution du taux au-delà de 2. Cependant, quand les changements sont faibles, les descripteurs maintiennent un taux d'appariement supérieur à $80\,\%$. DAISY donne des résultats inférieurs à $60\,\%$ d'appariement correct en deçà de 0.5 et au delà de 1.5. Tola et al. (2010) mentionnent que la normalisation des histogrammes indépendamment les uns des autres peut induire un effet étrange dans une application différente de l'appariement pour la création de carte $3D$.

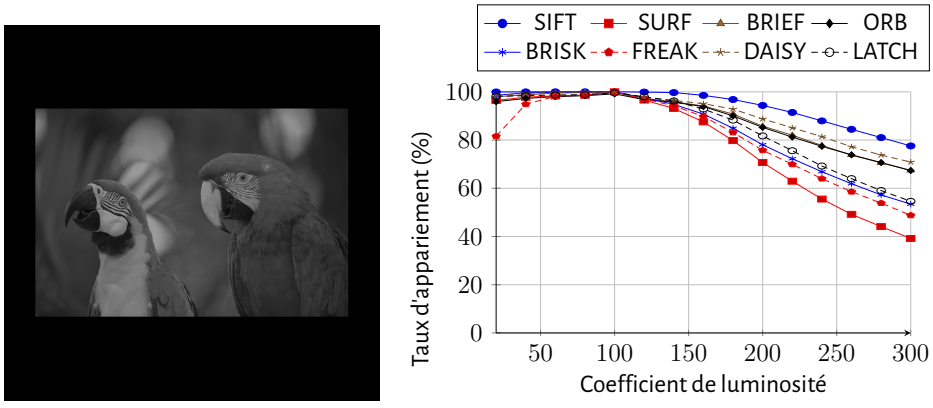


(a) Cisaillement horizontal : $\theta = 70°$ (b) Image initiale (c) Cisaillement vertical : $\theta = 30°$



(d) Cisaillement Horizontal (e) Cisaillement vertical

FIGURE 4.10 – **Invariance aux cisaillements**. Les premières figures rappellent l'effet visuel des transformations mesurées 4.10a et 4.10c, la seconde ligne présente les résultats d'invariance aux cisaillements 4.10d et 4.10e.



(a) Luminosité multipliée par 0.6 (b) Luminosité

FIGURE 4.11 – **Invariance à la luminosité**. La première figure rappelle un exemple de la transformation mesurée 4.11a, la seconde présente les résultats d'invariance à la luminosité 4.11b.

TABLE 4.3 – **Taux d'appariement après transformation par réflexion**.

	Horizontal	Vertical
SIFT	6.56 %	8.03 %
SURF	23.33 %	23.36 %
DAISY	12.86 %	19.28 %
BRIEF	6.92 %	9.89 %
ORB	6.19 %	8.33 %
BRISK	12.19 %	12.42 %
FREAK	11.63 %	11.60 %
LATCH	3.08 %	10.22 %

4.2.2.5 Invariance à la réflexion

La table 4.3 résume les résultats obtenus par réflexion. Une réflexion verticale a un impact moins fort pour l'œil humain qu'une réflexion horizontale qui met les objets sens dessus dessous (en l'occurrence sans dessus dessous). Les descripteurs semblent partager dans une certaine mesure ce phénomène. Le taux d'appariement des descripteurs DAISY et LATCH varie beaucoup selon qu'il s'agit d'une réflexion horizontale ou verticale. $13\,\%$ et $19\,\%$ pour DAISY en horizontal et vertical respectivement, $3\,\%$ et $10\,\%$ pour LATCH. Hormis ceux-ci, les autres descripteurs donnent un résultat similaire quel que soit l'axe de réflexion bien que l'appariement soit légèrement supérieur pour la réflexion verticale, excepté FREAK. SURF est légèrement au-dessus de $20\,\%$. BRISK et FREAK sont autour de $12\,\%$. Les autres sont en dessous. Dans l'ensemble les descripteurs ne sont pas invariants par réflexion, ce qui est le comportement attendu.

En résumé

Les résultats ne permettent pas de répondre à la question des invariances de manière binaire : invariance ou non. En effet, plus la déformation est forte plus le taux d'appariement diminue. L'invariance va donc avec une notion d'échelle de déformation. Pour tenter de répondre à la question des invariances supposées, nous avons donc évalué la moyenne des taux d'appariement sur toutes les transformations d'un type. Les résultats sont consignés dans la table 4.4. En gras, les invariances avancées par les auteurs. Nous remarquons qu'un score peut être attribué aussi aux invariances non attendues. Notons particulièrement le score de BRISK sur l'invariance aux cisaillements supérieur à son score luminosité. Pour les invariances attendues, les scores varient de $70\,\%$ à presque $100\,\%$. Dans l'ensemble SURF répond moins bien aux invariances que les autres descripteurs invariants. Cela est probablement dû à la simplification de calcul qu'impliquent les attributs qui le constituent. L'invariance la plus difficile à tenir sur l'échelle est clairement celle aux cisaillements. En effet, il s'agit d'un produit d'invariance d'échelle de rotation et d'une déformation non linéaire de l'image.

En résumé, des descripteurs répondent comme attendu aux transformations. Cependant, chaque descripteur a une plage d'efficacité. En dehors de cette plage, les résultats déclinent. Dans l'ensemble, les descripteurs ne sont pas invariants aux réflexions.

4.2.2.6 Impact de la taille de la zone d'intérêt

Avant de tester les invariances des descripteurs, nous étudions l'impact de la taille des points d'intérêt. Trois diamètres de zone d'intérêt sont choisis : 10, 50 et 100 pixels. La figure 4.12 résume le comportement des descripteurs. Il est évident que la taille du point d'intérêt influence le résultat.

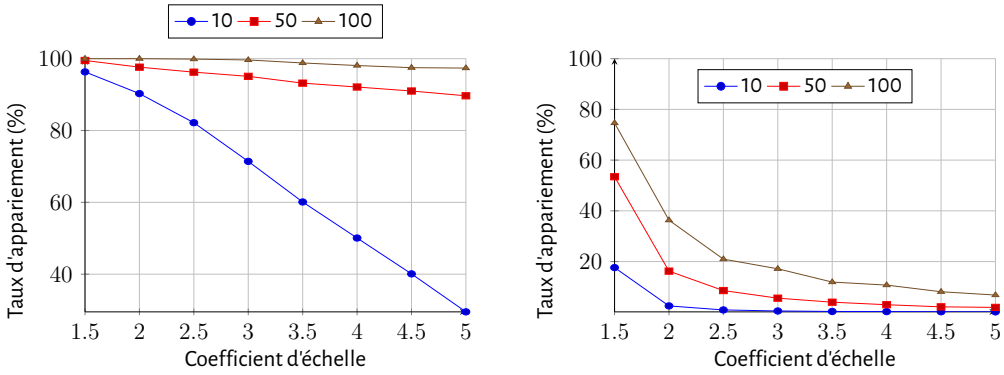
TABLE 4.4 – **Invariances des descripteurs calculés**. En gras, les invariances attendues par les auteurs.

	Échelle	Rotation	Cisaillement	Luminosité
SIFT Lowe (2004)	**94.31** %	**99.93** %	**81.83** %	**94.11** %
SURF Bay et al. (2008)	**90.70** %	**93.13** %	**77.59** %	**77.99** %
DAISY Tola et al. (2010)	21.42 %	14.82 %	**71.52** %	**90.44** %
BRIEF Calonder et al. (2010)	1.85 %	10.46 %	57.24 %	**88.74** %
ORB Rublee et al. (2011)	11.27 %	**99.14** %	49.94 %	**88.06** %
BRISK Leutenegger et al. (2011)	**98.30** %	**99.29** %	84.76 %	**83.71** %
FREAK Alahi et al. (2012)	**98.19** %	**98.49** %	**89.21** %	**80.55** %
LATCH Levi and Hassner (2015)	8.65 %	**93.25** %	42.72 %	**84.87** %

Plus la taille du point d'intérêt est grande, plus il y a d'informations stockées dans le descripteur, et donc un meilleur appariement.

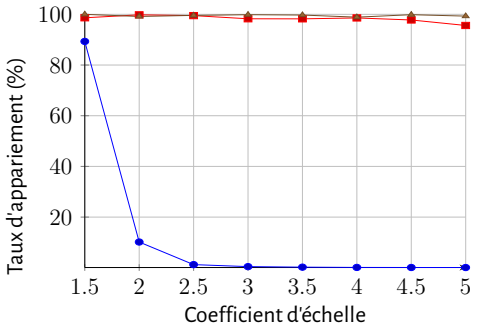
Les tendances des descripteurs peuvent se séparer en plusieurs catégories. Les descripteurs SIFT et SURF répondent similairement à la taille de la zone d'étude (figure 4.12a). Le taux d'appariement décroît régulièrement avec la diminution de la taille de la zone.

La figure 4.12b illustre le comportement de BRIEF, ORB, DAISY et LATCH. Cela est cohérent avec le tableau des invariances, table 4.2. En effet, ces descripteurs utilisent toujours la même fenêtre d'ana-



(a) SIFT (similaire à SURF)

(b) BRIEF (similaire à ORB, DAISY et LATCH)



(c) BRISK (similaire à FREAK)

FIGURE 4.12 – **Comparaison des résultats sur l'invariance d'échelle en fonction de la zone d'intérêt**. Tous les descripteurs sont testés avec différents diamètres de zone (10, 50 et 100 pixels). Tous les descripteurs y sont sensibles. Une plus grande zone contient plus d'informations. La capacité d'invariance à l'échelle diminue avec la taille des points d'intérêt dans 4.12a. Bien qu'une grande taille de point d'intérêt améliore les résultats, 4.12b montre une faible invariance à l'échelle pour ce type de détecteurs. Pour le cas 4.12c, le descripteur perd son invariance avec une taille trop petite car il y a chevauchement d'information.

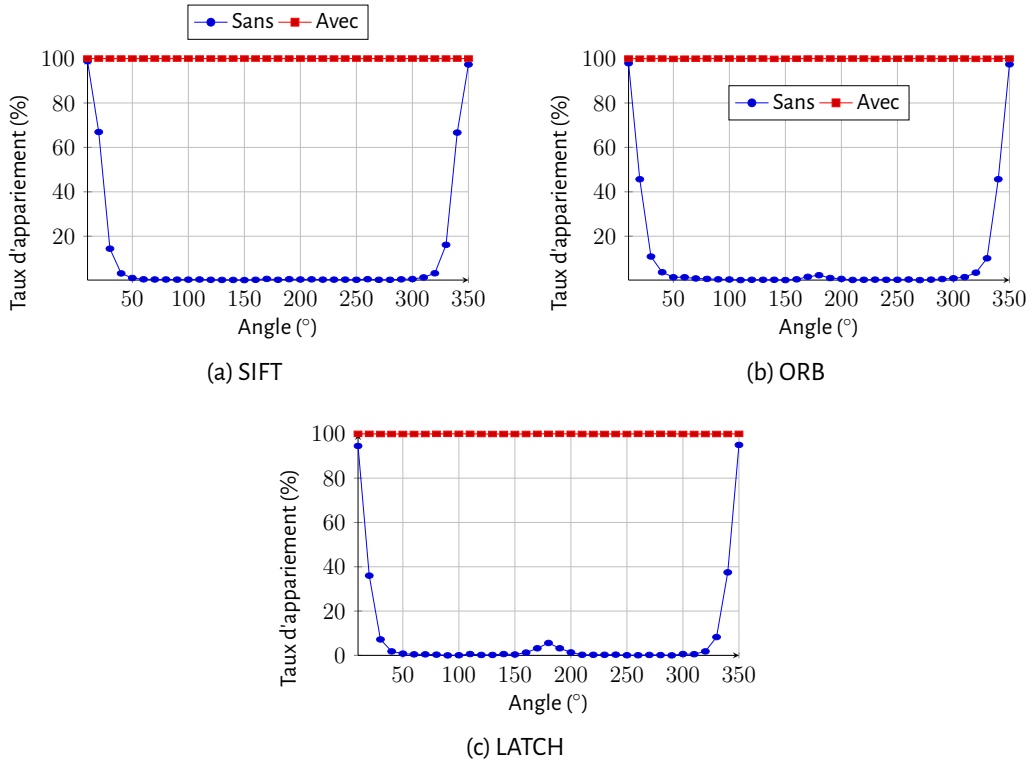
lyse pour décrire le point d'intérêt quelle que soit la taille donnée en entrée. Dans notre cas, plus la taille donnée est petite, plus nous risquons d'avoir un chevauchement des voisinages des points d'intérêt. Cela explique la chute du taux d'appariement avec la diminution de la taille.

Les comportements des descripteurs BRISK et FREAK se ressemblent. Celui de BRISK est présenté figure 4.12c. Le taux d'appariement chute drastiquement pour une taille de zone d'intérêt de 10 pixels. La zone étudiée par ces descripteurs est fixée à environ 30 pixels de diamètre soit supérieure à 10 pixels. Ainsi, il y a chevauchement des zones étudiées et donc mauvais appariement des points d'intérêt. Pour les autres tailles de régions d'intérêt, qui sont supérieures à la zone étudiée par les descripteurs, les résultats sont presque parfaits.

Les résultats présentant l'influence de la zone d'intérêt sur les autres transformations sont présentés en annexe G. D'une manière générale, les tailles 50 et 100 répondent bien, alors qu'avec un diamètre de 10 pixels, le taux d'appariement chute beaucoup. Souhaitant garder un maximum de points d'intérêt sans chevauchement, nous avons choisi d'utiliser pour les expérimentations une taille de zone d'intérêt de 50 pixels.

4.2.2.7 Influence de l'orientation principale de la zone d'intérêt

Certains descripteurs estiment l'angle du point d'intérêt, d'autres utilisent l'information fournie par le détecteur. Nous nous intéressons à l'impact de la connaissance ou non de l'angle à priori. SURF et FREAK calculent l'angle dans la chaîne du descripteur, il n'y a donc aucune variation que l'angle soit fourni ou non avant estimation du descripteur. BRIEF et DAISY n'intègrent aucune information d'angle, une information à priori n'a donc pas plus d'effet sur eux. Les autres descripteurs attendent une valeur d'angle déterminée lors de la détection des points d'intérêt. La figure 4.13 présente les résultats par descripteur de l'influence de la connaissance à priori de l'orientation du point d'intérêt.



(a) SIFT

(b) ORB

(c) LATCH

FIGURE 4.13 – **Comparaison des résultats d'invariance à la rotation avec ou sans à priori**. Quand l'angle correct est fourni, le taux d'appariement est presque parfait pour les descripteurs présentés. S'il n'est pas fourni, il n'y a pas d'appariement correct entre $40°$ et $310°$.

Nous constatons que la connaissance à priori de l'angle est une nécessité. Il faut noter que l'efficacité de ces descripteurs à être invariants à la rotation ne dépend finalement que de la capacité du détecteur à déterminer le “bon” angle.

Cela nous mène à deux questions : comment estimer correctement l'orientation de la zone d'intérêt, et l'angle estimé doit-il l'être au moment de la détection ou de la description ? Fan et al. (2012); Wang et al. (2011) proposent une réponse en libérant les descripteurs du besoin d'angle en effectuant une ligne de partage des eaux sur la région autour du point d'intérêt. En mesurant l'histogramme de la zone extraite, plus aucune information d'angle n'est nécessaire, que ce soit estimée par le détecteur ou le descripteur.

Lors des expériences précédentes, nous avons conservé la connaissance à priori de l'orientation principale de la zone d'intérêt.

Compte tenu des conditions dans lesquelles nous avons fait cette expérience, certains résultats peuvent sembler décevants. En effet, les conditions de transformation sont optimales et pourtant nous avons des taux d'appariement inférieur à $90\ \%$ pour l'invariance à la rotation de SURF par exemple.

Le temps de calcul est cohérent avec l'état de l'art, les descripteurs basés gradient sont plus lents que les descripteurs binaires. Une exception cependant, SURF est plus rapide que LATCH. Cela reste cohérent puisque SURF a été construit pour être rapide alors que LATCH nécessite la mesure d'une norme de Frobenius.

4.3 Un descripteur couleur déduit des conjectures de Julesz

Julesz s'est intéressé pendant toute sa carrière à étudier à quels paramètres était sensible l'œil humain lorsqu'il s'agit de discriminer des textures. Il fera plusieurs conjectures sur cette question :

> *“If two visual fields are presented simultaneously, in what properties must they differ in order to be discriminated ?”*

extrait de Julesz (1962). Julesz fut l'un des premiers chercheurs à s'intéresser à la perception de l'aspect non-uniforme d'une scène et à mettre en œuvre des expériences psychovisuelles permettant de mettre en place des conjectures sur la nature de la perception humaine des “textures”. De nombreux auteurs Bergen and Adelson (1988); Landy and Graham (2004); Rosenholtz (2014) ont suivi ses pas et utilisent encore les *principes* que Julesz a proposé au cours de sa carrière.

Nous en présenterons trois ayant conduit au développement de trois familles de descripteurs dits texture ou aspect non-uniforme. Le premier principe a donné lieu aux travaux de Haralick et al. (1973) sur la cooccurrence. Nous relierons les deux autres aux attributs qui leur ont succédé.

4.3.1 Première conjecture : cooccurrence

La première conjecture de Julesz (1965) énonce que les humains sont sensibles au contraste et aux différences locales, *i.e.* aux statistiques d'ordre un et deux.

En traduisant cette conjecture en termes d'attribut descriptif des variations locales, Haralick et al. (1973) définissent la notion de cooccurrence. Ils organisent ces cooccurrences en matrice de probabilité. La première dimension porte les valeurs de pixel et est donc reliée au moment d'ordre un. La seconde dimension est également reliée aux valeurs de pixel donc au moment d'ordre un. La probabilité jointe des deux dimensions définie par la cooccurrence donne accès indirectement au moment d'ordre deux (Unser (1986)) tel que Julesz (1962) l'a proposé.

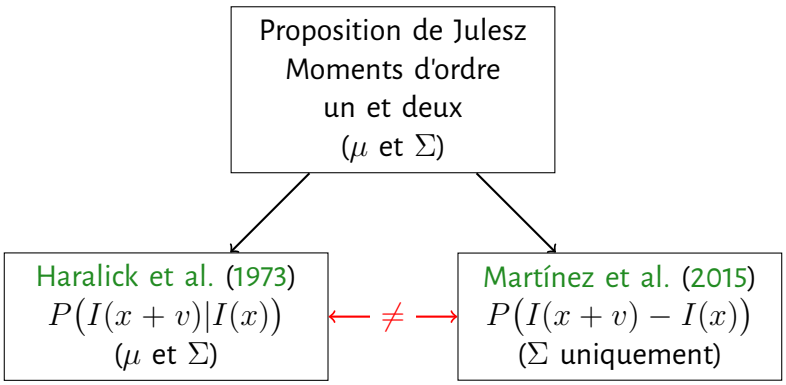


FIGURE 4.14 – **Illustration de la première conjecture de Julesz**

Dans les années 70, la mesure de similarité entre matrice était un problème complexe sur les ordinateurs de l'époque. Haralick a contourné le problème par une **mesure indirecte** de la similarité s'appuyant sur des moments statistiques extraits des matrices de cooccurrences. Il a ainsi proposé quatorze moments statistiques associés à chaque matrice dont la moyenne, la variance, l'entropie et certaines caractéristiques dépendant de la direction. Gotlieb and Kreyszig (1990) ajoutent des moments prenant en compte le contraste par exemple.

Les matrices de cooccurrences sont sensibles à la rotation puisque chaque orientation crée une matrice spécifique. L'intégration selon l'orientation des mesures de cooccurrences donnent accès au corrélogramme (fig. 4.5b). Par conséquent, celui-ci est invariant par rotation.

Plusieurs adaptations couleur ont été présentées. Drimbarean and Whelan (2001) utilisent de façon marginale les cooccurrences d'Haralick. Huang et al. (1997) proposent d'utiliser les corrélogrammes sur une quantification couleur.

Martínez et al. (2015) construisent un histogramme de différences couleur en s'affranchissant de la couleur moyenne de la zone d'intérêt. Ils reprennent ainsi la conjecture de Julesz mais perdent l'information de la valeur moyenne de l'image. Ils proposent les matrices d'occurrences de contraste couleur ($\mathrm{C_2O}$ pour *Colour Contrast Occurrences*). Elles expriment la probabilité $P(\overrightarrow{c_i c_j} = \overrightarrow{v})$ de trouver une différence couleur locale spécifique $\overrightarrow{c_i c_j} = \overrightarrow{v}$ séparée par un vecteur défini par une distance spatiale d et un angle θ. La différence couleur est définie dans l'espace couleur perceptuel $CIELAB$ afin de rester cohérent avec la vision humaine. Néanmoins l'expression n'est pas condamnée à rester dans cet espace. L'objectif est d'obtenir la probabilité des différences de couleur pour un vecteur v donné :

$$\mathrm{C_2O}(d,\theta) \; : \; \begin{cases} P\Big(\Delta\big(I(x), I(x+v)\big) = \delta\Big), \\ \quad \text{avec } v \in \mathbb{R}^2, \|v\| = d \text{ et } \widehat{v} = \theta, \\ \quad \text{et } \delta \in \mathbb{R}^3, \delta = (\Delta L, \Delta a, \Delta b), \end{cases} \tag{4.8}$$

où P correspond à la probabilité de trouver une paire de couleurs ayant une différence spatiale de v et une différence couleur de δ.

Concrètement, la matrice $\mathrm{C_2O}$ est une densité de probabilité exprimée dans un espace $(\Delta_L, \Delta_a, \Delta_b)$. Une texture stationnaire crée une seule différence couleur dont la probabilité est non nulle : $P\big(\Delta(I(x), I(x+v) = 0\big) = 1$. Quelques exemples de matrices $\mathrm{C_2O}$ sont présentés figure 4.15. Le problème ici est l'attribut extrait. La transformation d'un histogramme $3D$ en un histogramme $1D$ via une distance hélicoïdale au centre limite l'information conservée. Même si la mesure de volume extraite par Richard et al. (2016a) est plus cohérente, elle n'est pas encore complète au sens de la première conjecture de Julesz.

Le problème de toutes ces propositions est la modélisation des données qui se fait par une paramétrisation de la matrice de cooccurrence. Cela implique une perte d'information et transforme les données dans un espace où la mesure n'est pas naturelle (cf. chap. 3).

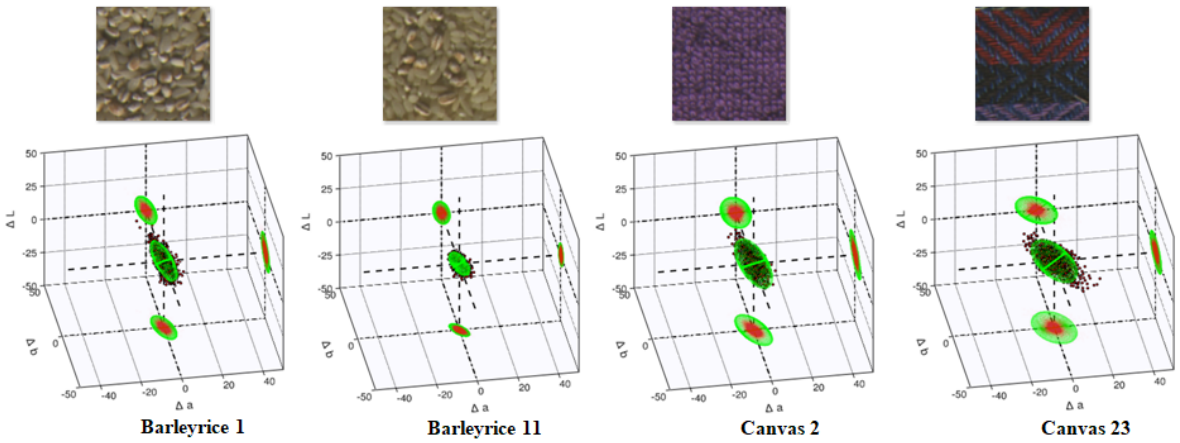
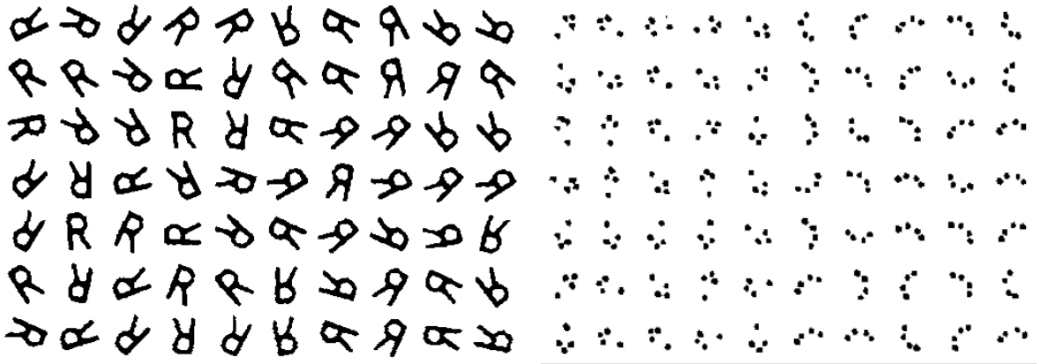


FIGURE 4.15 – **Exemples de matrices** $\mathrm{C_2O}$. Quelques textures couleur de la base *Outex* et leurs distributions $\mathrm{C_2O}$ pour $d = 1$ et $\theta = 0$.



(a) Non discriminé par l'œil humain (b) Discriminé par l'œil humain

FIGURE 4.16 – **Textures dont les statistiques d'ordre un et deux sont identiques**. Il est difficile de voir la différence d'orientation des R dans la figure 4.16a alors que la différence est évidente pour la figure 4.16b. La segmentation attendue pour l'image 4.16a est disponible en annexe H.1. Images extraites de Julész et al. (1978).

Certains auteurs, Provenzi et al. (2016) s'intéressent aux statistiques spatio-chromatiques d'ordre deux dans l'espace des transformées de Fourier.

4.3.2 Deuxième conjecture : ordres supérieurs à deux

Rapidement, Julész et al. (1978) ont montré que les statistiques d'ordre deux ne sont pas suffisantes pour toutes les surfaces non-uniformes. La figure 4.16 propose deux images présentant deux textures avec des statistiques d'ordre un et deux identiques. Sur la figure 4.16a, la différence d'orientation des R entre les deux moitiés horizontales de l'image n'apparaît que lorsqu'on la cherche. Au contraire, les deux textures se distinguent très bien sur la figure 4.16b.

Bien que la question sur l'ordre statistique nécessaire et suffisant pour décrire deux surfaces non-uniformes discernables par un observateur humain soit toujours ouverte, différentes propositions prenant en compte des ordres supérieurs à deux ont été présentées.

Les longueurs de plage, introduites par Galloway (1974), mesurent les longueurs des zones uniformes de l'image dans une direction donnée et sont conservées dans une matrice des longueurs en fonction des valeurs de pixels (fig. 4.17). Une scène non-uniforme orientée impliquera de grandes longueurs dans une direction et de faibles dans la direction perpendiculaire. Le descripteur est consti-

tué d'attributs de type moments statistiques extrait de la matrice. Chu et al. (1990) extraient d'autres attributs jouant sur le poids de l'intensité du pixel. Tang (1998) réduit la taille du descripteur en appliquant une analyse en composante principale sur la matrice de longueurs de plage. Xiaoyan et al. (2009) les appliquent dans un contexte médical. Bony and Fernandez-Maloigne (2016) proposent une adaptation couleur des longueurs de plage. Ils se placent dans l'espace $CIELAB$ et emploient le seuil de la différence juste visible comme limite des zones uniformes. Dans les matrices de longueurs de plage, l'information contient des statistiques d'ordre supérieur à deux. En effet, les matrices de longueurs de plage contiennent les différences entre plusieurs pixels dans une même direction. Celle-ci contiennent donc des informations d'ordre supérieur à deux.

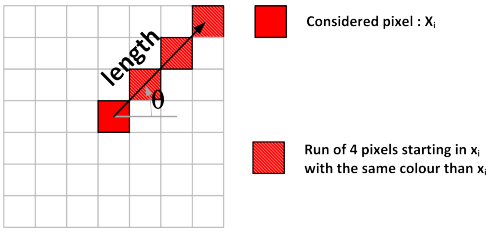


FIGURE 4.17 – **Illustration de l'approche longueur de plage**. Image extraite du cours texture de Noël Richard 2019.

Les décompositions fréquentielles et analyses en ondelettes s'intéressent aussi à des statistiques d'ordre plus grand que deux. Soulard et al. (2013) proposent une adaptation couleur de la décomposition en ondelettes prenant en compte les phases des canaux couleur lors de la transformation en quaternion. L'analyse fréquentielle vise à étudier la nature des variations entre les pixels. Dès que plusieurs fréquences sont gardées, l'information liée aux statistiques d'ordre supérieur à deux est présente.

4.3.3 Troisième conjecture : texton

Suite au travail sur l'ordre des statistiques nécessaire à la définition d'une texture pour la perception humaine, Julesz and Bergen (1983) proposent la notion de texton définie ainsi :

Textons appear to be the basic units of preattentive texture perception, when textures are viewed in a quick glance with no further effort or analysis.

Il serait l'élément fondamental à l'analyse des textures par l'œil humain dans le cas de la vision pré-attentive. La vision pré-attentive est la vision qui a lieu avant qu'une tâche spécifique soit assignée à l'utilisateur ou que le cerveau ait désigné une information comme plus importante que les autres. Le texton est donc un motif probabiliste ou déterministe qui représente une zone restreinte et "unitaire"de la région d'intérêt.

Ojala et al. (1994) reprennent cette idée de texton, pour en créer un descripteur. Ils proposent les *local binary pattern* (LBP). Chaque pixel est comparé à ses huit pixels adjacents pour la création du texton. L'ensemble des textons ainsi calculés forment un histogramme.

Guo et al. (2010) adaptent les LBP pour les rendre invariants à la rotation à l'image des corrélogrammes dans le cas des cooccurrences. Le corrélogramme, présenté par Huang et al. (1997), effectue une moyenne des cooccurrences sur l'angle de la direction. Il garde la notion de voisin mais s'affranchit de leur ordre. Il permet l'accès à une moyenne entre le centre et ses voisins. Au contraire, les LBP perdent l'information moyenne mais gardent d'un certaine manière la variance des voisins au centre. La figure 4.19 présente le parallèle qui peut être fait entre les deux approches. Elles diffèrent notamment sur l'extraction des attributs et des distances utilisées dans la littérature, appropriée ou non à la nature des données.

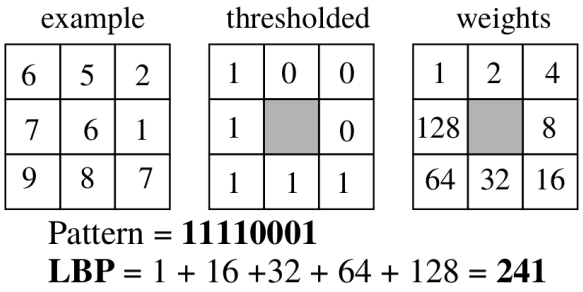


FIGURE 4.18 – **Illustration de l'approche LBP**. Image extraite de Mäenpää et al. (2000).

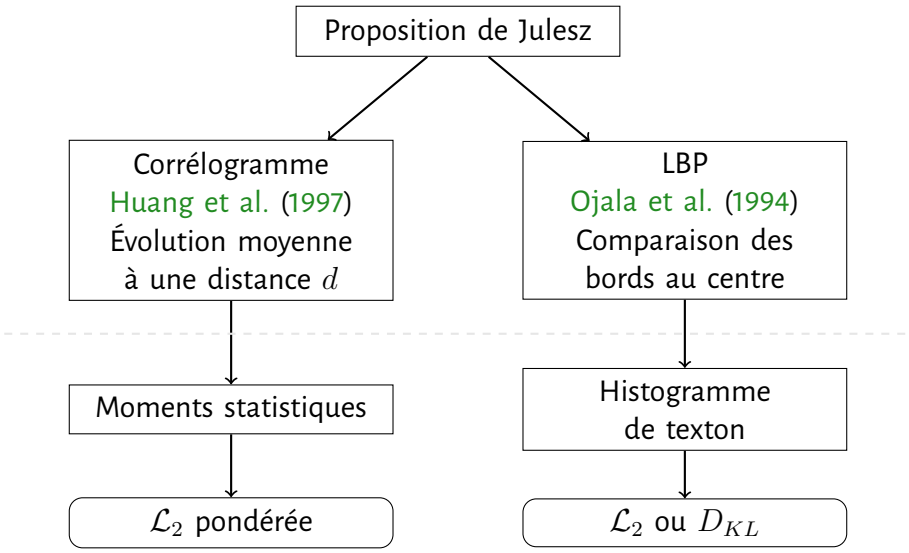


FIGURE 4.19 – **Similarités et différences entre les corrélogrammes et les LBP**. Les deux méthodes s'intéressent aux pixels autour du point d'intérêt cependant les attributs extraits diffèrent et fournissent des informations complémentaires.

La plupart des descripteurs de cette catégorie sont mono-échelles donc liés à une distance d spatiale d'analyse. Pour une adaptation multi-échelles, il faut combiner les réponses pour plusieurs distances d. Ojala et al. (2002) proposent une adaptation multi-résolution invariant à la rotation. Le corrélogramme permet aussi de regarder l'information à plusieurs échelles.

Zhang et al. (2012) reprennent l'idée des LBP mais comparent les réponses des filtres de Gabor sur l'image.

4.3.4 Relocated colour contrast occurrence

La transition aux images couleurs pose problème au niveau de l'aspect paire descripteur/mesure de similarité. Les approches marginales ou inter-canaux produisent un descripteur simple mais amplifient la difficulté en réunissant un ensemble de données. Il existe peu d'expressions vectorielles de l'aspect non-uniforme. Parmi elles, le $\mathrm{C_2O}$ produit la distribution de différences couleur de la région d'intérêt.

Richard et al. (2016a) ont montré le potentiel métrologique de l'approche $\mathrm{C_2O}$ bien que les résultats de classification n'aient pas été au niveau attendu. Nous proposons ici de reprendre le $\mathrm{C_2O}$ et de l'associer à une métrique de similarité adaptée.

Comme nous l'avons déjà dit, le $\mathrm{C_2O}$ représente une distribution de probabilité couleur. Dans le chapitre précédent, nous avons montré que face à ce type de données attribut, l'emploi de la divergence de Kulback-Leibler était non seulement justifié mais préconisé. Comme dans le chapitre

précédent, nous modélisons chaque distribution par une Gaussienne multivariée afin d'accélérer la mesure de similarité et de réduite la taille du descripteur.

Cependant, comme nous l'avons mis en avant dans le schéma 4.14, le descripteur $\mathrm{C_2O}$ ne prend pas en compte la moyenne de la région d'intérêt. Nous proposons donc le descripteur $\mathrm{RC_2O}$ pour *Relocated Colour Contrast Occurrences*. Ses attributs se composent de la moyenne de la région d'intérêt et de la matrice de variance-covariance de la distribution $\mathrm{C_2O}$:

$$\mathrm{RC_2O}(d,\theta) \ : \begin{cases} P\Big(\Delta\big(I(x),I(x+v)\big)+\mathbb{E}(I(x))=\delta\Big), \\ \text{avec } v\in\mathbb{R}^2, \|v\|=d \text{ et } \widehat{v}=\theta, \\ \text{et } \delta\in\mathbb{R}^3, \end{cases} \tag{4.9}$$

où P correspond à la probabilité de trouver une paire de couleurs ayant une différence spatiale de v et une différence couleur de δ. La couleur moyenne de la région d'intérêt est μ. Nous retrouvons donc tous les éléments nécessaires pour reconnaître les statistiques d'ordres un et deux proposées par la première conjecture de Julesz (fig. 4.20).

Sur la figure 4.20, nous observons que le $\mathrm{C_2O}$ n'est pas au même niveau que les deux autres attributs puisque ce dernier perd l'information de la valeur moyenne. Les méthodes employées pour extraire le descripteur diffèrent pour les trois méthodes. La modélisation par une Loi Gaussienne que nous proposons dans ce chapitre n'étant pas obligatoire. Il est cependant indispensable de passer par une quantification pour le calcul des cooccurrences sous peine d'obtenir des matrices creuses qu'ils sera difficile de différencier. En effet, les matrices de cooccurrences sont de taille $nbC \times nbC$ où nbC correspond au nombre de valeurs que peut prendre le pixel. L'histogramme obtenu pour le $\mathrm{C_2O}$ ou le $\mathrm{RC_2O}$ est, au maximum, de la taille du nombre de pixels ($+n$ pour le $\mathrm{RC_2O}$, n nombre de canaux).

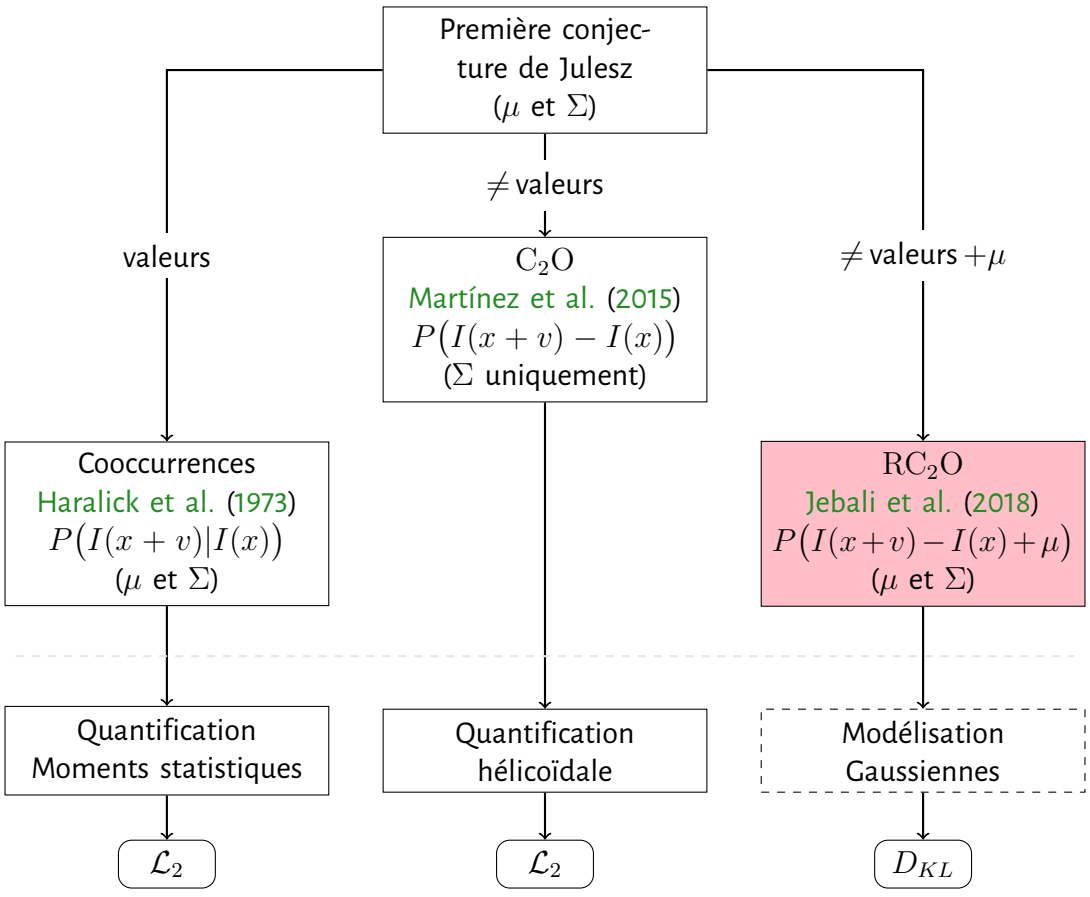
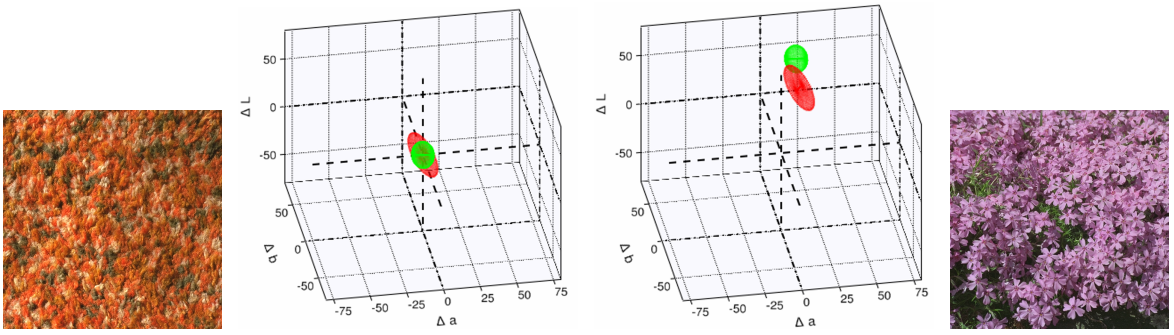


FIGURE 4.20 – **Comparaison des approches s'inspirant de la première conjecture de Julesz**. Les cooccurrences comme le $\mathrm{RC_2O}$ prennent en compte les moments d'ordre un et deux. Ils diffèrent dans l'extraction de leur attributs et la distance/similarité utilisée.



(a) *Fabric* (b) $\mathrm{C_2O}$ (c) $\mathrm{RC_2O}$ (d) *Flowers*

FIGURE 4.21 – **Différence entre les distributions C_2O et $\mathrm{RC_2O}$.** Les ellipsoïdes modélisent la distribution C_2O et $\mathrm{RC_2O}$ pour $d = 1$ et $\theta = 0$ pour les images *Fabric* (ellipse verte) et *Flowers* (ellipse rouge) de la base *VisTex*. Les distributions de la proposition $\mathrm{RC_2O}$ induisent une plus grande différence.

La matrice $\mathrm{C_2O}$ décrit les variations internes d'une texture. Quand la texture est stationnaire, la distribution $\mathrm{C_2O}$ associée est centrée à l'origine de l'espace de représentation. Pour les applications en Contrôle Qualité, toutes les textures sont similaires et possèdent une moyenne commune. Cependant, nous cherchons à adapter le propos à la classification de textures dans des bases de données. Dans ce cas, la moyenne couleur de la texture peut varier sur chaque image. De nombreuses études sur la discrimination de texture montrent que la moyenne est une caractéristique importante à la discrimination (Singh and Hemachandran (2012)).

La figure 4.21 présente la similarité évaluée en utilisant les distributions $\mathrm{C_2O}$ (fig. 4.21b) ou les distributions $\mathrm{RC_2O}$ (fig. 4.21c). Nous observons que la seconde solution embarquant la moyenne implique une différence plus grande que celle basée uniquement sur les différences de variations locales.

4.4 Preuve de l'approche

Pour comparer les capacités de caractérisation/discrimination des descripteurs, Mikolajczyk et al. (2005); Mikolajczyk and Schmid (2005) ont proposé une méthode pour comparer les détecteurs de point d'intérêt et les descripteurs. Dans ces articles, les auteurs introduisent une nouvelle base de données avec une vérité terrain. Pour comparer les descripteurs, Mikolajczyk and Schmid (2005) comptent le nombre d'appariements corrects sur le nombre total d'appariements. Miksik and Mikolajczyk (2012) ont mis cette étude à jour avec de nouveaux détecteurs et descripteurs testés ensemble pour déterminer l'impact des associations et comparer de nouveaux descripteurs binaires.

D'autres comparatifs existent pour des tâches dédiées : Gauglitz et al. (2011) s'intéressent au suivi d'objets, Mandal et al. (2015) se spécialisent sur la reconnaissance faciale, Hietanen et al. (2015) comparent détecteurs et descripteurs pour la classification d'objets.

Nous nous intéressons depuis le début de ce manuscrit à l'utilisation de la couleur. Même si le nombre d'articles spécifiques à la couleur est moindre dans ce domaine, plusieurs évaluations de descripteurs couleur ont comparé les performances relatives des descripteurs dans ce contexte : van de Sande et al. (2010); Burghouts and Geusebroek (2009); Malekabadi et al. (2018).

Dans la suite de cette section, nous allons comparer les descripteurs utilisés et produits avec le *Local Binary Pattern*. Les approches $\mathrm{C_2O}$ et $\mathrm{RC_2O}$ résolvant les limites des approches basées sur des cooccurrences, nous avons choisi de ne pas inclure cette approche afin d'alléger les analyses de résultats.

La problématique posée pour les différentes comparaisons des descripteurs de l'aspect non-uniforme est au-delà du choix de la paire{descripteur, mesure de similarité} celle du choix des bases de données à utiliser. Par nature, l'aspect non-uniforme se caractérise par une distribution spatio-chromatique particulière. Nous attendons donc qu'une base d'évaluation s'assure à la fois d'explorer toutes les fréquences spatiales possibles et toutes les distributions couleur possibles. Cette contrainte n'est pas gérée avec les bases de données réelles de la littérature. Elle ne l'est pas non plus par la base de données de Gaussiennes $3D$ utilisée aux deux chapitres précédents puisqu'elle n'est pas construite pour contenir une notion de spatialité absolument nécessaire pour évaluer un descripteur d'aspect non-uniforme. Nous commencerons donc nos expérimentations avec une base de synthèse. C'est une base de données de textures fractales couleur.

4.4.1 Base de données artificielles

Le cœur de la proposition est l'exploitation d'une base d'images fractales couleur. L'intérêt des images fractales est tout d'abord d'assurer la présence d'énergie à toutes les fréquences spatiales possibles dans l'image ; grâce au paramètre de réglage de la complexité : le coefficient de Hurst. Celui-ci permet de gérer les rapports d'amplitude entre ces fréquences suivant une loi :

$$\frac{1}{f^{2H+1}}, \tag{4.10}$$

où f est la fréquence et H le coefficient de Hurst. Plus d'information sur la génération est fournie en annexe I. Deuxièmement, l'aspect génération aléatoire permet de contrôler la plage de variations chromatiques explorées à chaque nouvelle synthèse. Ainsi les contraintes spécifiées peuvent être facilement résolues.

La seconde question naturelle dans la mise en place d'une base d'images de synthèse est celle du choix de l'espace couleur de génération. La seule bonne réponse serait l'espace physique de génération de la scène,mais cette solution n'est pas encore accessible. Les images seront générées dans $CIELAB$. L'idée ici est d'anticiper sur des études psychophysiques avec des images dont les différences couleur générées l'ont été dans une espace cohérent avec la perception de ces différences.

Les attributs seront établis dans trois espaces différents, l'espace RGB_{FV} du capteur (l'espace dans lequel la mesure couleur est effectuée), l'espace $CIERGB$ qui correspond à l'espace de référence en RGB (normalisé par la CIE en 1931) et enfin l'espace $CIELAB$. Il est attendu que les résultats obtenus dans $CIELAB$ soient supérieurs à ceux obtenus dans les deux autres espaces.

Dans le cas du LBP et son adaptation à la couleur, nous suivons l'approche développée par Porebski et al. (2013) exploitant une technique de type $CCMA$ (*Cross Channel Marginal Approach*). Cette technique n'a de sens que dans l'espace $CIERGB$. Nous calculerons aussi les LBP en marginal (méthode CMA) dans les espaces RGB_{FV} et $CIELAB$.

Comparons la dimension des signatures entre les descripteurs étudiés. Les LBP extraient un histogramme de 256 classes par comparaison étudiée (3 canaux et 6 canaux croisés lorsque ceux-ci sont calculés). La signature du LBP est donc 3 ou 9 histogrammes accolés. Ces histogrammes ne peuvent être considérés comme un histogramme nD dense puisqu'ils conservent les réponses binaires de chaque pixels. Ils ne peuvent donc être modélisés pour diminuer la taille de la signature. Les attributs des descripteurs $\mathrm{C_2O}$ et $\mathrm{RC_2O}$ sont des histogrammes de différences couleur de nbC nombre de classes par canal. Leurs tailles sont donc de $3 \times nbC$. Cependant, ces histogrammes sont denses et peuvent être modélisés par une loi Gaussienne comme nous le faisons. La taille du descripteur devient alors $9 + 3$ pour le $\mathrm{C_2O}$ et $9 + 3 + 3$ pour le $\mathrm{RC_2O}$. Il faut neuf paramètres pour caractériser la matrice de variance, les trois valeurs suivantes correspondent à la moyenne de la matrice de différences (nulle si la texture est stationnaire). Dans le cas du $\mathrm{RC_2O}$, nous ajoutons la moyenne couleur de l'image. Nous obtenons donc une taille de 12 ou 15 éléments ce qui est beaucoup plus réduit que pour les LBP.

4.4.1.1 Protocole des expérimentations

Nous disposons d'une base d'images fractales dont les paramètres de génération sont :

— μ : la valeur moyenne recherchée ;

— Σ : la matrice de variance-covariance induisant les corrélations inter-canaux ;

— H : le coefficient de Hurst expliquant la complexité relative de l'image.

Plus H tend vers zéro, plus l'image est complexe.

Rappelons que l'estimation de la dimension fractale n'est pas le propos de l'étude. Ici, nous nous intéressons à la recherche d'image similaire pour une même échelle spatiale d'analyse. Dans cette expérimentation, chaque image générée est un échantillon unique. Il n'existe donc pas deux images identiques, elles sont tout au plus similaires avec un même jeu de paramètres et dépendent du processus aléatoire de génération.

Quatre expérimentations sont proposées. La première expérience vise à discriminer un changement de valeur moyenne. Considérant des images de même Σ et H, nous cherchons à vérifier que les attributs permettent de discriminer des images de valeurs moyennes différentes. Les attributs $\mathrm{C_2O}$ et LBP ne travaillent qu'avec des mesures de différences, il est attendu des limites dans leur performances.

La deuxième expérience vise à discriminer un changement de matrice de covariance. Dans ce test, toutes les images ont une même valeur moyenne couleur μ cible et la même complexité H. Puisque la valeur moyenne n'est plus un facteur différenciant, le $\mathrm{C_2O}$ est attendu avec de meilleures performances à niveau égal avec le $\mathrm{RC_2O}$.

La troisième expérience vise à discriminer un changement de coefficient de Hurst. Dans cette expérimentation, nous fixons les valeurs de μ et Σ et nous cherchons à discriminer les images de complexité différentes. Pour les mêmes raisons que précédemment, nous attendons que $\mathrm{C_2O}$ et $\mathrm{RC_2O}$ donnent de bon résultats.

Dans la dernière expérimentation, nous cherchons les images similaires dans la base, *i.e.* les images ayant le même μ, Σ et H. Nous disposons de cinq images pour chaque configuration présentée.

La base possède 400 images, 100 par moyenne cible. La moitié des images a été générée avec une matrice de variance Σ_k et l'autre moitié avec une variance Σ_l. Les matrices peuvent varier d'un μ_i à l'autre. 5 images par μ_i et Σ_k ont été générées avec 10 coefficients de Hurst différents.

Nous avons effectué les calculs sur quatre angles : $0°$, $45°$, $90°$ et$135°$; et quatre distances : 1, 3, 5 et 7 pixels. Les images ont été générées de façon isotrope. Nous devrions donc retrouver les mêmes propriétés sur chacun des angles mesurés. La figure 4.25 le montre et c'est pourquoi nous présenterons, par la suite uniquement la moyenne des résultats sur les quatre angles. Les variations sur les distances étant faibles ou négligeables, nous réalisons aussi une moyenne sur les quatre distances. Le LBP est calculé sur ses huit voisins connexes donc à une distance $d = 1$ comme il a été présenté dans sa première version par Ojala et al. (1994).

Critère d'évaluation

Nous classons les données grâce à une mesure des k plus proches voisins. Cela permet de mesurer la précision pour un k donné :

$$P@k = \frac{\#\mathcal{TP}}{k}, \tag{4.11}$$

où $\mathcal{TP}$ correspond à l'ensemble des k-premières images bien classées. Puisque chaque classe possède 5 images, nous mesurerons le $P@5$. Cette mesure ne prend pas en compte le classement des bonnes images après les k sélectionnés. La précision moyenne (AveP pour *average precision*) permet

TABLE 4.5 – **Récapitulatif des paramètres de l'expérience**. Le tableau indique le nombre de cas différents pour chaque variable.

	Base de données fractales
H	10 valeurs de coefficient de Hurst
μ	4 moyennes
Σ	2 matrices de variance

	Descripteurs		
	$\mathrm{C_2O}$	LBP	$\mathrm{RC_2O}$
Espaces couleur	3	3	3
Distances	4	1	4
Angles	4	$\varnothing$	4

de prendre en compte toutes les images de la classe, nous calculons aussi cette mesure. Pour une base de taille n ayant l éléments dans chaque classe, la précision moyenne se calcule ainsi :

$$AveP = \sum_{k \in \mathcal{T}_k} P@k, \tag{4.12}$$

où $\mathcal{T}_k$ est l'ensemble des valeurs k dont la dernière entrée est une image de la classe cherchée.

Ces deux mesures qualifient la qualité de classification d'une seule image de la base. Les mesures généralement utilisées en classification sont les moyennes respectives ($MP@5$ et MAP) de ces deux mesures sur l'ensemble des images de la base.

4.4.1.2 Discrimination selon la couleur moyenne

Ici, nous fixons le coefficient de Hurst et la matrice de covariance utilisée pour la génération des images fractales. L'objectif est donc de déterminer la séparabilité des données lorsque la moyenne de l'image varie.

La table 4.6 présente les résultats de classification. Les descripteurs $\mathrm{RC_2O}$ et LBP donnent 100% de bonne classification. La paire descripteur $\mathrm{C_2O}$/divergence de Kullback-Leibler ne prend pas en compte la moyenne de l'image dans la mesure de similarité entre images. Cela explique le résultat plus faible qu'obtient ce descripteur lorsque l'objectif de classification est la discrimination des moyennes des images.

Pour les autres tests, la moyenne est fixée pour les mesures de classification éliminant ainsi le défaut du descripteur $\mathrm{C_2O}$.

Cette question semble triviale, pourtant elle ramène à la question de la mesure de similarité entre deux surfaces couleurs. Doit-on disposer d'une question unique et d'une métrique de similarité associée ou bien de deux descripteurs (l'un couleur et l'autre texture), et dans ce cas, perdre la validité métrologique de la mesure de similarité pour la remplacer par une mesure empirique.

Le $\mathrm{RC_2O}$ a été construit pour intégrer naturellement les deux aspects du problème. Le $\mathrm{C_2O}$ n'intègre pas la référence à la moyenne, de même que le LBP.

Les résultats obtenus correspondent bien aux attentes, avec un LBP qui résout parfaitement le problème posé. L'explication tient dans le fait que le nombre d'images exploitées est trop réduit avec un nombre de configuration trop petit. Dès lors, les mesures croisées permettent d'identifier les configuration choisies.

TABLE 4.6 – **Classification sur la base de données fractales avec H et Σ constants**. Le taux de bonne classification est parfait pour $\mathrm{RC_2O}$ et LBP, un peu plus faible pour le $\mathrm{C_2O}$ puisque la moyenne de l'image n'est pas prise en compte lors de la mesure de similarité.

		$MP@5$	MAP
$\mathrm{C_2O}$	$CIELAB$	97.3 %	98.7 %
	RGB_{FV}	97.2 %	98.6 %
	$CIERGB$	97.2 %	98.6 %
LBP	$CMA\text{-}CIELAB$	79.9 %	88.9 %
	$CMA\text{-}RGB_{FV}$	69.2 %	81.7 %
	$CCMA\text{-}RGB$	100 %	100 %
$\mathrm{RC_2O}$	$CIELAB$	100 %	100 %
	RGB_{FV}	100 %	100 %
	$CIERGB$	100 %	100 %

TABLE 4.7 – **Classification sur la base de données fractales avec μ et H constants**. Cette fois-ci le $\mathrm{C_2O}$ donne de très bons résultats. Au contraire, le $\mathrm{RC_2O}$ chute plus qu'attendu à cause d'une variation non négligeable des valeurs moyennes des images.

		$MP@5$	MAP
$\mathrm{C_2O}$	$CIELAB$	86.9 %	94.9 %
	RGB_{FV}	90.2 %	95.1 %
	$CIERGB$	90.2 %	95.1 %
LBP	$CMA\text{-}CIELAB$	65.9 %	79.8 %
	$CMA\text{-}RGB_{FV}$	63.5 %	77.9 %
	$CCMA\text{-}RGB$	76.2 %	86.4 %
$\mathrm{RC_2O}$	$CIELAB$	68.7 %	81.5 %
	RGB_{FV}	70.4 %	82.9 %
	$CIERGB$	70.4 %	82.9 %

4.4.1.3 Discrimination selon la matrice de covariance

Ce test conserve les images de même moyenne et de même coefficient de Hurst. L'objectif est de mesurer le pouvoir discriminant des couples $\{$Descripteur;Similarité$\}$ lorsque la covariance des images générées varie.

La table 4.7 présente la moyenne des résultats de classification sur toutes les images. Les meilleurs taux de bonne classification ($MP@5 = 90.6\%$ et $MAP = 95.3\%$) sont obtenus avec le descripteur $\mathrm{C_2O}$ prouvant ainsi son pouvoir discriminant lorsque la moyenne est fixe.

Le descripteur LBP en marginal sur les espaces orthogonaux ne permet pas de distinguer les différentes matrices de covariances. Si le descripteur de type $CCMA$ fait des comparaisons intra-canaux et inter-canaux, cela ne lui permet pas plus de distinguer des images générées avec différentes matrices de covariance. Cependant ces résultats sont meilleurs que dans les deux autres espaces. L'utilisation des informations inter-canaux améliore le taux de bonne classification. Cela serait-il le cas pour les espaces $CIELAB$ et RGB_{FV} qui sont sensés être orthogonaux ? Le descripteur de type LBP n'est pas robuste dans le cas de la discrimination ente matrice de covariance.

Nous avons présenté le descripteur $\mathrm{RC_2O}$ comme plus abouti que le $\mathrm{C_2O}$. Cependant, le taux de bonne classification du $\mathrm{RC_2O}$ est inférieur à celui du LBP pour ce test. Les images sont générées dans l'espace des transformées de Fourier quaternioniques depuis l'espace $CIELAB$ avec une moyenne cible. Nous avons vérifié après création, les moyennes, et leurs écarts-type, obtenus sur les 100 images par moyennes μ_i cible. Les résultats sont récapitulés dans la table 4.8. L'écart-type corres-

TABLE 4.8 – **Vérification des moyennes et écarts-type par μ générés**. L'écart-type sur les 100 images générées implique un ΔE qui, parfois, excède le seuil de différence juste visible.

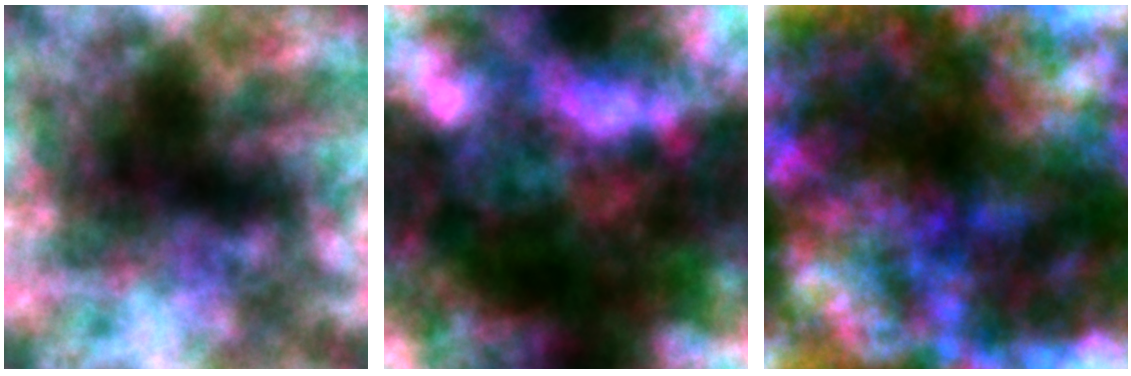
	Moyenne mesurée			Écart-type			ΔE
	L	a	b	L	a	b	
μ_1	52.2	5.6	-8.6	1.8	2.1	0.7	2.8
μ_2	35.1	63.4	-9.4	2.7	2.2	0.8	3.6
μ_3	48.3	0.2	40.5	6.4	1.3	3.0	7.2
μ_4	31.4	11.3	-13.0	3.3	3.7	4.4	6.6

pondant à la moyenne de l'écart à la moyenne, nous prenons sa norme $\mathcal{L}_2$ pour représenter un ΔE moyen entre les images générées. Ce ΔE varie entre 2.0 et 6.8. Si la première valeur est en dessous du seuil de différence visible par l'œil humain, ce n'est pas le cas des trois autres valeurs supérieures à 3. De tels écarts entre les moyennes d'un même μ_i expliquent les mauvais résultats du descripteur $\mathrm{RC}_2\mathrm{O}$ qui prend en compte la moyenne de l'image. À la différence du LBP, il est plus sensible à ces variations de moyennes.

La figure 4.22 présente deux images appartenant à la même classe (fig. 4.22b et 4.22a) et une image, dont μ et H sont identiques, mais avec une valeur de Σ différente (fig. 4.22c). L'image centrale (fig. 4.22b) sert de référence dans cette comparaison. Sous les deux autres figures (fig. 4.22a et 4.22c), nous avons indiqué la mesure du ΔE avec l'image de référence. Bien que les deux images les plus à droite aient été générées avec des matrices de covariances différentes (et appartiennent donc à deux classes différentes), elles semblent visuellement plus proches que les deux images de gauche. Cela montre bien l'importance de prendre en compte la moyenne dans la mesure de similarité. Cela démontre aussi les limites de la base de données fractales actuelle.

4.4.1.4 Discrimination selon le coefficient de Hurst

Continuons l'analyse des moyennes des images de la base de données mais du point de vue du coefficient de Hurst. La figure 4.23 présente deux images appartenant à la même classe (fig. 4.23b et 4.23a) et une image de μ et Σ identiques mais avec une valeur de H différente (fig. 4.23c). L'image



(a) μ_1 et Σ_1, $\mu = (45.7, 3.7, -8.8)$, $\Delta E(4.22b, 4.22a) = 18.0$.

(b) Image de Référence : $H = 0.9$, $\mu = (29.3, 10.2, -12.0)$.

(c) μ_1 et Σ_2, $\mu = (31.3, 9.0, -10.3)$, $\Delta E(4.22b, 4.22c) = 2.9$.

FIGURE 4.22 – **Incohérence des moyennes générées** en $CIELAB$ par rapport à la théorie. Chacune de ces images a été générée avec la même moyenne cible μ_1. La moyenne obtenue en pratique n'est pas toujours proche de cette valeur, impliquant des ΔE entre les moyennes des images non négligeables.

TABLE 4.9 – **Classification sur la base de données fractales avec μ et Σ constants**. La complexité de l'image semble être plus difficile à différencier. Nous étudions ici des descripteurs mono-échelles ce qui explique la faible discrimination de la complexité.

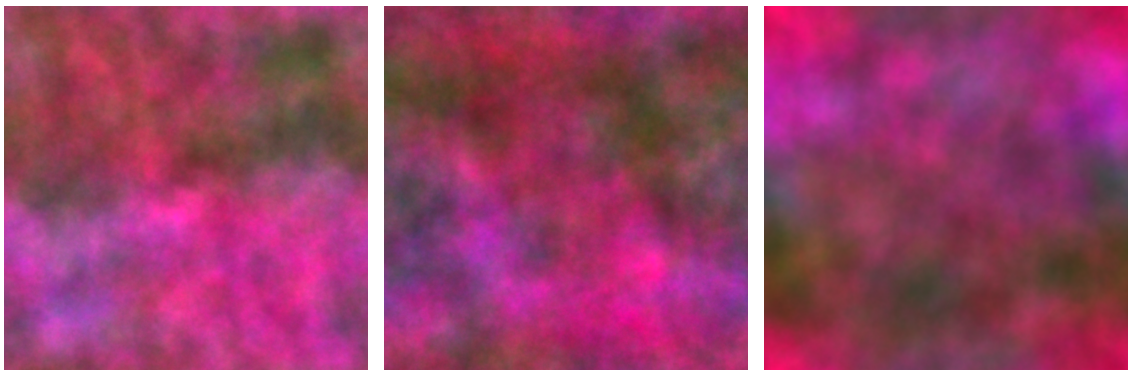
		$MP@5$	MAP
C_2O	$CIELAB$	73.9 %	83.5 %
	RGB_{FV}	71.4 %	81.3 %
	$CIERGB$	71.4 %	81.3 %
LBP	$CMA\text{-}CIELAB$	93.0 %	95.9 %
	$CMA\text{-}RGB_{FV}$	96.0 %	97.8 %
	$CCMA\text{-}RGB$	68.7 %	79.0 %
RC_2O	$CIELAB$	47.0 %	57.0 %
	RGB_{FV}	45.8 %	55.7 %
	$CIERGB$	45.8 %	55.7 %

centrale (fig. 4.23b) sert de référence dans cette comparaison. Sous les deux autres figures (fig. 4.23a et 4.23c), nous avons indiqué la mesure du ΔE avec l'image de référence. Même si l'image de la figure 4.23a à un ΔE supérieur à celle de la figure 4.23c, la complexité de l'image à tendance à nous faire associer les deux images de même coefficient de Hurst. Nous montrons avec cet exemple l'importance des statistiques d'ordre deux qui amènent une information sur l'arrangement spatial des données. Nous pouvons aussi en déduire que le descripteur RC_2O ne pourra pas donner de résultat probant puisque la variation de moyenne intra-classe est identique ou supérieure à la variation inter-classe.

Cette expérience fait varier uniquement le coefficient de Hurst et laisse la moyenne et la covariance visées fixes. Avec l'hypothèse que la complexité d'une image est liée à la fréquence des variations couleur dans l'espace. Plus les fréquences sont hautes, plus l'image est complexe. Nous cherchons donc à discriminer la complexité de l'image liée au H. Plus H est faible, plus l'image est complexe.

La table 4.9 récapitule les moyennes des résultats de classification pour μ et Σ constants. Comme attendu, le descripteur RC_2O donne une bonne classification aux alentours de 50%. Les descripteurs C_2O et LBP ont des résultats très proches d'environ 70% pour la précision à 5 et 80% pour la MAP.

La figure 4.24 présente les moyennes de la précision à 5 et de la précision moyenne pour les trois

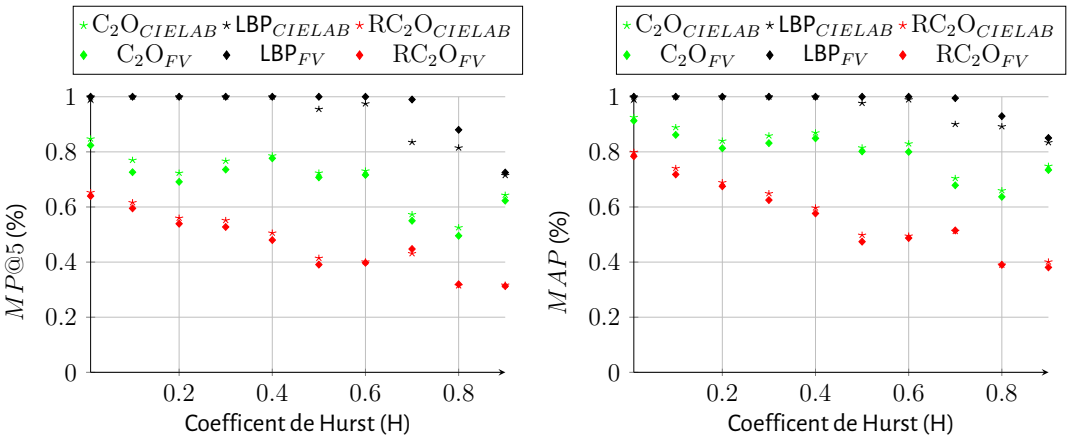


(a) μ_1, Σ_1 et $H = 0.6$, $\mu = (39.5, 67.0, -9.1)$, $\Delta E(4.23b, 4.23a) = 7.1$.

(b) μ_1, Σ_1 et $H = 0.6$, $\mu = (34.2, 62.5, -9.1)$.

(c) μ_1, Σ_1 et $H = 0.9$, $\mu = (32.7, 61.0, -8.6)$, $\Delta E(4.23b, 4.23c) = 2.1$.

FIGURE 4.23 – **Incohérence des moyennes générées** en $CIELAB$ par rapport à la théorie. Chacune de ces images a été générée avec un $\mu = (50, 50, 0)$ fixe. La moyenne obtenue n'est pas toujours proche de cette valeur, impliquant des ΔE entre les moyennes des images non négligeables.



(a) $P@5$ moyen par coefficient de Hurst (b) MAP par coefficient de Hurst

FIGURE 4.24 – **Variations des précisions avec le coefficient de Hurst**. Le taux de bonne classification chute avec l'augmentation de H, *i.e.* avec la diminution de la complexité. Le LBP qui est calculé à une distance de 1, est très performant sur les H faibles (0.01 et 0.1). Le C_2O lui est performant sur des échelles plus hautes puisque ces distances de calcul sont 3, 5 et 7.

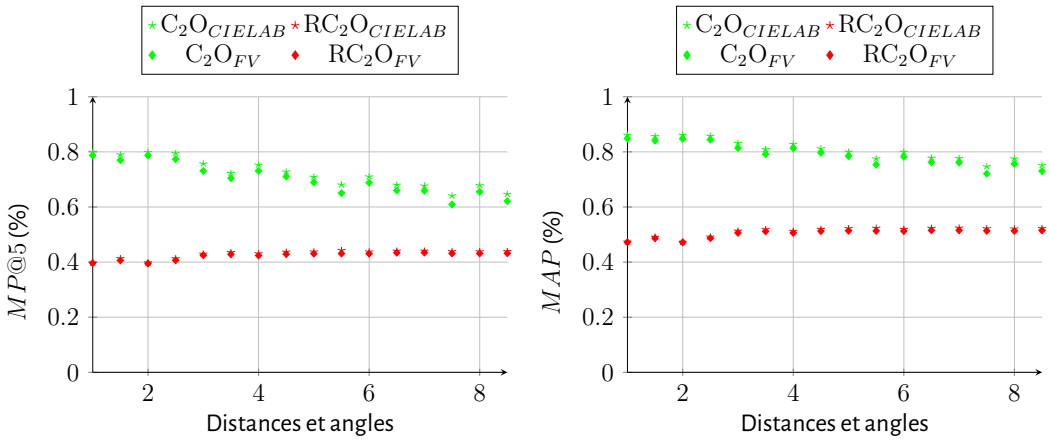
TABLE 4.10 – **Classification sur la base de données fractales**. Le C_2O donne les meilleurs résultats et les LBP en $CCMA$ chutent semblant indiquer les limites de cette méthode pour une telle base. La variance des moyennes inter-classe justifie le faible score du RC_2O.

		$MP@5$	MAP
C_2O	$CIELAB$	72.3 %	80.8 %
	RGB_{FV}	70.2 %	79.1 %
	$CIERGB$	70.2 %	79.1 %
LBP	$CMA\text{-}CIELAB$	55.0 %	67.1 %
	$CMA\text{-}RGB_{FV}$	43.3 %	54.2 %
	$CCMA\text{-}RGB$	60.5 %	70.2 %
RC_2O	$CIELAB$	42.9 %	51.1 %
	RGB_{FV}	42.3 %	50.4 %
	$CIERGB$	42.3 %	50.4 %

descripteurs en fonction du coefficient de Hurst. Nous remarquons une tendance générale de diminution de bonne classification avec l'augmentation de H, c'est-à-dire avec la diminution de la complexité. Or plus la complexité diminue, plus l'échelle permettant de voir les variations représentatives de l'aspect non-uniforme devra être grande. Or les LBP ne s'intéressent qu'aux huit voisins connexes et les distances choisies pour C_2O et RC_2O sont au maximum de 7, soit inférieures à la distance nécessaire pour discriminer les complexités plus faibles. Choisir des distances suffisamment grandes induirait le problème inverse, les complexités fortes ne seraient plus discriminées. Cela montre la nécessité d'aborder certains problèmes de façon multi-échelles. Ce qui est typiquement le cas pour une base d'images fractales.

4.4.1.5 Classification sur toutes les images

Maintenant que nous avons analysé le comportement de classification selon la variation de chaque paramètre un à un, intéressons nous à la base complète. Nous ne pouvons pas nous attendre à de meilleurs résultats que les plus faibles obtenus puisque les problèmes rencontrés par élément



(a) $P@5$ moyen pour les distances et angles choisis (b) MAP moyen pour les distances et angles choisis

FIGURE 4.25 – **Variations du C_2O et RC_2O en fonction de la distance et de l'angle**. Les quatre premières valeurs correspondent aux quatre angles pour $d = 1$, puis les quatre suivantes à $d = 3$, etc. L'écart de mesure entre les angles est très faible prouvant ainsi que les textures fractales sont bien générées de manière isotrope. Les variations sont un peu plus marquées sur les distances mais restent faibles.

variant vont s'additionner. La table 4.10 présente les résultats de classification sur toute la base de données fractales. Le descripteur LBP présente la plus grosse diminution de bonne classification (perte de 8% environ) par rapport aux résultats de la table 4.9. Le RC_2O chute aussi de 3.5% pour la $MP@5$ et de 5.1% pour le MAP. Le C_2O est moins impacté et chute de seulement 2% au maximum. Il donne donc ici les meilleurs résultats de classification.

Notons que les deux espaces couleurs RGB utilisés ne présentent aucune différence de résultats. Les espaces $CIERGB$ et RGB_{FV} ne sont séparés que par une transformation linéaire, ce qui explique que la mesure de divergence associée à la moyenne et la covariance n'induise pas de différence. Les résultats dans l'espace $CIELAB$ sont meilleurs sauf dans le cas où l'objectif est la séparation des images de différentes moyennes. Cet espace varie entre -100 et 100 alors que les espaces RGB sont conservés entre 0 et 1. L'écart de moyenne est donc plus important dans l'espace $CIELAB$ induisant des divergences plus fortes et donc une chute du taux de bonne classification.

En conclusion, en souhaitant montrer l'apport du *Relocated Colour Contrast Occurrences* sur cette base de données, nous avons montré les limites de la base sans pouvoir prouver la valeur de l'approche proposée. Dans la sous-section suivante, nous étudions cette proposition sur des bases de données réelles.

4.4.2 Classification de textures

4.4.2.1 Protocole de l'expérimentation

Nous comparons notre méthode sur quatre bases de données difficiles : *VisTex*, *STex*, *Outex TC13* et *KTH-TIPS2-b* (Table 4.11).

Chacune des ces bases présente différentes classes d'images. La moitié des images de chaque classe sert pour l'apprentissage, l'autre sert à mesurer le taux de bonne classification. L'image à classifier est affectée à la même classe que son plus proche voisin de l'ensemble apprentissage (donc nécessairement pas lui-même). Cela permet de prendre en compte la variabilité des images du monde naturel (typiquement la végétation).

La base de données *Outex* est une large collection de 68 classes. Comme l'ont fait Arvis et al.

TABLE 4.11 – **Bases utilisées pour les expériences**. Les références initiales et les entrées bibliographiques qui y font appel et présentant les meilleurs résultats pour chacune d'entre elles.

Bases	**Ire référence**	**Référence de Classification**
VisTex	MIT (1995)	Pham et al. (2017),Richard et al. (2016b), Maliani et al. (2014),Hauta-Kasari et al. (1996), Martínez et al. (2015),Alvarez and Vanrell (2012) Byeon et al. (2014),Mäenpää and Pietikäinen (2004).
STex	Kwitt and Meerwald	Pham et al. (2017),Richard et al. (2016b), Mäenpää and Pietikäinen (2004),Maliani et al. (2014), Hauta-Kasari et al. (1996),Martínez et al. (2015).
Outex TC13	Yliopisto	Pham et al. (2017),Richard et al. (2016b), Byeon et al. (2014),Sandid and Douik (2016), Maliani et al. (2014),Mäenpää and Pietikäinen (2004), Hauta-Kasari et al. (1996),Martínez et al. (2015). Alvarez and Vanrell (2012),Florindo et al. (2016b).
KTH-TIPS2-b	Caputo et al. (2010)	Sandid and Douik (2016),Florindo et al. (2016b), Song et al. (2016),Khan et al. (2015),Nguyen et al. (2016).

TABLE 4.12 – **Divergence D_{KL} inter et intra-classe pour les images de la base *Outex***.

OUTEX	Barleyrice 1	Barleyrice 11	Canvas 2	Canvas 23
Barleyrice 1	**0**	2.052	50.634	39.782
Barleyrice 11	2.052	**0**	90.435	72.330
Canvas 2	50.634	90.435	**0**	17.233
Canvas 23	39.782	72.330	17.233	**0**

(2004), chaque image est découpée en 16 sous images de 128×128 pixels dont la moitié est utilisée pour l'apprentissage et l'autre moitié pour la classification.

La figure 4.15 montre quelques textures de la base *Outex* ainsi que leurs distributions $\mathrm{C_2O}$. Visuellement, les deux images de la classe *Barleyrice*, *Barleyrice 1* et *Barleyrice 11* sont très proches. Cette similarité se retrouve dans leurs distributions $\mathrm{C_2O}$. C'est confirmé par la divergence entre les deux distributions : $D_{KL} = 2.052$ (table 4.12). Les divergences sont plus grandes avec les images de la classe *Canvas*, par exemple pour *Canvas 2*, elle est de $D_{KL} = 90.435$.

La base de données *VisTex* MIT (1995) est construite de 54 images qui ont une taille initiale de 512×512 pixels. Les images sont prises sans éclairage normalisé. Comme précédemment, elles seront découpées en 16 sous-images, obtenant ainsi 864 sous images.

La figure 4.26 présente quelques classes de cette base avec leurs distributions $\mathrm{C_2O}$, la table 4.13 montre les divergences D_{KL} associées. La distance intra-classe est toujours nulle, vu qu'il s'agit de la même image et donc de la même distribution. Les divergences inter-classe varient beaucoup. Les textures *Fabric* et *Food* se ressemblent et leur divergence est faible $D_{KL} = 5.025$.

La base de données *STex* est composée de 476 images de textures couleurs, dont les conditions d'acquisition ne sont pas accessibles. Chaque image est découpée en 16 sous-images sans chevauchement.

La base de données *KTH-TIPS2-b* Caputo et al. (2010) contient 11 textures de 4 échantillons cha-

TABLE 4.13 – **Divergence D_{KL} pour des images de la base *VisTex*.**

VisTex	Fabric	Flowers	Food	Leaves
Fabric	**0**	184.052	5.025	81.305
Flowers	184.052	**0**	99.175	24.311
Food	5.025	99.175	**0**	34.327
Leaves	81.305	24.311	34.327	**0**

TABLE 4.14 – **Divergence D_{KL} pour les variations d'échelle d'une feuille de laitue** extraite de la base *KTH-TIPS2-b*.

KTH-TIPS2-b	Échelle 1	Échelle 6	Échelle 9
Échelle 1	**0**	5.201	4.323
Échelle 6	5.201	**0**	0.696
Échelle 9	4.323	0.696	**0**

cune. Il y a un total de 4752 images, 9 vues à différentes échelles, 4 conditions d'éclairage, et 3 poses. La figure 4.27 présente la texture feuille de laitue sur trois échelles. Les distributions $\mathrm{C_2O}$ associées sont similaires entre l'échelle 6 et 9. Le tableau des divergences est donné table 4.14. Toutes les valeurs sont faibles ce qui s'explique par la similarité des textures vu qu'il s'agit de la même classe. Nous pouvons en conclure que l'attribut de texture $\mathrm{C_2O}$ est invariant aux changements d'échelle.

4.4.2.2 Performance de classification

Dans cette section, nous présentons les taux de classification obtenus pour notre proposition de *Relocated Colour Contrast Occurrences* comparé aux résultats publiés récemment (tab. 4.15).

Tout d'abord, nous pouvons observer que pour trois des quatre bases de données, le $\mathrm{RC_2O}$ utilisé avec la divergence de Kullback-Leibler donne les meilleurs taux de classification : *VisTex* (100%), *STex* (87.24%) et *KTH-TIPS2-b* (96.21%). Pour la base *Outex*, le meilleur taux de classification est obtenu

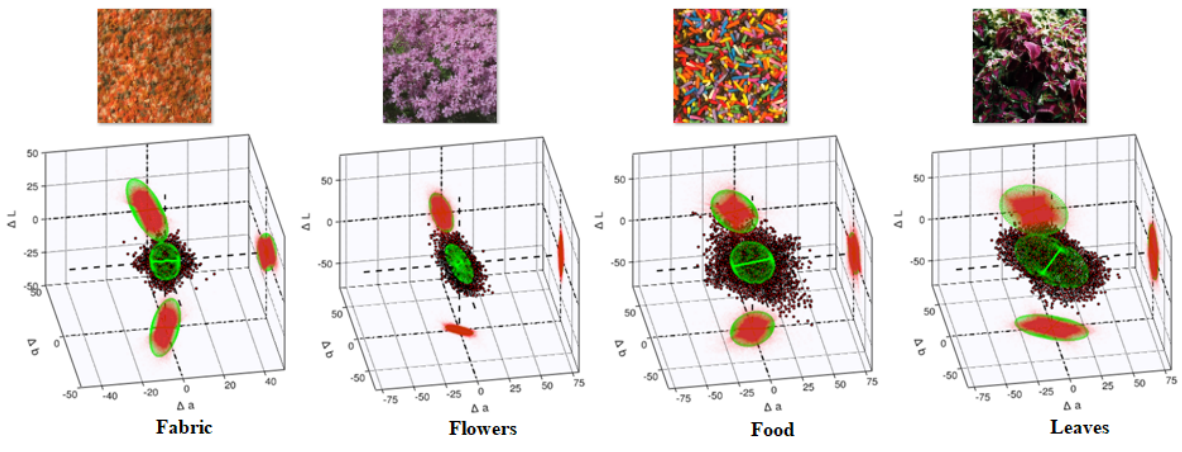


FIGURE 4.26 – **Quelques images de texture de la base *VisTex*, et leurs distributions** $\mathrm{C_2O}$ ($d = 1, \theta = 0$).

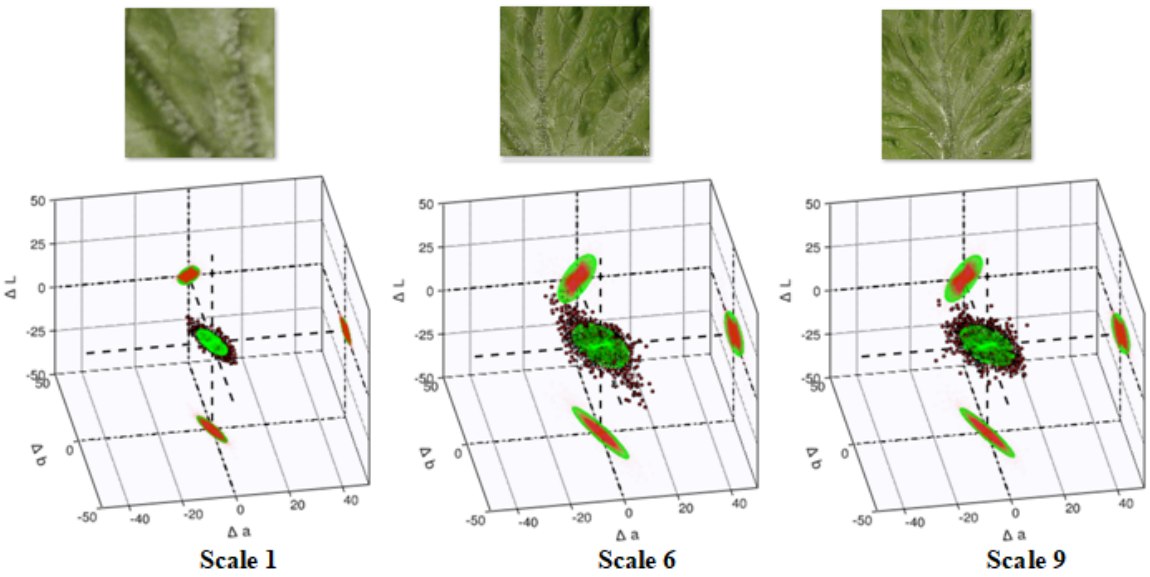


FIGURE 4.27 – **Textures feuille de laitue de la base *KTH-TIPS2-b* et leurs distributions $\mathrm{C_2O}$ associées** pour $d = 1, \theta = 0$.

TABLE 4.15 – **Comparaison des taux de classification(%) obtenus sur les différentes bases présentées**. Notre approche réalise de très bonnes performances, même face à des approches de type réseaux de neurones convolutionnels.

Méthodes	*VisTex*	*STex*	*Outex*	*KTH-TIPS2-b*
LED+ED Pham et al. (2017)	94.70	80.08	76.67	-
CLP Richard et al. (2016b)	97.70	83.90	82.10	-
LSTM network Byeon et al. (2014)	99.09	-	94.7	-
3D-ASDH Sandid and Douik (2016)	-	-	**95.8**	91.3
SMGD Maliani et al. (2014)	97.5	77.6	89.7	-
GLACI Mäenpää and Pietikäinen (2004)	88.42	41.91	87.67	-
CCMA Hauta-Kasari et al. (1996)	91.89	57.58	83.67	-
BF+CLBP+S/M+ScatNet Nguyen et al. (2016)	-	-	-	78.09
FC-CNN+FV-CNN Song et al. (2016)	-	-	-	83.3
STD Alvarez and Vanrell (2012)	98.89	-	90.3	-
Compact DITC Khan et al. (2015)	-	-	-	69
3D connectivity Florindo et al. (2016a)	-	-	84.26	83.54
$\mathbf{C_2O}$ Martínez et al. (2015)	99.3	76.65	82.64	-
$\mathbf{RC_2O_{DKL}}$	**100**	**87.24**	92.4	**96.21**

par la méthode *3D-ASDH* avec (95.8%). Le second meilleur taux est celui obtenu par notre méthode (92.4%).

Deuxièmement, nous pouvons comparer le gain entre l'ancienne version de $\mathrm{C_2O}$ et celle-ci utilisée avec la divergence d_{KL}. Pour la base de données *Outex*, les contenus représentent essentiellement des produits manufacturés avec une variation couleur intra-texture faible mais de grandes variations entre les différentes textures. Dans ce cas, le taux de bonne classification augmente d'environ 12%, ce qui est cohérent vu l'apport de la proposition relocalisée. Les autres bases de données contiennent plus d'images d'objets naturels, la variation locale est donc plus prononcée, le gain est donc plus faible mais toujours positif.

Il est intéressant de noter que plus le contenu de la base de données devient complexe avec l'augmentation d'images du monde naturel, plus le taux de bonne classification diminue. C'est le cas de la base *STex*. Dans ce cas, la plupart des images sont non stationnaires.

Pour finir, notons que nous comparons le $\mathrm{RC_2O}$, avec des approches, pour certaines, basées apprentissage profond. La paire descripteur/mesure de similarité que nous proposons classifie mieux que les méthodes de Song et al. (2016); Nguyen et al. (2016), par exemple, sur la base de données *KTH-TIPS2-b*. Cela montre l'importance de bien choisir son descripteur et le critère de similarité à lui associer. Puisqu'une simple classification plus proche voisin permet d'obtenir un très bon score de classification.

4.5 Synthèse

Dans ce chapitre, nous nous sommes intéressés aux descripteurs de points d'intérêt. Ayant mis en avant l'importance de la famille de données pour le choix de la distance dans le chapitre précédent, nous avons examiné l'état de l'art au travers du filtre du type de données utilisées. Nous ne trouvons qu'un seul type de données : l'histogramme. Celui-ci peut être ou non modélisé et s'adapte à la nature des données d'entrée (valeur de l'image ou gradient de celle-ci par exemple).

Tous les descripteurs sont construits pour être invariants à certaines transformations. Nous proposons donc une étude vérifiant les invariances annoncées par les auteurs. Pour cela, nous avons créé une base de données maîtrisant les paramètres d'invariance et la localisation des points d'intérêt afin d'être dans des conditions aussi aseptisées que possible pour la réalisation de l'expérience. Cette dernière a montré que les invariances n'étaient jamais parfaites mais pouvaient être quantifiées. Les descripteurs étudiés présentent de bons résultats sur les invariances avancées par les auteurs respectifs.

Nous avons ensuite proposé une relecture du travail de Julesz. Ce dernier proposa trois conjectures majeures au cours de sa carrière. Nous avons rattaché différentes approches de descripteur à ces conjectures. Nous avons ensuite proposé le descripteur $\mathrm{RC_2O}$ qui reprend toutes les hypothèses de la première conjecture de Julesz. Ce dernier est la forme plus complète du $\mathrm{C_2O}$. Ces deux approches sont à coupler avec une divergence de Kullback-Leibler puisqu'il s'agit d'une distribution de différences couleur.

Ensuite, nous avons éprouvé ces deux approches en regard du célèbre LBP dans sa version $CCMA$. La première épreuve de classification s'est faite sur une base de données artificielles fractales. Les résultats ont montré l'efficacité des descripteurs $\mathrm{C_2O}$ et $\mathrm{RC_2O}$ ainsi que les limites du $\mathrm{C_2O}$ qui ne prend pas en compte les statistiques d'ordre un. L'étude a aussi mis en avant des limites de la base de données. La génération des images se fait avec une moyenne cible qui n'est pas atteinte avec une erreur suffisamment faible. Ce biais impacte les résultats du $\mathrm{RC_2O}$ attendus.

Une classification sur quatre bases de données réelles a enfin été réalisée suivant un protocole séparant les données d'apprentissage des données à classifier. Les résultats obtenus ici avec le $\mathrm{RC_2O}$ sont plus que prometteurs, dépassant parfois des méthodes basées apprentissage profond.

Résumé

- Création d'une taxonomie des descripteurs en fonction de la nature des données insérées dans l'histogramme constituant l'attribut associé au descripteur.
- Mise en place d'un protocole permettant de vérifier les invariances des descripteurs.
- Mise en évidence du fait que la réponse aux invariances n'est pas simplement binaire mais présente un % d'invariance.
- Proposition d'un nouveau descripteur couleur basé sur la relecture des conjectures de Julesz liant la psychophysique et le traitement d'image.
- Validation des descripteurs proposés sur une base de données artificielles fractales.
- Preuve du bien-fondé de la paire $\{$descripteur,mesure de similarité$\}$ dans le cas de classification sur quatre bases de données réelles (*VisTex*, *STex*, *Outex TC13* et *KTH-TIPS2-b*).

La couleur est le sujet principal de cette thèse. Ce travail de thèse a cherché à répondre à la problématique suivante : **en quoi une prise en compte métrologiquement valide de la couleur améliore-t-elle l'analyse et le traitement des images couleur ?** Par couleur nous entendons son sens générique faisant référence à l'ensemble du spectre visible et pourquoi pas au-delà dans le cas des images hyper-spectrales. Le cadre de la thèse est l'analyse d'images similaires, ce qui se fait majoritairement dans un contexte de classification et d'indexation. Les outils nécessaires à ces traitements, sont les détecteurs et les descripteurs de points d'intérêt sans oublier la mesure de similarité. Notre objectif a été de proposer des approches métrologiques en relation avec la physique de la scène ou la perception de l'observateur.

Pour résoudre cette question avec ces contraintes, nous avons abordé le problème à trois niveaux sur l'échelle de traitement de l'image. Tout d'abord, une analyse a porté sur le gradient couleur qui est la première étape de nombre de traitements. Puis, nous nous sommes intéressés aux détecteurs de points d'intérêt qui sont traditionnellement gérés en niveaux de gris. Enfin, nous nous sommes focalisés sur l'analyse de la texture via les descripteurs qui sont des outils de plus haut niveau, surtout lors de leur association avec une mesure de similarité permettant des tâches d'indexation ou de classification.

Conclusion

Parmi les premières étapes de traitement d'images, le calcul du gradient est un point de départ naturel. Souhaitant adapter le traitement d'images à la couleur, nous avons commencé par nous intéresser aux gradients couleur. L'un des plus utilisés est celui de Di Zenzo (1986) qui s'appuie sur la matrice d'auto-corrélation spatiale. Nous avons montré les limites de sa proposition basée sur l'hypothèse d'orthogonalité de l'espace RGB. Nous avons montré que cette hypothèse n'était pas fondée dans le cadre des capteurs RGB ou des différents espaces RGB existants. Nous avons donc proposé une adaptation de sa formulation qui prend en compte la corrélation entre les canaux. Cette nouvelle écriture est **générique : indépendante des caractéristiques du capteur et du nombre de canaux présents dans l'image**. Afin de démontrer l'apport de cette écriture nous avons créé un protocole de test maîtrisant la “difficulté” d'extraction d'un gradient couleur. **Le protocole pose l'hypothèse que plus les distributions couleur sont entremêlées plus l'extraction d'un bord est délicate**. Les résultats montrent l'apport du *full-vector gradient* lorsque les distributions se chevauchent fortement. Dans les cas de bases moins complexes, les approches existantes sont tout aussi efficaces. Nous avons aussi proposé d'utiliser le *full-vector gradient* dans un cadre perceptuel appliqué à la déficience visuelle des couleurs. Ce résultat préliminaire est un point d'entrée pour de futurs travaux sur la perception visuelle.

Armé d'un gradient couleur ayant fait ses preuves, nous avons montré son importance dans la définition des points d'intérêt. Une revisite de l'état de l'art nous a permis de mettre en avant le manque de prise en compte vectorielle de la couleur dans ce type d'outils. Nous avons proposé d'**adapter le détecteur de Harris and Stephens (1988) à la couleur et au multi-échelles**. Avec la base artificielle créée pour le gradient couleur et la base Oxford habituellement utilisée pour mesurer la qualité d'un détecteur, nous avons montré l'apport d'un détecteur de points d'intérêt couleur pour des images complexes. En particulier, nous avons montré la forte spécificité des détecteurs couleur et leur capacité à ne pas surévaluer le nombre de points d'intérêt. Ce qui peut s'avérer crucial pour les tâches d'indexation et d'appariements $3D$. Nous nous sommes ensuite intéressés à l'adéquation possible entre les points saillants regardés par les humains et les points d'intérêt extraits par les détecteurs. Les résultats obtenus semblent nettement plus prometteurs que notre première attente et appellent à de nouvelles expérimentations psychovisuelles dédiées.

Pour aller plus loin dans l'analyse du contenu “sémantique”, nous avons étudié les descripteurs de points d'intérêt associés à une mesure de similarité. L'analyse de l'état de l'art, nous a permis de présenter une taxonomie des descripteurs en fonction de la nature des données qu'ils utilisent. Les invariances sont des qualités recherchées dans les descripteurs de points d'intérêt. Nous avons mis en place un protocole permettant de confirmer le pourcentage d'invariances d'un descripteur dans des conditions de transformation idéales. Nous avons ensuite fait une relecture du travail de Julesz qui a conduit à **compléter le descripteur** $\mathrm{C_2O}$ basé sur des différences couleur donnant **naissance au** $\mathrm{RC_2O}$. Préalablement, nous avons montré **l'indissociabilité qu'il y avait entre un descripteur de points d'intérêt et le type de distance de similarité** qui devait lui être associé. Pour cela, nous avons reformulé mathématiquement la relation entre la nature des objets à mesurer et la distance à y associer. Cela nous a permis de mettre en avant certaines incohérences des approches de l'état de l'art. La plupart des descripteurs de points d'intérêt proposent d'extraire un histogramme de données de la zone d'intérêt. Nous proposons d'employer la divergence de Kullback-Liebler qui appartient à la famille de mesure de similarité pouvant être utilisée dans le cas de données de probabilité. Nous avons validé la proposition dans un contexte de données artificielles, puis confirmé sa cohérence sur des images couleur de données réelles. Enfin, nous avons corroboré l'efficacité d'une paire $\{$descripteur, mesure de similarité$\}$ avec l'étude des descripteurs proposés ($\mathrm{C_2O}$ et $\mathrm{RC_2O}$), sur une base de données synthétiques fractales puis sur des bases de données réelles.

Dans cette thèse, nous avons proposé trois outils prenant en compte la couleur de façon complète et métrologique à trois échelles de traitement différentes. Nous avons montré l'apport de ces outils pour des images couleur dont la complexité se rapproche de celle des images naturelles. L'importance de l'analyse des outils sur différents niveaux est de pouvoir, à terme, maîtriser toute la chaîne de traitement, sous contraintes métrologiques. Nous espérons ainsi améliorer encore le résultat lorsque toutes les approches seront combinées.

Perspectives

Chaque outil proposé présente des perspectives qui lui sont propres, bien que l'objectif final soit de pouvoir combiner toutes ces propositions en une chaîne complète d'indexation. En effet nous avons proposé un détecteur et un descripteur de points d'intérêt. À l'aide de la mesure de similarité associée, tous les éléments sont présents. Cependant, le détecteur extrait plusieurs points d'intérêt impliquant plusieurs descriptions. Si nous savons mesurer la similarité entre deux points d'intérêt, comment combiner de façon métrologique les différentes descriptions qui caractérisent l'image ou les différentes mesures de similarité entre paires de points d'intérêt ? Il aurait été facile de choisir un critère pour présenter quelques résultats d'indexation. Cependant, prouver que les critères choisis sont cohérents avec la physique de l'objet ou la perception de l'humain est un autre travail de longue durée.

Amélioration des bases de données

Les résultats de la base de données synthétiques varient beaucoup avec le choix des paires de couleurs et l'appariements des distributions. Nous aurions beaucoup à gagner en augmentant le nombre de couleurs et de distributions pour atteindre une stabilité des résultats. Sachant qu'augmenter les paires de couleurs entraînera nécessairement une augmentation du nombre de cas se discriminant aisément en niveaux de gris.

Les données actuelles de la base fractale de synthèse présentent aussi des limites. En effet, nous avons montré que la variabilité des moyennes couleur était trop grande pour s'appuyer sur ces moyennes. De plus, le nombre de cas présentés est trop petit pour couvrir de façon suffisante l'espace

couleur. La base nécessiterait d'être ajustée, pour obtenir une variance plus faible sur les moyennes, puis complétée pour explorer une plus grande complexité spatio-chromatique.

Harris$_{ME}$ et Harris$_{S_2}$ appliqué au multi ou hyper-spectral

Pour la proposition du détecteur couleur, nous employons le *full-vector gradient*. Celui-ci est indépendant du nombre de canaux, nous avons d'ailleurs montré des exemples de gradients mesurés sur des images multi et hyper-spectrales. Cette généricité s'étend aux détecteurs Harris$_{ME}$ et Harris$_{S_2}$ que nous avons proposé. Cependant, aucune preuve de concept sur son emploi en multi ou hyper-spectral n'a été présentée. Un manque qui sera comblé dans les mois à venir.

Compléter le full-vector gradient

La proposition du *full-vector gradient* permet une mesure de direction spatiale du gradient, comportant une norme et une direction. Dans cette thèse nous n'avons exploité que les résultats de norme. L'étude de la direction spatiale du gradient via son angle reste à faire. De plus, certaines zones du cerveau permettant la vision, utilisent l'angle comme information principale (Chossat and Faugeras (2009)).

De plus la proposition du *full-vector gradient*, ne présente qu'une analyse spatiale du gradient, nous perdons toute l'information sur la direction couleur du gradient. Or connaître cette direction (par exemple d'un rouge à un vert ou d'un rouge à un bleu) peut permettre d'associer des bords entre eux lors d'une mesure de segmentation par exemple. Il serait donc intéressant de retrouver cette information par un calcul de différence entre pixels par exemple. Compte tenu des filtres utilisés, quelles positions pour les pixels ?

Expériences psychovisuelle

Les travaux réalisés peuvent être complétés par plusieurs expériences psychovisuelles. L'écriture générique du *full-vector gradient* nous a amené à proposer son utilisation dans le domaine perceptuel au travers d'une étude sur la vision des daltoniens ou sur l'adéquation possible entre points d'intérêt et points saillants. Cela pourrait mener à des expériences visant à confirmer ou infirmer les hypothèses prises. En effet, malgré le support mathématique de notre approche, nous n'avons pas vérifié si les gradients mesurés dans l'espace $CIELMS$ était en correspondance avec la perception humaine déficiente ou non. Pour évaluer cette correspondance, la présentation de l'image avec plusieurs choix de mesure de gradient peut être une approche. Demander à l'utilisateur de dessiner les bords, peut en être une autre, sans oublier pour autant que les gradients et les bords sont des éléments différents.

Nous avons proposé d'employer la divergence de Kullback-Liebler comme mesure de similarité sur les données de type densité de probabilité. Cette utilisation montre des résultats concordants avec nos attentes et notre analyse de la perception. Cela n'en fait pas une preuve. Il serait nécessaire de prouver que cette mesure conserve l'ordonnancement perceptuel. Pour cela impossible d'employer des bases d'images réelles. En effet, l'observateur humain fera nécessairement une analyse du contenu et aura tendance à associer des images de contenu sémantique identique. Il faut donc une base d'images de synthèse. Puisque nous proposons de modéliser les données par une Gaussienne, une telle loi peut être utilisée pour la création de la base. Une expérience psychovisuelle pourrait ensuite permettre de valider l'ordonnancement entre la divergence proposée et la perception humaine.

Malgré toute l'envie développée dans ce travail, il reste encore tant de choses à écrire, développer, mesurer…

Teinte *vs* luminosité

La méthode initiée par Carron and Lambert (1994) est l'une des premières à prendre en compte la couleur autrement que comme trois canaux indépendants.

Tout d'abord, les auteurs effectuent un changement d'espace de l'espace RGB vers l'espace HLS. La première coordonnée correspond à la luminosité de l'image. La seconde est la teinte. C'est une coordonnée angulaire. La troisième coordonnée représente la saturation. Ce sont des coordonnées cylindriques. Avec cet espace, nous pouvons introduire l'idée principale du gradient de Carron. Si la saturation est forte alors l'information teinte est pertinente, sinon l'information luminosité présente plus d'importance.

En pratique, la luminosité et la saturation sont calculées avec le filtre de Sobel. Pour la teinte, qui représente un angle, le calcul est effectué avec cette équation :

$$\Delta(I(x,c_1),I(x',c_1)) = p(I(x,c_3),I(x',c_3))[I(x,c_1) - I(x',c_1)]_\pi \tag{A.1}$$

où c_1 et c_3 représentent respectivement la teinte et la saturation, avec :

$$p(I(x,c_3),I(x',c_3)) = \sqrt{\alpha(I(x,c_3)).\alpha(I(x',c_3))}, \tag{A.2}$$

et $\alpha(S)$ est donné par :

$$\alpha(I(x,c_3)) = \frac{1}{\pi}\left(\frac{\pi}{2} + \arctan(\beta(I(x,c_3) - S_0)\right) \tag{A.3}$$

β contrôle la pente et S_0 fixe le point d'inflexion. La figure A.1 présente trois courbes différentes de $\alpha(S)$.

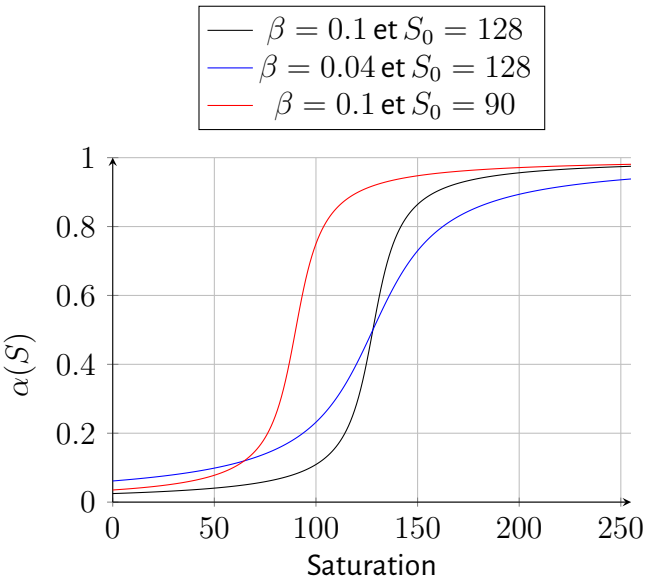


FIGURE A.1 – **Modèle de courbes** $\alpha(S)$

Les trois gradients sont ensuite combinés :

$$\nabla C = \left(\begin{array}{c} \left|\frac{\partial I(x,c_1)}{\partial x_1}\right| + (1-p)\left|\frac{\partial I(x,c_3)}{\partial x_1}\right| + (1-p)\left|\frac{\partial I(x,c_2)}{\partial x_1}\right| \\ \left|\frac{\partial I(x,c_1)}{\partial x_2}\right| + (1-p)\left|\frac{\partial I(x,c_3)}{\partial x_2}\right| + (1-p)\left|\frac{\partial I(x,c_2)}{\partial x_2}\right| \end{array} \right), \tag{A.4}$$

où $I(x, c_2)$ correspond à la coordonnée luminosité et p est le coefficient de l'équation (A.1). Lorsque nous utiliserons ce gradient, ce sera avec les coefficients $\beta = 0.1$ et $S_0 = 128$ comme le propose Carron (1995).

Évaluation des gradients couleur pour des images générées dans $CIELAB$

Comme dans le chapitre présentant les gradients couleur, nous reprenons les deux expériences évaluant l'impact de la forme de la distribution puis l'évaluation générale des gradients.

La scène enregistrée par le capteur n'est pas générée dans l'espace d'acquisition, mais dans le domaine physique où les interactions lumière-matière sont définies. Pour le moment, aucun modèle de génération de texture n'existe dans le domaine spectral. Dans cette expérience, nous considérons des images générées en $CIELAB$. Cet espace couleur permet de contrôler les paramètres aléatoires en respectant la vision humaine tout en vérifiant que les couleurs appartiennent au gamut de l'espace $CIERGB$.

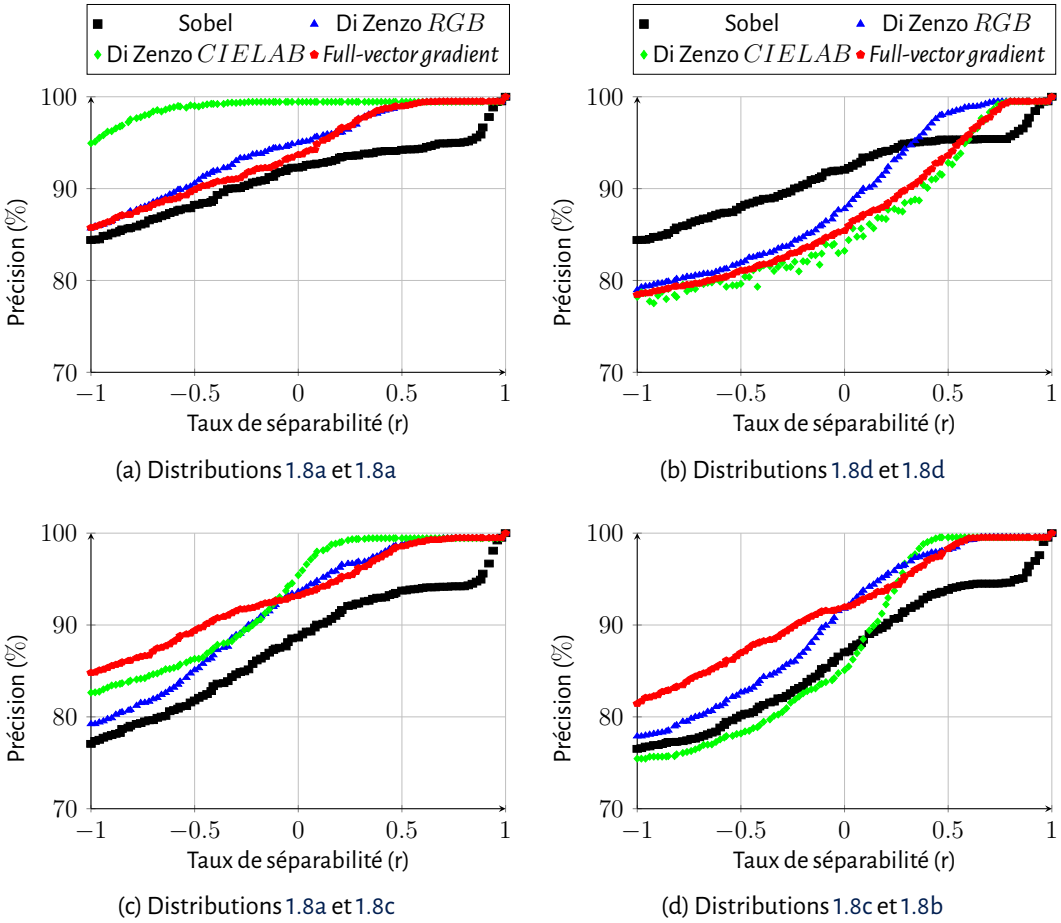
Nous comparerons quatre gradients dans cette expérience : FVG, Sobel, Di Zenzo calculé dans $CIERGB$ et dans l'espace $CIELAB$ où sont générées les images. Nous attendons des résultats similaires à l'expérience en RGB. Di Zenzo en $CIELAB$ devrait donner de meilleurs résultats que les autres quand les distributions sont sphériques. Le *full-vector gradient* devrait être meilleur dans les autres cas.

La figure B.1a montre que Di Zenzo en $CIELAB$ donne de meilleurs résultats puisque l'espace $CIELAB$ a été construit pour être orthogonal. Quand la distribution est plus complexe et que le taux de séparabilité diminue, les limites de l'approche apparaissent clairement (fig. B.1b à B.1d). Le *full-vector gradient* montre son potentiel sur les taux de séparabilité négatifs quand les distributions sont très entremêlées (fig. B.1c et B.1d). La distribution présentée figure 1.8d est principalement dirigée sur l'axe luminosité lors d'une génération dans $CIELAB$. En effet, la rotation autour du premier axe est de $9°$. L'ellipsoïde étant très allongée, la variation couleur se fera majoritairement sur l'axe luminosité et la transformation en niveaux de gris requise par le gradient de Sobel dans notre protocole écrasera encore les différences qui auraient pu être présentes sur les axes couleur. Cela explique pourquoi la figure B.1b présente le gradient de Sobel comme le plus performant. Nous retrouverons cette tendance sur toutes les courbes présentant la distribution de la figure 1.8d présentées en annexe D.

Ces figures présentent des résultats pour une paire de distributions à chaque sous-figure. Comme dans le chapitre 1, nous avons aussi effectué une analyse statistique sur 500 images (tab. B.1). Si l'on considère les approches qui ne se font pas dans l'espace $CIELAB$, le *full-vector gradient* est toujours le plus précis. Les écarts-type associés à ses résultats sont cependant les plus élevés.

Comme nous l'attendions, la formulation de Di Zenzo dans l'espace $CIELAB$ obtient de très bons résultat pour des taux de séparabilité élevé ($r > 1$) donc des images faciles à segmenter.

Avec ces deux expériences, nous sommes face à un problème de validation. Pour évaluer la précision de l'extraction du gradient, nous avons besoin de deux régions à segmenter et nous devons explorer le niveau de difficulté des distributions couleur. Cette difficulté peut se définir de différentes manières, par l'entropie ou l'organisation fractale par exemple. La question principale reste l'espace de génération des images. L'espace couleur idéal pour la génération de texture est le domaine spectral avec un spectre continu. Actuellement, aucune génération couleur permet cette contrainte. L'espace $CIELAB$ est un espace où la différence de couleur est perceptuellement uniforme ce qui le rend intéressant. Un autre espace est le $CIEXYZ$ qui est plus proche de la physique de l'acquisition mais dont la relation avec un espace RGB est linéaire, lui donnant les mêmes contraintes.



(a) Distributions 1.8a et 1.8a

(b) Distributions 1.8d et 1.8d

(c) Distributions 1.8a et 1.8c

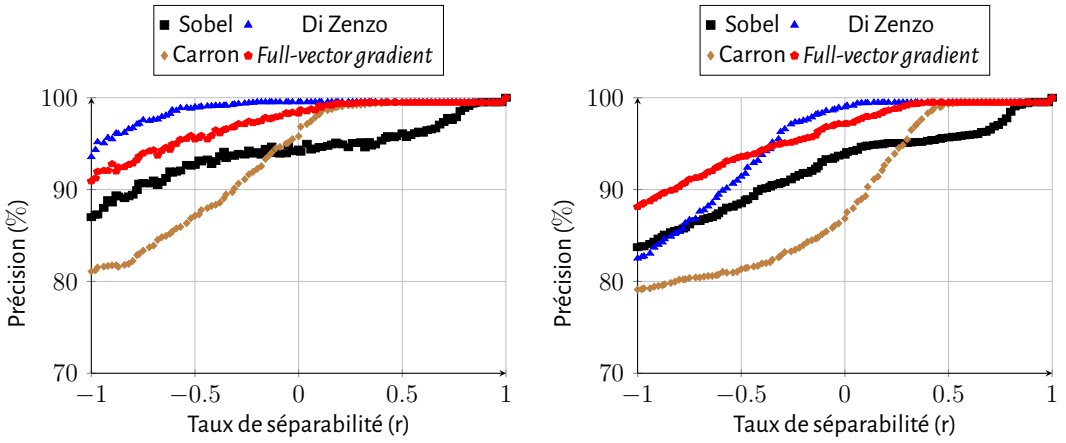
(d) Distributions 1.8c et 1.8b

FIGURE B.1 – **Exemple d'extraction de bord pour des images générées avec** $\Delta E = 3$. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.

TABLE B.1 – **Moyennes et écarts-type de la précision** pour un taux de séparabilité donné avec des images générées dans $CIELAB$.

	Di Zenzo $CIELAB$		Di Zenzo RGB		FVG		Sobel	
r	Moy	E-T	Moy	E-T	Moy	E-T	Moy	E-T
-1	76.60	*0.18*	77.18	0.24	**78.28**	0.34	77.33	0.24
-0.5	78.94	*0.38*	79.39	0.43	**80.80**	0.59	79.38	0.44
-0.25	81.17	*0.56*	81.27	0.60	**82.77**	0.76	81.02	0.58
-0.15	82.43	*0.65*	82.17	0.68	**83.69**	0.84	81.84	*0.65*
0	84.70	0.81	83.69	0.81	**85.18**	0.94	83.08	*0.76*
0.15	*87.71*	1.01	85.48	0.91	**86.88**	1.07	84.55	*0.88*
0.25	*90.04*	1.13	86.85	0.99	**88.05**	1.14	85.68	*0.95*
0.5	*96.09*	*0.97*	91.17	1.35	**91.75**	1.29	88.84	1.13

Impact de la forme de distribution pour des images générées en RGB



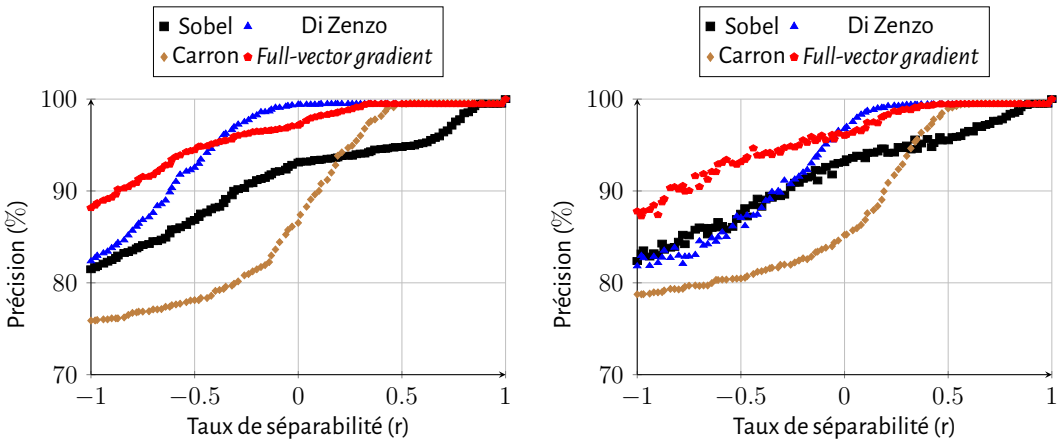
(a) Distributions 1.8a et 1.8a

(b) Distributions 1.8a et 1.8b

(c) Distributions 1.8a et 1.8c

(d) Distributions 1.8a et 1.8d

FIGURE C.1 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



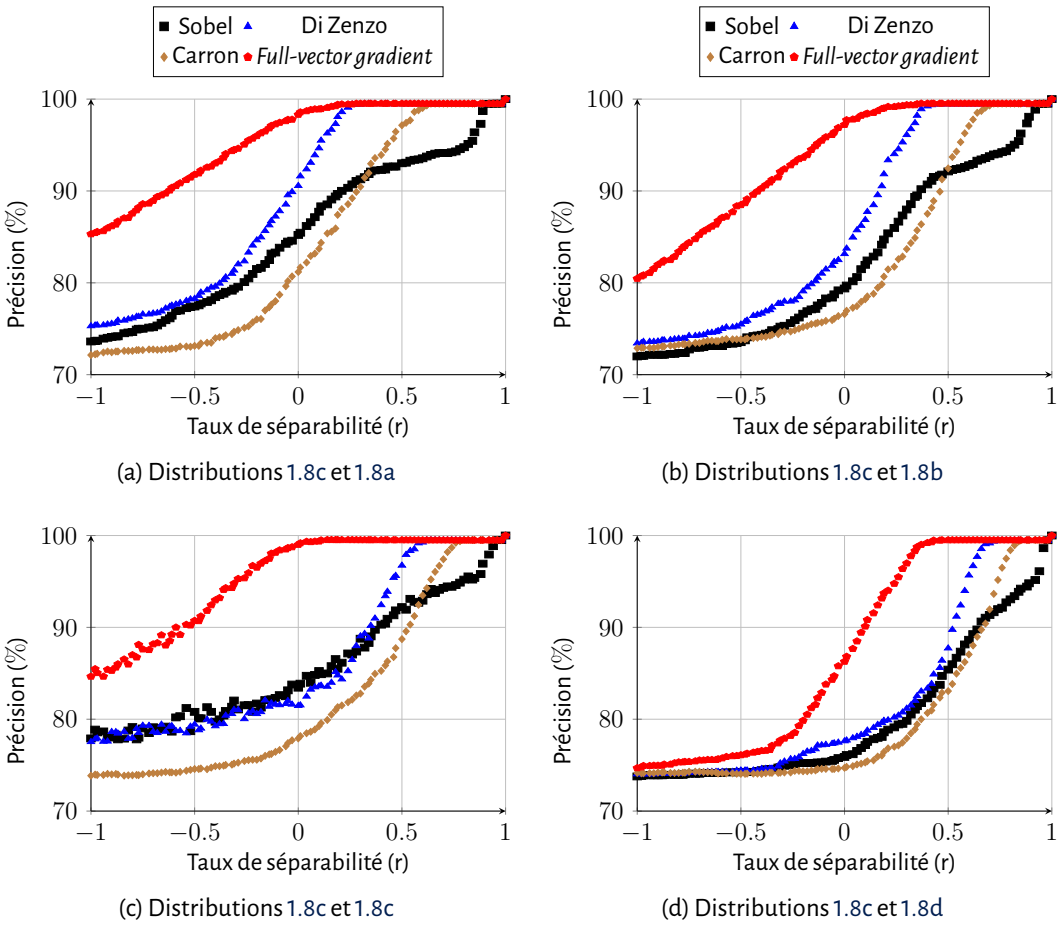
(a) Distributions 1.8b et 1.8a

(b) Distributions 1.8b et 1.8b

(c) Distributions 1.8b et 1.8c

(d) Distributions 1.8b et 1.8d

FIGURE C.2 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



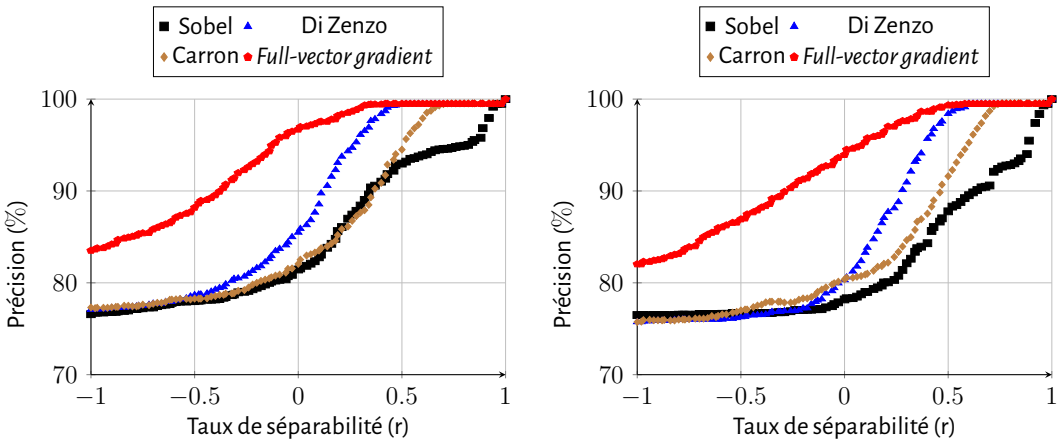
(a) Distributions 1.8c et 1.8a

(b) Distributions 1.8c et 1.8b

(c) Distributions 1.8c et 1.8c

(d) Distributions 1.8c et 1.8d

FIGURE C.3 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



(a) Distributions 1.8d et 1.8a

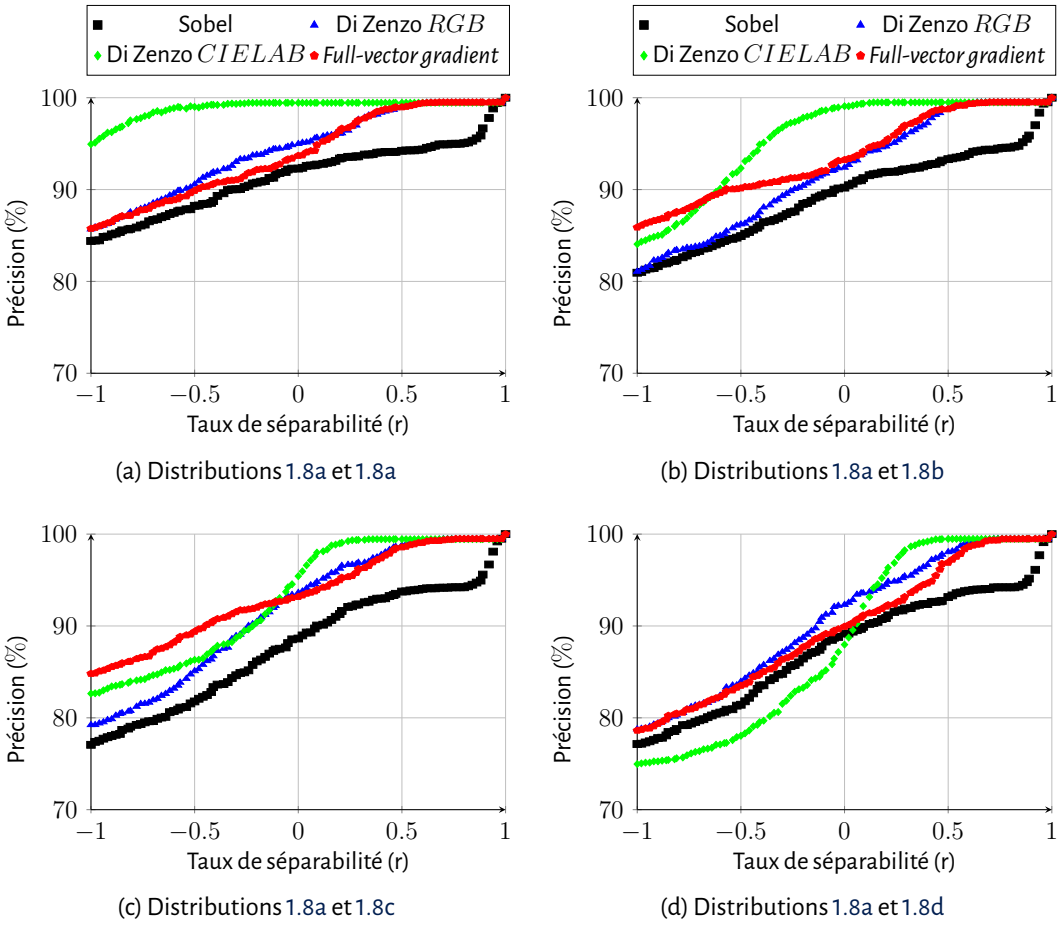
(b) Distributions 1.8d et 1.8b

(c) Distributions 1.8d et 1.8c

(d) Distributions 1.8d et 1.8d

FIGURE C.4 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.

Impact de la forme de distribution pour des images générées en $CIELAB$



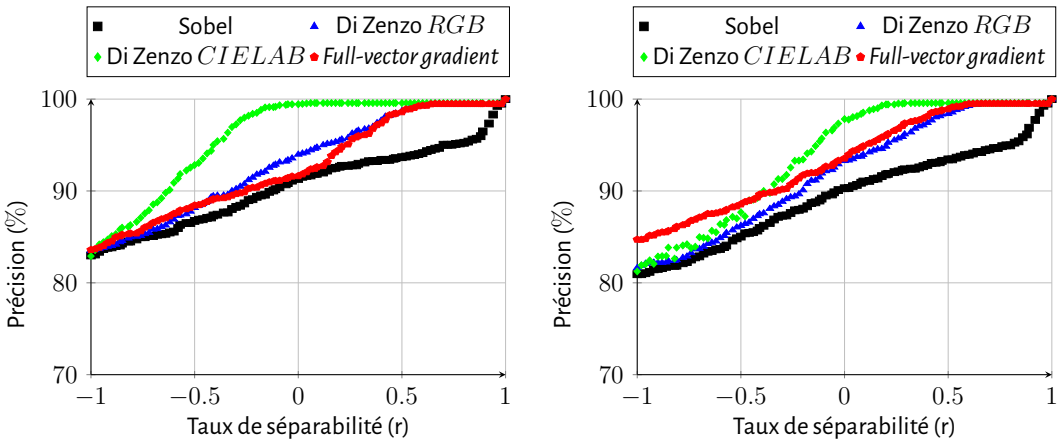
(a) Distributions 1.8a et 1.8a

(b) Distributions 1.8a et 1.8b

(c) Distributions 1.8a et 1.8c

(d) Distributions 1.8a et 1.8d

FIGURE D.1 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



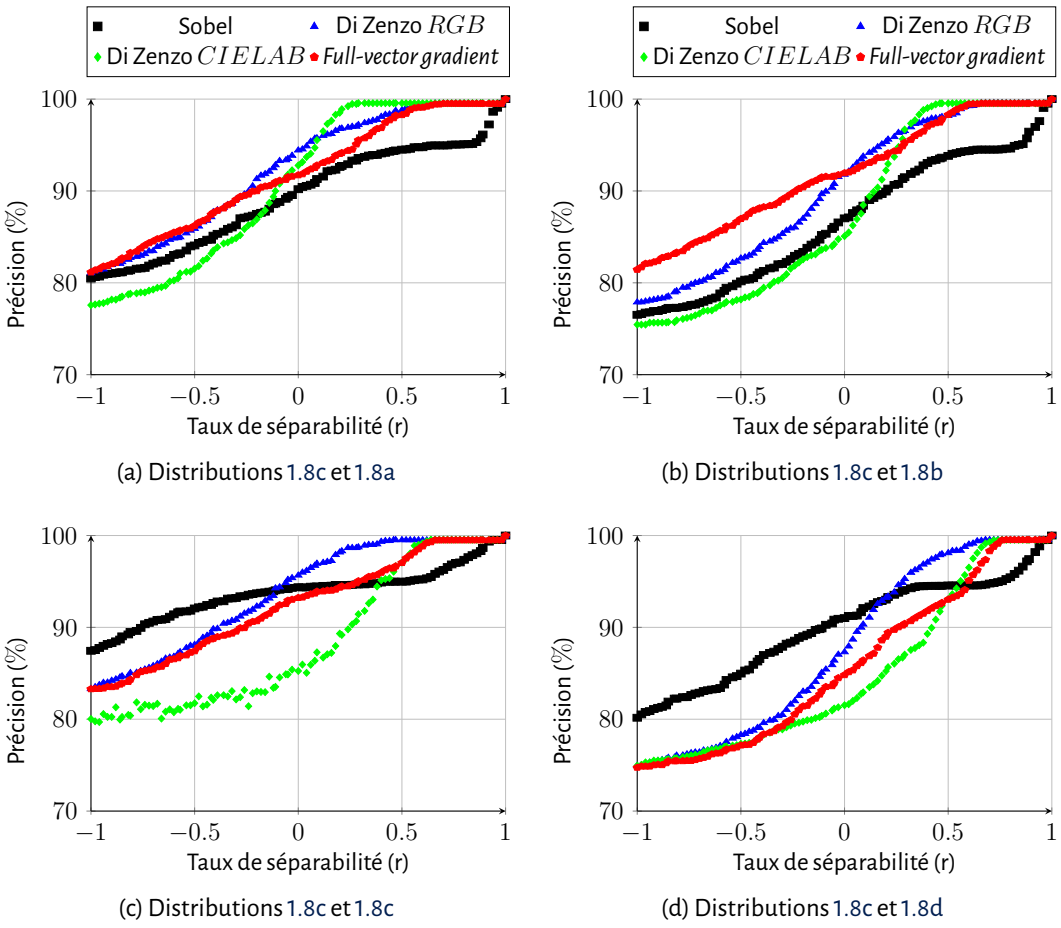
(a) Distributions 1.8b et 1.8a

(b) Distributions 1.8b et 1.8b

(c) Distributions 1.8b et 1.8c

(d) Distributions 1.8b et 1.8d

FIGURE D.2 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



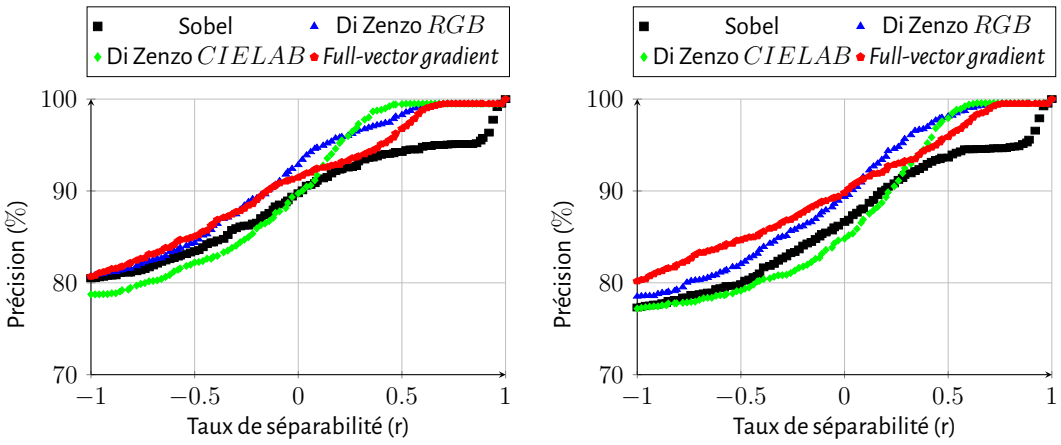
(a) Distributions 1.8c et 1.8a

(b) Distributions 1.8c et 1.8b

(c) Distributions 1.8c et 1.8c

(d) Distributions 1.8c et 1.8d

FIGURE D.3 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.



(a) Distributions 1.8d et 1.8a

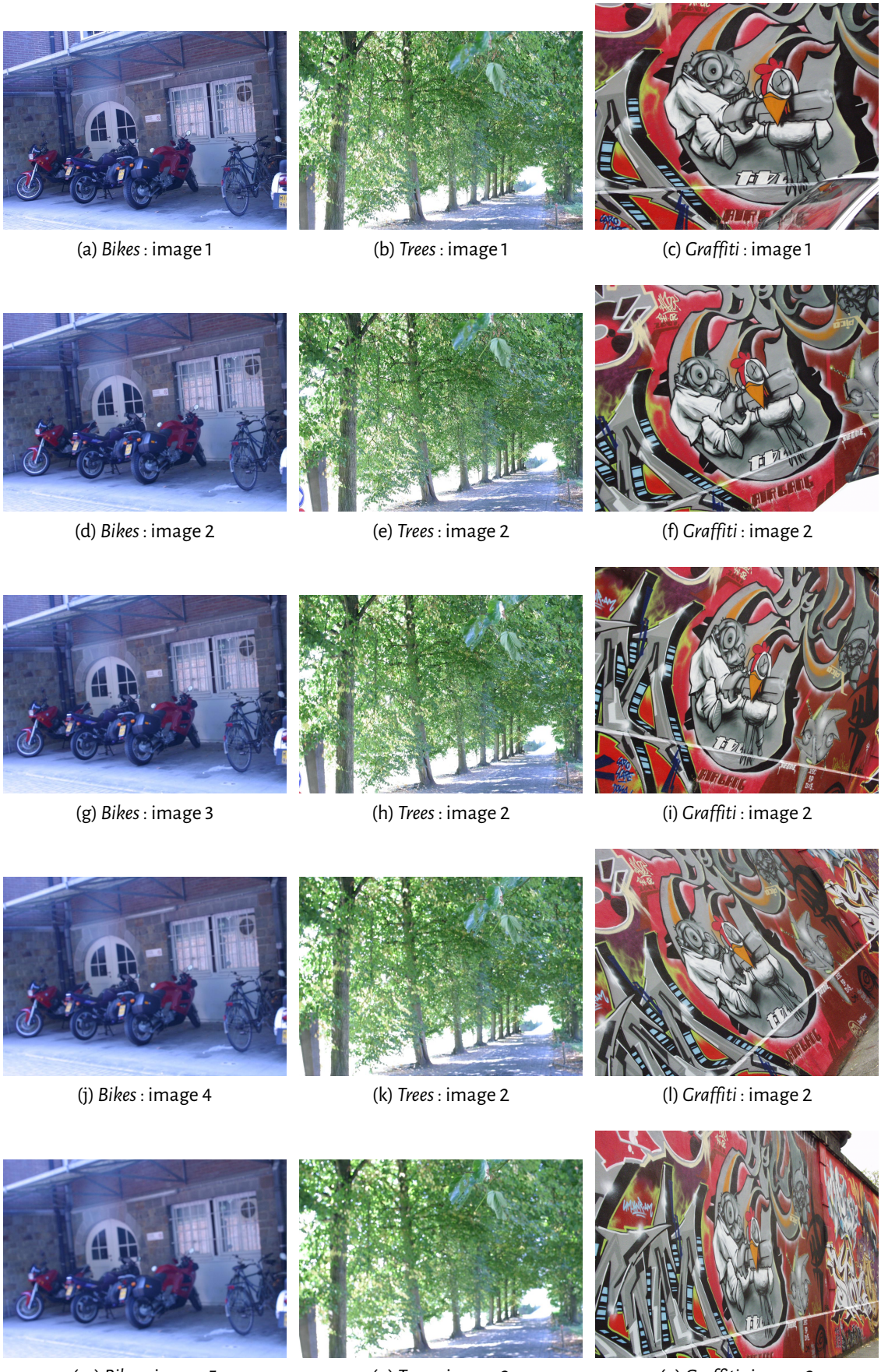
(b) Distributions 1.8d et 1.8b

(c) Distributions 1.8d et 1.8c

(d) Distributions 1.8d et 1.8d

FIGURE D.4 – **Résultat de séparabilité du gradient pour différentes formes de distributions**. Seulement deux distributions sont utilisées par sous-figure. Les formes des distributions utilisées sont données par sous-figure et font référence à la figure 1.8.

Images de la base Oxford utilisées



(a) *Bikes* : image 1 (b) *Trees* : image 1 (c) *Graffiti* : image 1

(d) *Bikes* : image 2 (e) *Trees* : image 2 (f) *Graffiti* : image 2

(g) *Bikes* : image 3 (h) *Trees* : image 2 (i) *Graffiti* : image 2

(j) *Bikes* : image 4 (k) *Trees* : image 2 (l) *Graffiti* : image 2

(m) *Bikes* : image 5 (n) *Trees* : image 2 (o) *Graffiti* : image 2

FIGURE E.1 – **Images de la base Oxford**.



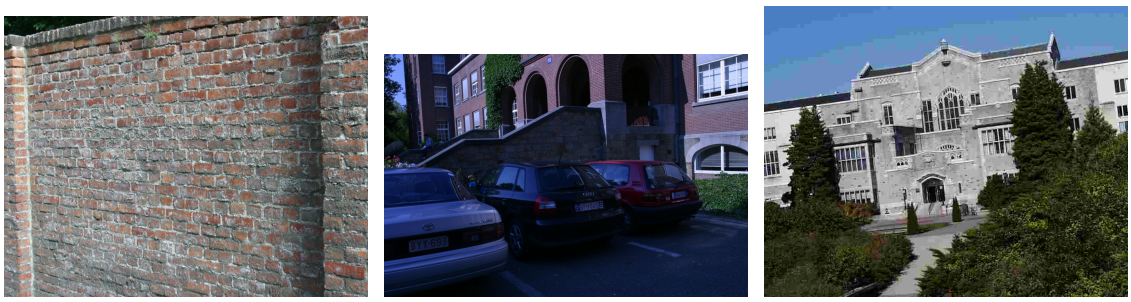
(a) *Wall* : image 1 (b) *Leuven* : image 1 (c) *UBC* : image 1



(d) *Wall* : image 2 (e) *Leuven* : image 2 (f) *UBC* : image 2



(g) *Wall* : image 3 (h) *Leuven* : image 2 (i) *UBC* : image 2



(j) *Wall* : image 4 (k) *Leuven* : image 2 (l) *UBC* : image 2



(m) *Wall* : image 5 (n) *v* : image 2 (o) *UBC* : image 2

FIGURE E.2 – **Images de la base Oxford**.

Tests d'hypothèse sur le modèle Gaussien des images couleur

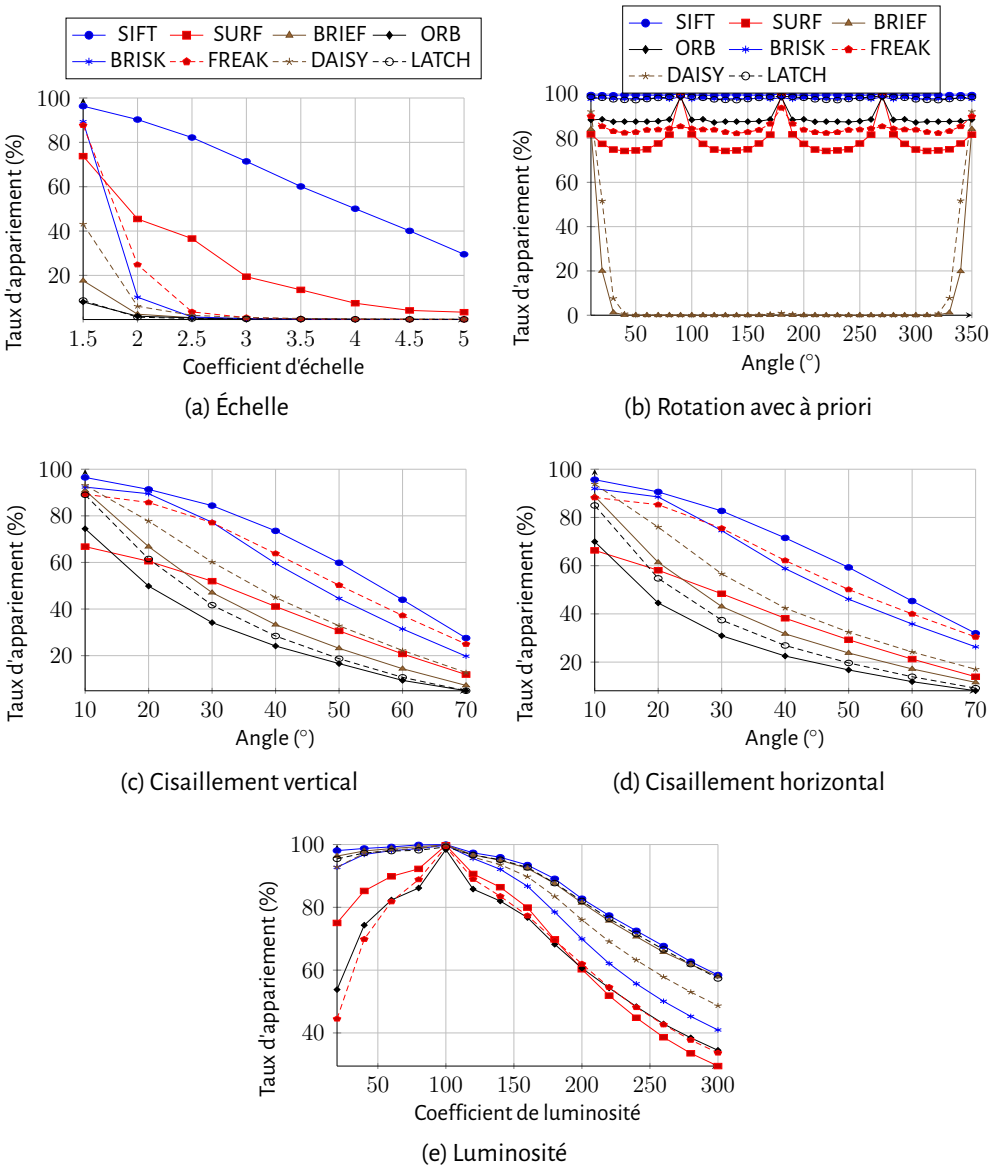
TABLE F.1 – **Tests d'hypothèse de normalité pour les images de textures proches**.

		RGB	$CIELAB$
	χ^2 de Pearson	1.3×10^{-198600}	1.4×10^{-5998337}
	Cramér-von Mises	1.2×10^{-41}	7.0×10^{-38}
	Jarque-Bera	8.9×10^{-10744}	3.8×10^{-11330}
	Mardia kurtosis	4.5×10^{-205006}	1.4×10^{-156595}
	Mardia asymétrie	5.6×10^{-96425}	2.4×10^{-101599}
	χ^2 de Pearson	3.1×10^{-201793}	1.5×10^{-6533278}
	Cramér-von Mises	2.7×10^{-42}	7.2×10^{-38}
	Jarque-Bera	1.5×10^{-11011}	3.8×10^{-11330}
	Mardia kurtosis	8.0×10^{-533003}	1.4×10^{-156595}
	Mardia asymétrie	7.7×10^{-178529}	2.4×10^{-101599}
	χ^2 de Pearson	1.2×10^{-164949}	8.9×10^{-6915031}
	Cramér-von Mises	1.7×10^{-41}	6.0×10^{-38}
	Jarque-Bera	3.8×10^{-11330}	3.8×10^{-11330}
	Mardia kurtosis	1.4×10^{-156595}	1.4×10^{-156595}
	Mardia asymétrie	2.4×10^{-101599}	2.4×10^{-101599}
	χ^2 de Pearson	4.8×10^{-346202}	$3.2 \times 10^{-12669932}$
	Cramér-von Mises	3.3×10^{-40}	2.5×10^{-38}
	Jarque-Bera	1.6×10^{-22195}	3.8×10^{-11330}
	Mardia kurtosis	2.2×10^{-300238}	1.4×10^{-156595}
	Mardia asymétrie	1.8×10^{-147035}	2.4×10^{-101599}

TABLE F.2 – **Tests d'hypothèse de normalité pour les images de textures différentes**.

		RGB	$CIELAB$
	χ^2 de Pearson	9.8×10^{-112982}	$1.9 \times 10^{-10492114}$
	Cramér-von Mises	2.7×10^{-42}	7.8×10^{-38}
	Jarque-Bera	1.3×10^{-7640}	1.6×10^{-22195}
	Mardia kurtosis	4.0×10^{-28022}	2.2×10^{-300238}
	Mardia asymétrie	3.2×10^{-32701}	1.8×10^{-147035}
	χ^2 de Pearson	3.1×10^{-241829}	$4.9 \times 10^{-17451478}$
	Cramér-von Mises	6.6×10^{-42}	3.9×10^{-38}
	Jarque-Bera	1.0×10^{-9596}	1.6×10^{-22195}
	Mardia kurtosis	5.2×10^{-4066}	2.2×10^{-300238}
	Mardia asymétrie	6.0×10^{-17163}	1.8×10^{-147035}
	χ^2 de Pearson	3.7×10^{-770900}	$3.8 \times 10^{-50204272}$
	Cramér-von Mises	1.1×10^{-43}	9.6×10^{-38}
	Jarque-Bera	1.2×10^{-4222}	1.6×10^{-22195}
	Mardia kurtosis	9.3×10^{-55960}	2.1×10^{-300238}
	Mardia asymétrie	1.4×10^{-36788}	1.8×10^{-147035}
	χ^2 de Pearson	3.6×10^{-190277}	$3.8 \times 10^{-50204272}$
	Cramér-von Mises	2.6×10^{-42}	9.6×10^{-38}
	Jarque-Bera	5.9×10^{-3893}	1.6×10^{-22195}
	Mardia kurtosis	4.6×10^{-3057}	2.2×10^{-300238}
	Mardia asymétrie	2.2×10^{-33144}	1.8×10^{-147035}

G.1 Influence de la taille de la zone d'intérêt par transformation



(a) Échelle

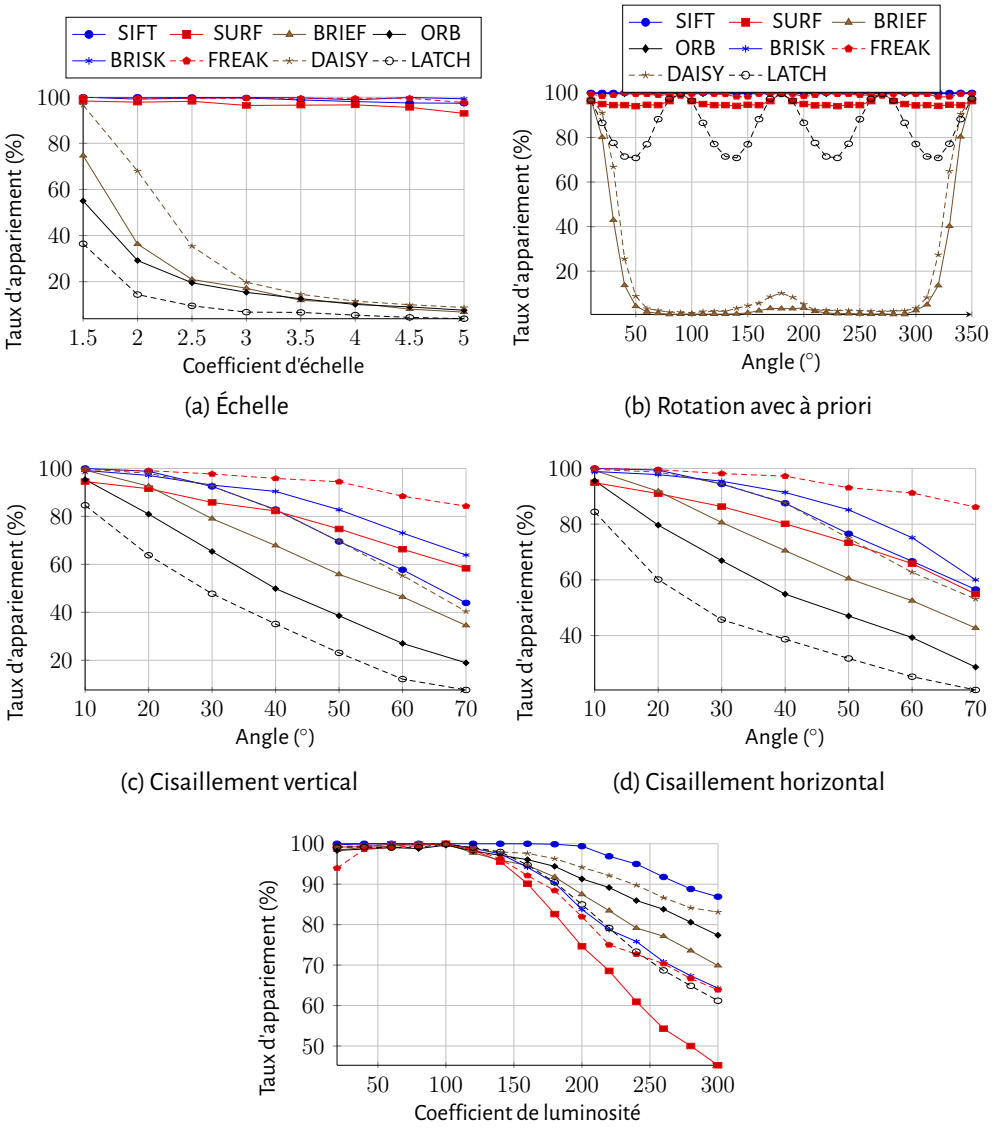
(b) Rotation avec à priori

(c) Cisaillement vertical

(d) Cisaillement horizontal

(e) Luminosité

FIGURE G.1 – **Comparaison des résultat d'invariances des descripteurs** pour un diamètre de zone d'intérêt de 10 pixels.



(a) Échelle

(b) Rotation avec à priori

(c) Cisaillement vertical

(d) Cisaillement horizontal

(e) Luminosité

FIGURE G.2 – **Comparaison des résultat d'invariances des descripteurs** pour un diamètre de zone d'intérêt de 100 pixels.

G.2 Influence de la taille de la zone d'intérêt par descripteur

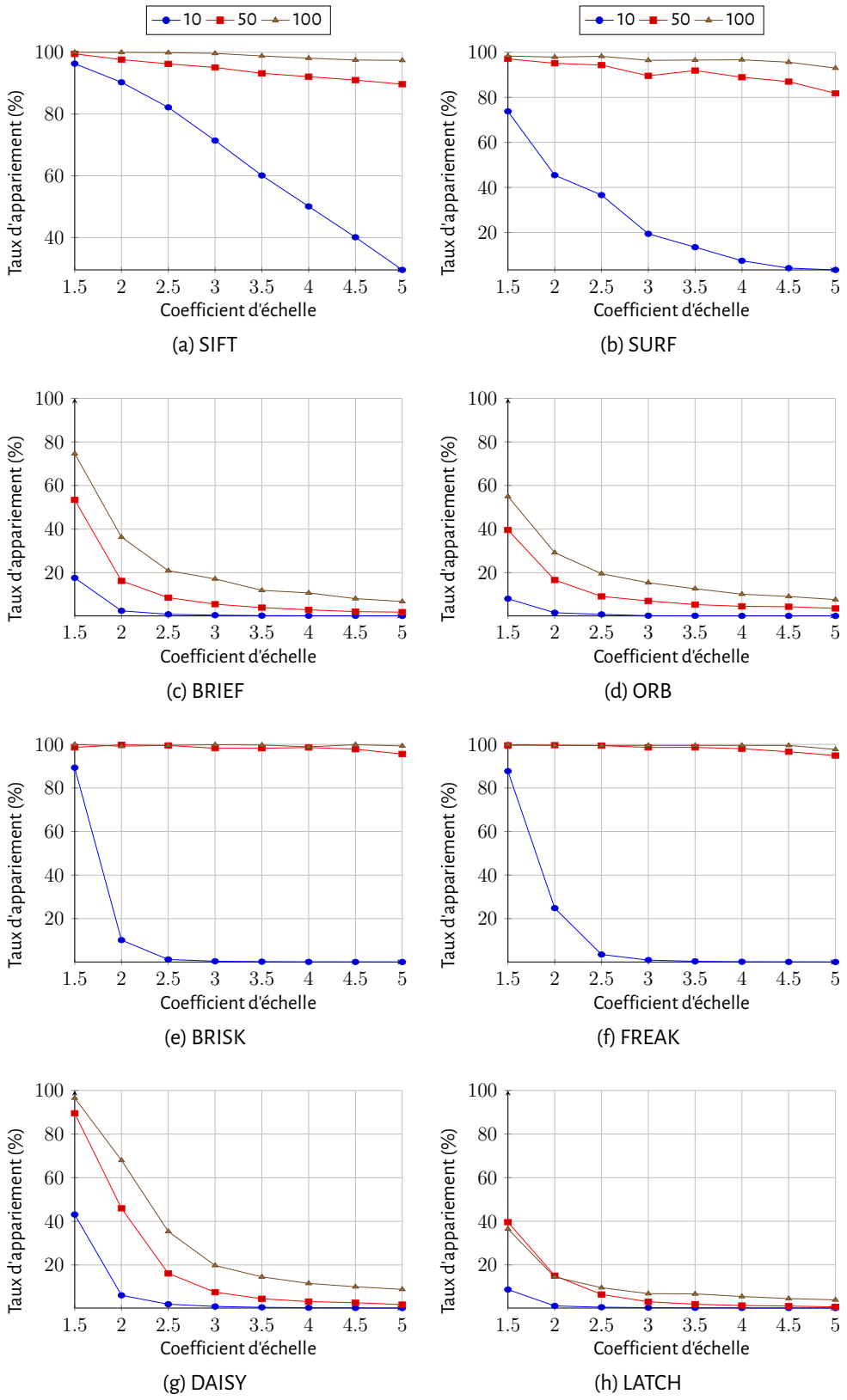
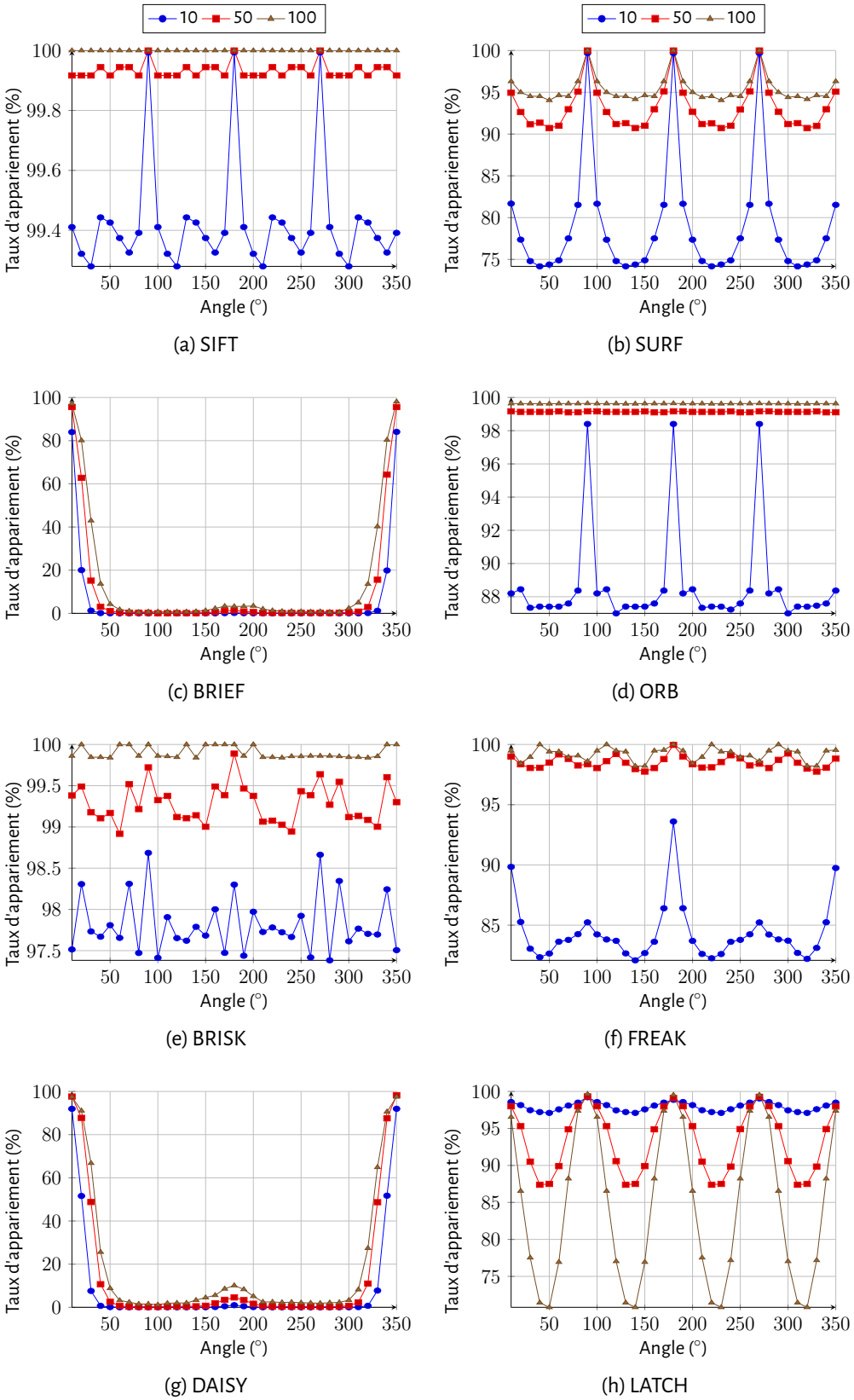


FIGURE G.3 – **Comparaison des résultats sur l'invariance d'échelle** en fonction de la zone d'intérêt.



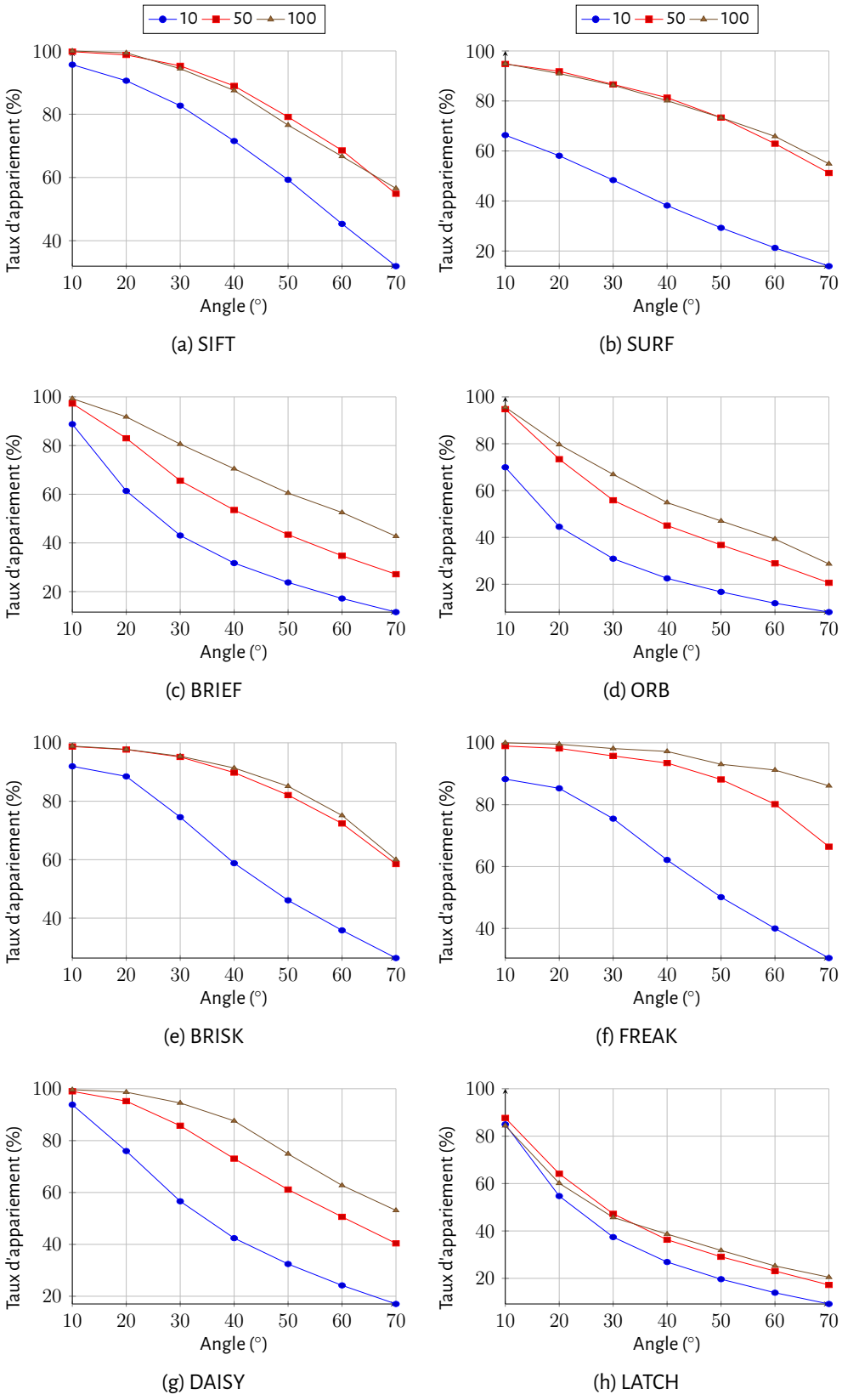
(a) SIFT (b) SURF

(c) BRIEF (d) ORB

(e) BRISK (f) FREAK

(g) DAISY (h) LATCH

FIGURE G.4 – **Comparaison des résultats sur l'invariance de rotation** en fonction de la zone d'intérêt.



(a) SIFT (b) SURF

(c) BRIEF (d) ORB

(e) BRISK (f) FREAK

(g) DAISY (h) LATCH

FIGURE G.5 – **Comparaison des résultats sur l'invariance de cisaillement horizontal** en fonction de la zone d'intérêt.

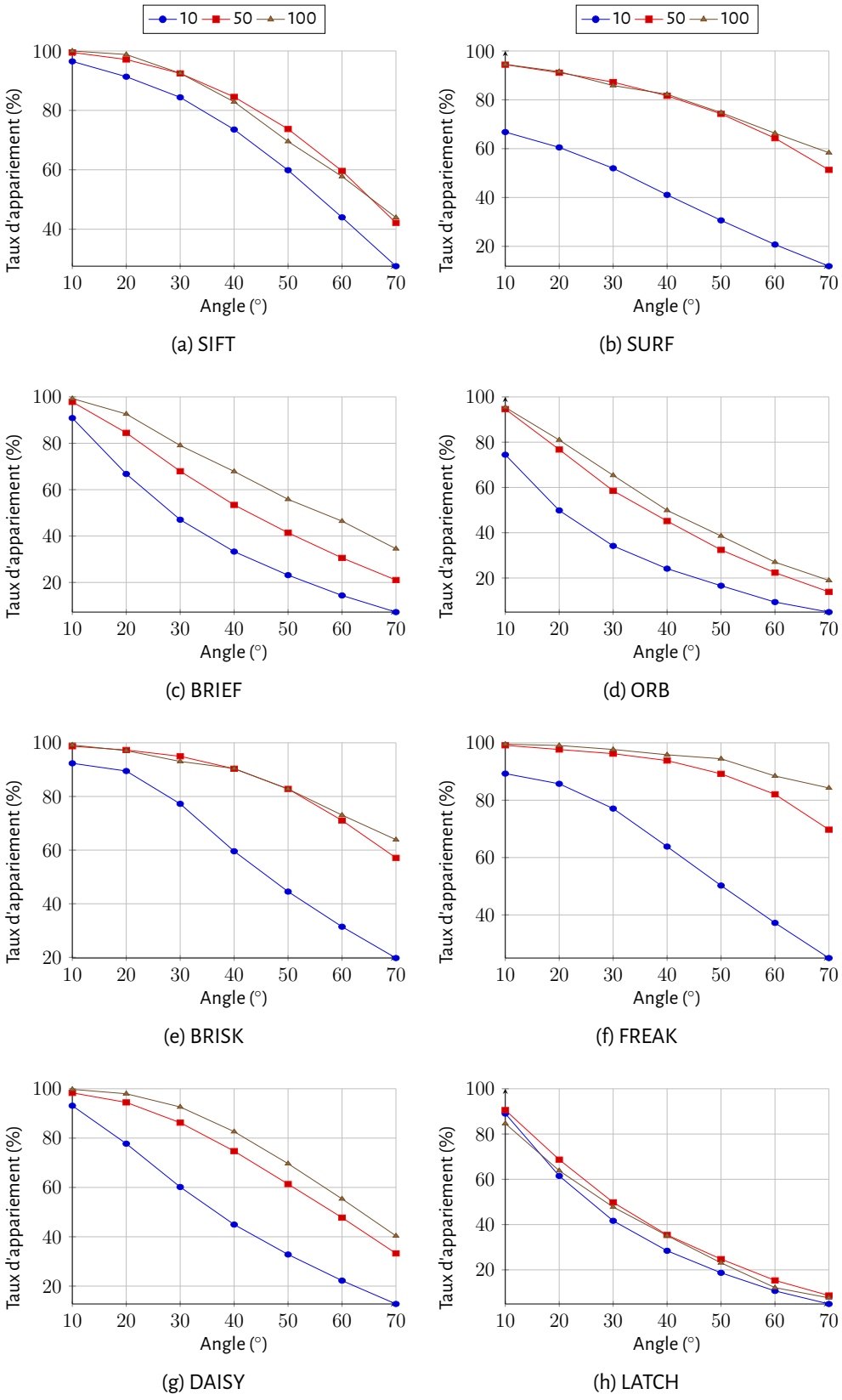
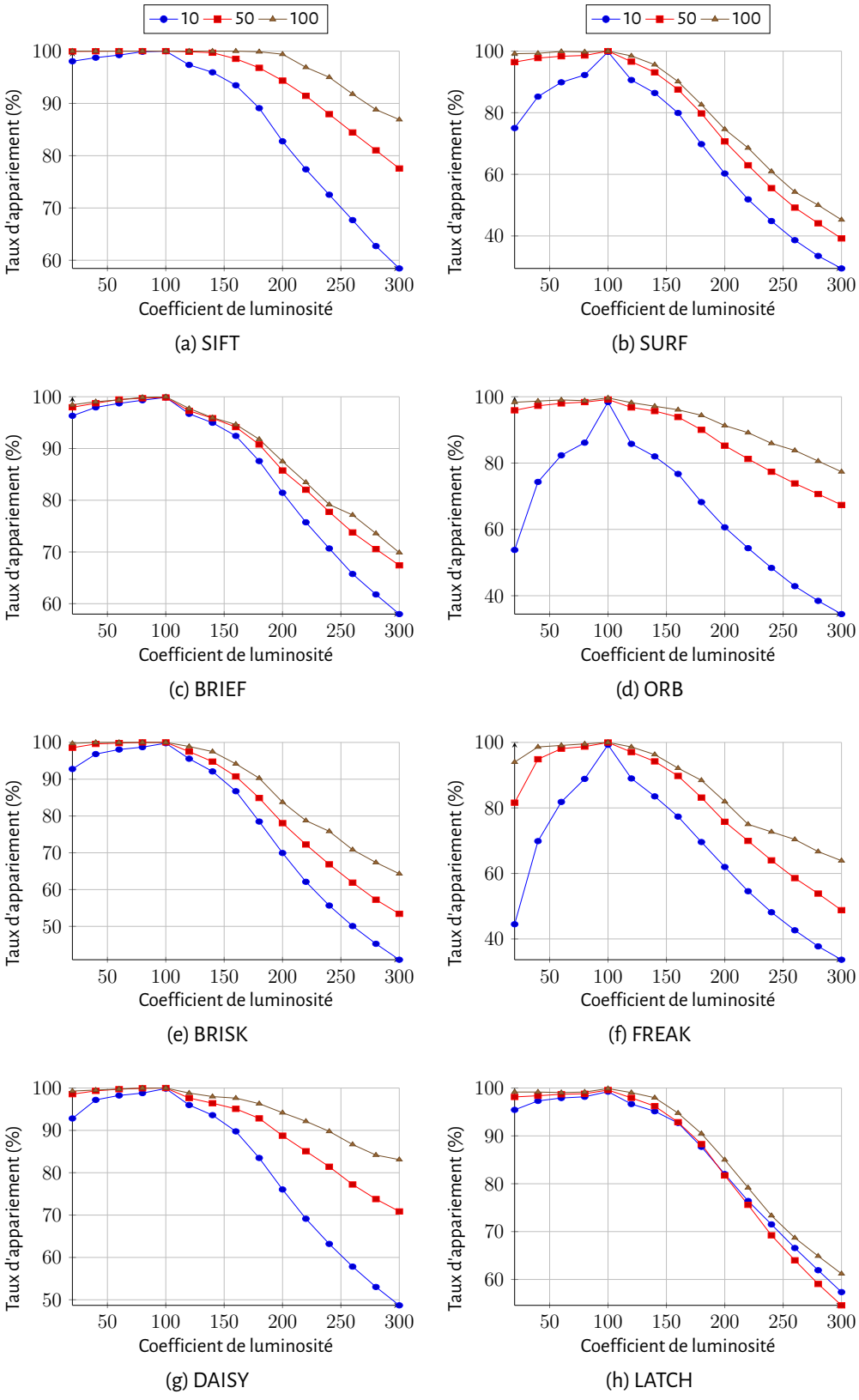


FIGURE G.6 – **Comparaison des résultats sur l'invariance de cisaillement vertical**. en fonction de la zone d'intérêt



(a) SIFT (b) SURF

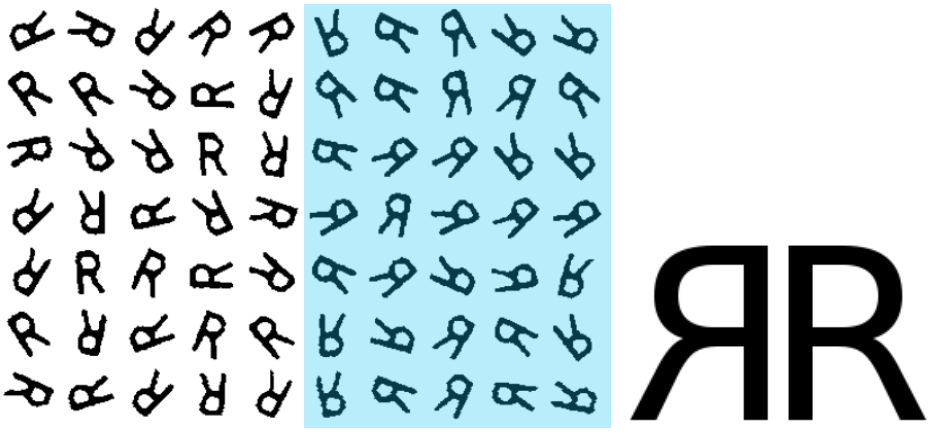
(c) BRIEF (d) ORB

(e) BRISK (f) FREAK

(g) DAISY (h) LATCH

FIGURE G.7 – **Comparaison des résultats sur l'invariance à la luminosité** en fonction de la zone d'intérêt.

Deuxième conjecture de Julesz



(a) Séparation des deux textures (b) Motif à différentier

FIGURE H.1 – **Solution de segmentation pour les statistiques d'ordre deux identiques**. Images extraites de Julész et al. (1978) puis modifiées .

Construction de la base de données synthétiques fractales

Les objets de type fractals se définissent par une auto-similarité dans les échelles : chaque partie ressemble au “tout”. La mesure utilisée pour évaluer la complexité de ces objets est la dimension de Hausdorff Falconer (2004). Mandelbrot associe d'autres caractéristiques à la description de l'aspect des objets fractals, telles que la lacunarité par exemple. Dans ce travail, nous nous concentrons sur les aspects "texture fine" induite par la dimension fractale, et non à l'aspect variation basse fréquence induite par la lacunarité ou la succolarité.

Prenons un exemple pour illustrer cette notion de dimension fractale. Les poumons sont considérés comme des objets naturels respectant pour partie le modèle fractal. Ils représentent des surfaces arrangées dans l'espace 3D en plis et replis. Cette construction induit que le poumon soit plus considéré comme un volume, que comme une surface. Ce qui se traduit par un rapport surface/volume occupé important.

Une dimension fractale est une mesure comprise entre la dimension topologique entière N de l'objet étudié et $N+1$. Dans notre cas, la dimension sera entre 5 (2 dimension spatiale et 3 pour la couleur) et 6.

Modèle fractal

Dans l'idée de générer des fractales vectorielles couleur, nous nous sommes appuyés sur le modèle du bruit Brownien fractionnaire introduit par Kolmogorov (1940) cité par Mandelbrot and Van Ness (1968). Ce modèle de bruit appartient au groupe des bruits de type $\frac{1}{f}$. Considérons une variable aléatoire multivaluée $I(x)$ telle que ses variations spatiales $I(y)-I(x)$ suivent des lois Gaussiennes multivaluées de moyennes nulles dont les variances dépendent de la distance entre x et y et de la complexité fractale définie par le coefficient de Hurst :

$$\mathbb{E}\left((I(y)-I(x))^2\right) \propto \|y-x\|_2^{2H}, \tag{I.1}$$

où $\mathbb{E}$ est l'espérance statistique, x et y sont des positions spatiales de pixel et $H \in]0,1[$ est le coefficient de Hurst.

Dans le cas des générateurs de terrains ou de surfaces dans des espaces 3D, la dimension fractale sera liée à la rugosité de la surface générée. Dans le cas des signaux unidimensionnels, la dimension fractale induit des signaux de bruit, dits colorés. Dans notre cas, les images couleurs générées ressemblent à des images de bruit dont la complexité de contenu est fonction de cette dimension fractale.

Il existe deux principales méthodes pour générer des images fractales suivant ce bruit brownien fractionnaire, la méthode du point médian proposée par Voss (1988) et une méthode basée sur la décomposition fréquentielle du signal présentée par Saupe (1988). La technique employée pour cette base de données est la seconde. Il s'agit de modéliser le signal fréquentiel selon l'équation (I.2), puis de transformer le signal dans l'espace spatial pour obtenir l'image fractale.

Le générateur fractal choisi est de type synthèse fréquentielle Voss (1988), permettant un contrôle direct des densités spectrales de puissance des images générées. Il est de plus relativement isotrope. Il a été adapté au domaine couleur par l'utilisation de quaternions. La densité spectrale de puissance ainsi générée suit une loi de type :

$$S(f) = \frac{1}{f^{2H+E}}, \tag{I.2}$$

où E représente la dimension de l'espace de la variable aléatoire multivaluée (ici 5). La dimension de Haussdorf, abusivement appelé dimension fractale, de ce modèle ne dépend que du paramètre H : $D = E+1-H$. L'analyse d'une surface fractale nécessite par définition une approche multi-échelles permettant d'estimer l'évolution des mesures selon les échelles dans un plan log-log.

L'ensemble des quaternions

Afin de prendre en compte la couleur de façon vectorielle, une représentation à base de quaternions est utilisée. L'ensemble des quaternions a été présenté par Hamilton en 1843. Les quaternions sont des nombres complexes dont les coordonnées sont complexes. Ils peuvent s'écrire sous la forme :

$$p = a + bi + cj + dk, \tag{I.3}$$

où a, b,c et d sont des réels et i, j et k symbolisent les imaginaires. Ces trois symboles sont liés par la relation :

$$i^2 = j^2 = k^2 = ijk = -1. \tag{I.4}$$

Cet ensemble est très utilisé dans le domaine couleur puisqu'il présente trois imaginaires, comme les trois canaux nécessaires à la couleur. Ainsi, en prenant une partie réelle (a) nulle, et en associant à chaque imaginaire un canal, nous pouvons effectuer des traitements vectoriels couleur :

$$I(x, y) = C_1(x, y)i + C_2(x, y)j + C_3(x, y)k, \tag{I.5}$$

C_i représentant chaque canal couleur. L'espace couleur utilisé n'est pas fixé, un espace RGB est souvent utilisé mais l'espace $CIELAB$ est plus approprié par l'orthogonalité des axes qui le composent.

La transformée de Fourier couleur

L'ensemble des quaternions a notamment été utilisé pour créer une transformée de Fourier vectorielle couleur. Elle se base sur la transformée de Fourier classique en $2D$ discret :

$$F(u, v) = \sum_{x,y \in I} \exp\left(-2\pi\omega(ux + vy)\right), \tag{I.6}$$

où ω représente l'imaginaire employé. Cependant, l'imaginaire ω à utiliser pour le calcul peut varier selon les auteurs et leurs intentions. Cela peut être simplement i, j, k ou une combinaison des trois. En RGB, pour obtenir une transformation inversible, il est recommandé de prendre $\omega = \frac{i+j+k}{\sqrt{3}}$. Dans $CIELAB$, $\omega = L$, correspondant à la direction L de l'espace, génère une transformée inversible.

La base de données artificielles fractales

L'objectif est de créer des images dont les couleurs ont été créées vectoriellement, qui sont similaires sans être identiques et qui suivent un modèle fractal. Pour cela, la transformée de Fourier est utilisée. Celle-ci implique une amplitude, une phase et un axe dépendants de la transformée utilisée. Cette base de données a été générée dans $CIELAB$ pour permettre une étude psychophysique des textures. Dans ce cas, ω a été pris égal à L.

La distribution couleur sans information spatiale sera donnée par une Gaussienne multivaluées $3D$. Puisque la Gaussienne reste de même type après transformée de Fourier, la loi Gaussienne choisie sera mesurée dans l'espace de Fourier. Faire varier la phase permet d'obtenir des images similaires mais non identiques. Une loi uniforme est choisie pour la sélection des valeurs de phase. L'amplitude de la transformée de Fourier est ensuite filtrée de façon à respecter la loi fractale choisie :

$$\|T(f)\|^2 \propto \frac{1}{f^{2H+1}}. \tag{I.7}$$

Une fois le signal fréquentiel créé, il suffit d'utiliser la transformée de Fourier inverse pour obtenir l'image couleur fractale souhaitée.

Bibliographie

Alaa E Abdel-Hakim and Aly A Farag. Csift : A sift descriptor with color invariant characteristics. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1978–1983. IEEE, 2006.

Motilal Agrawal, Kurt Konolige, and Morten Rufus Blas. Censure : Center surround extremas for real-time feature detection and matching. In *European Conference on Computer Vision*, pages 102–115. Springer, 2008.

Cuneyt Akinlar and Cihan Topal. Colored : Color edge and segment detection by edge drawing (ed). *Journal of Visual Communication and Image Representation*, 44 :82–94, 2017.

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

Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. Kaze features. In *European Conference on Computer Vision*, pages 214–227. Springer, 2012.

S. Alvarez and M. Vanrell. Texton theory revisited : A bag-of-words approach to combine textons. *Pattern Recogn*, 45(12) :4312 – 4325, 2012.

V. Arvis, C. Debain, M. Berducat, and A. Benassi. Generalization of the co-occurrence matrix for colour images : Application to colour texture segmentation. *Image Anal Estereol*, 23 :63–72, 2004.

Jaakko Astola, Petri Haavisto, and Yrjo Neuvo. Vector median filters. *Proceedings of the IEEE*, 78(4) : 678–689, 1990.

Mathieu Aubry, Bryan Russell, and Josef Sivic. Visual geo-localization of non-photographic depictions via 2d–3d alignment. In *Large-Scale Visual Geo-Localization*, pages 255–275. Springer, 2016.

Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. Surf : Speeded up robust features. In *European conference on computer vision*, pages 404–417. Springer, 2006.

Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3) :346–359, June 2008. ISSN 10773142. doi : 10.1016/j.cviu.2007.09.014.

Raquel Bello-Cerezo, Francesco Bianconi, Antonio Fernández, Elena González, and Francesco Di Maria. Experimental comparison of color spaces for material classification. *Journal of Electronic Imaging*, 25(6) :061406–061406, 2016.

Adnane Belmamoun, Mohammed El Hassouni, and Ahmed Hammouch. On selection and combination of relevant color components for edge detection. *Procedia Technology*, 17 :764–771, 2014.

Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré. Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2) :A1111–A1138, 2015.

James R Bergen and Edward H Adelson. Early vision and texture perception. *Nature*, 333(6171) :363–364, 1988.

Alexandre Bony and Christine Fernandez-Maloigne. Classification of painting techniques with color run-length matrices. In *Color and Imaging Conference*, volume 2016, pages 157–162. Society for Imaging Science and Technology, 2016.

Anna Bosch, Andrew Zisserman, and Xavier Muñoz. Scene classification using a hybrid generative/discriminative approach. *IEEE transactions on pattern analysis and machine intelligence*, 30(4) : 712–727, 2008.

Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3) :200–217, 1967.

Gertjan J Burghouts and Jan-Mark Geusebroek. Performance evaluation of local colour invariants. *Computer Vision and Image Understanding*, 113(1) :48–62, 2009.

W. Byeon, M. Liwicki, and T.M. Breuel. Texture classification using 2d lstm networks. *IEEE, 22nd Int. C. Pattern Recogn. (ICPR)*, 2014.

Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. Brief : Binary robust independent elementary features. *Computer Vision ECCV 2010*, pages 778–792, 2010.

John Canny. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 679–698, 1986.

Yang Cao, Changhu Wang, Liqing Zhang, and Lei Zhang. Edgel index for large-scale sketch-based image search. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 761–768. IEEE, 2011.

B. Caputo, E. Hayman, M. Fritz, and J.O. Eklundh. Classifying materials in the real world. *Image Vision Comput*, 28 :150–163, 2010.

Guillaume Carlier, Vincent Duval, Gabriel Peyré, and Bernhard Schmitzer. Convergence of entropic schemes for optimal transport and gradient flows. *SIAM Journal on Mathematical Analysis*, 49(2) : 1385–1418, 2017.

Philippe Carré, Patrice Denis, and Christine Fernandez-Maloigne. Spatial color image processing using clifford algebras : application to color active contour. *Signal, Image and Video Processing*, 8(7) : 1357–1372, 2014.

Thierry Carron. *Segmentations d'images couleur dans la base Teinte-Luminance-Saturation : approche numérique et symbolique*. PhD thesis, Chambéry, 1995.

Thierry Carron and Patrick Lambert. Color edge detector using jointly hue, saturation and intensity. In *Image Processing, 1994. Proceedings. ICIP-94., IEEE International Conference*, pages 977–981, 1994.

H. Chatoux, N. Richard, F. Lecellier, and C. Fernandez-Maloigne. Full-vector gradient for multispectral or multivariate images. *IEEE Transactions on Image Processing*, 28(5) :2228–2241, May 2019. ISSN 1057-7149. doi : 10.1109/TIP.2018.2883794.

Shu-Jie Chen, Hui-Liang Shen, Chunguang Li, and John H Xin. Normalized total gradient : a new measure for multispectral image registration. *IEEE Transactions on Image Processing*, 27(3) :1297–1310, 2018.

Xin Chen and Houjin Chen. A novel color edge detection algorithm in rgb color space. In *Signal Processing (ICSP), 2010 IEEE 10th International Conference on*, pages 793–796. IEEE, 2010.

Yi-An Chen, Chia-Hsin Chan, and Wen-Jiin Tsai. Creak : Color-based retina keypoint descriptor. In *Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)*, page 252. The Steering Committee of The World Congress in Computer Science, Computer …, 2016.

Pascal Chossat and Olivier Faugeras. Hyperbolic planforms in relation to visual edges and textures perception. *PLoS Computational Biology*, 5(12) :e1000625, 2009.

A Chu, Chandra M Sehgal, and James F Greenleaf. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, 11(6) :415–419, 1990.

David Connah and Graham D Finlayson. Using local binary pattern operators for colour constant image indexing. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2006, pages 60–64. Society for Imaging Science and Technology, 2006.

Jason J Corso and Gregory D Hager. Coherent regions for concise and stable image description. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 184–190. IEEE, 2005.

Harald Cramer. On the composition of elementary errors : second paper : statistical applications. *Scandinavian Actuarial Journal*, 1928(1) :141–180, 1928.

Aldo Cumani. Edge detection in multispectral images. *CVGIP : Graphical models and image processing*, 53(1) :40–51, 1991.

Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

Hilda Deborah, Noël Richard, Jon Yngve Hardeberg, and Christine Fernandez-Maloigne. Assessment protocols and comparison of ordering relations for spectral image processing. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(4) :1253–1265, 2018.

Patrice Denis, Philippe Carre, and Christine Fernandez-Maloigne. Spatial and spectral quaternionic approaches for colour images. *Computer Vision and Image Understanding*, 107(1) :74–87, 2007.

Silvano Di Zenzo. A note on the gradient of a multi-image. *Computer vision, graphics, and image processing*, 33(1) :116–125, 1986.

Leidy P Dorado-Munoz, Miguel Velez-Reyes, Amit Mukherjee, and Badrinath Roysam. A vector sift detector for interest point detection in hyperspectral imagery. *IEEE transactions on Geoscience and Remote sensing*, 50(11) :4521–4533, 2012.

Alexandru Drimbarean and Paul F Whelan. Experiments in colour texture analysis. *Pattern recognition letters*, 22(10) :1161–1167, 2001.

Fritz Ebner. *Derivation and modelling hue uniformity and development of the IPT color space*. PhD thesis, 1998.

Matthias Joachim Ehrhardt and Simon R Arridge. Vector-valued image processing by parallel level sets. *IEEE Transactions on Image Processing*, 23(1) :9–18, 2014.

Adrian N Evans and Xin U Liu. A morphological gradient approach to color edge detection. *Image Processing, IEEE Transactions on*, 15(6) :1454–1463, 2006.

Ivo Everts, Jan C Van Gemert, and Theo Gevers. Evaluation of color spatio-temporal interest points for human action recognition. *IEEE Transactions on Image Processing*, 23(4) :1569–1580, 2014.

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

Kenneth Falconer. *Fractal geometry : mathematical foundations and applications*. John Wiley & Sons, 2004.

Bin Fan, Fuchao Wu, and Zhanyi Hu. Rotationally invariant descriptors using intensity order pooling. *IEEE transactions on pattern analysis and machine intelligence*, 34(10) :2031–2045, 2012.

Julien Fauqueur, Nick Kingsbury, and Ryan Anderson. Multiscale keypoint detection using the dual-tree complex wavelet transform. In *Image Processing, 2006 IEEE International Conference on*, pages 1625–1628. IEEE, 2006.

Lu Feng, Zhuangzhi Wu, and Xiang Long. Fast image diffusion for feature detection and description. *International Journal of Computer Theory and Engineering*, 8(1) :58, 2016.

Mariela Atausinchi Fernandez, Rubens M Lopes, and Nina ST Hirata. Image segmentation assessment from the perspective of a higher level task. In *Graphics, Patterns and Images (SIBGRAPI), 2015 28th SIBGRAPI Conference on*, pages 111–118. IEEE, 2015.

Joao B Florindo, Gabriel Landini, and Odemir M Bruno. Three-dimensional connectivity index for texture recognition. *Pattern Recognition Letters*, 84 :239–244, 2016a.

Joao Batista Florindo, Lucas Assirati, and Odemir Martinez Bruno. Locally enhancing fractal descriptors by using the non-additive entropy. *Pattern Recognition Letters*, 70 :32–37, January 2016b. ISSN 01678655. doi : 10.1016/j.patrec.2015.11.002.

Lidia Forlenza, Patrick Carton, Domenico Accardo, Giancarmine Fasano, and Antonio Moccia. Real time corner detection for miniaturized electro-optical sensors onboard small unmanned aerial systems. *Sensors*, 12(1) :863–877, 2012.

Per-Erik Forssén. Maximally stable colour regions for recognition and matching. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.

William T. Freeman and Edward H Adelson. The design and use of steerable filters. *IEEE transactions on pattern analysis and machine intelligence*, (9) :891–906, 1991.

William T Freeman and Michal Roth. Orientation histograms for hand gesture recognition. In *International workshop on automatic face and gesture recognition*, volume 12, pages 296–301, 1995.

Mary M Galloway. Texture analysis using grey level run lengths. *NASA STI/Recon Technical Report N*, 75, 1974.

Steffen Gauglitz, Tobias Höllerer, and Matthew Turk. Evaluation of Interest Point Detectors and Feature Descriptors for Visual Tracking. *International Journal of Computer Vision*, 94(3) :335–360, September 2011. ISSN 0920-5691, 1573-1405. doi : 10.1007/s11263-011-0431-5.

Baris Gecer, George Azzopardi, and Nicolai Petkov. Color-blob-based cosfire filters for object recognition. *Image and Vision Computing*, 57 :165–174, 2017.

J-M Geusebroek, Rein Van den Boomgaard, Arnold W. M. Smeulders, and Hugo Geerts. Color invariance. *IEEE Transactions on Pattern analysis and machine intelligence*, 23(12) :1338–1350, 2001.

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

Jacob Goldberger, Shiri Gordon, Hayit Greenspan, et al. An efficient image similarity measure based on approximations of kl-divergence between two gaussian mixtures. In *ICCV*, volume 3, pages 487–493, 2003.

Rafael C Gonzalez, Richard E Woods, et al. *Digital image processing 2d edition*. Prentice hall Upper Saddle River, NJ, 2001.

Calvin C Gotlieb and Herbert E Kreyszig. Texture descriptors based on co-occurrence matrices. *Computer Vision, Graphics, and Image Processing*, 51(1) :70–86, 1990.

Valérie Gouet, Philippe Montesinos, Rachid Deriche, and Danielle Pelé. Evaluation de détecteurs de points d'intérêt pour la couleur. In *12eme Congres Francophone AFRIF–AFIA de Reconnaissance des Formes et Intelligence Artificielle*, pages 257–266, 2000.

Robert O Green, Michael L Eastwood, Charles M Sarture, Thomas G Chrien, Mikael Aronsson, Bruce J Chippendale, Jessica A Faust, Betina E Pavri, Christopher J Chovit, Manuel Solis, et al. Imaging spectroscopy and the airborne visible/infrared imaging spectrometer (aviris). *Remote sensing of environment*, 65(3) :227–248, 1998.

Zhenhua Guo, Lei Zhang, and David Zhang. A completed modeling of local binary pattern operator for texture classification. *IEEE Transactions on Image Processing*, 19(6) :1657–1663, 2010.

Richard W Hamming. Error detecting and error correcting codes. *Bell Labs Technical Journal*, 29(2) : 147–160, 1950.

Allan G Hanbury and Jean Serra. Morphological operators on the unit circle. *IEEE Transactions on Image Processing*, 10(12) :1842–1850, 2001.

Robert M Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5) : 786–804, 1979.

Robert M Haralick, Karthikeyan Shanmugam, et al. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621, 1973.

Chris Harris and Mike Stephens. A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Citeseer, 1988.

Markku Hauta-Kasari, Jussi Parkkinen, T Jaaskelainen, and Reiner Lenz. Generalized co-occurrence matrix for multispectral texture analysis. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on*, volume 2, pages 785–789. IEEE, 1996.

Jared Heinly, Enrique Dunn, and Jan-Michael Frahm. Comparative evaluation of binary features. In *Computer Vision–ECCV 2012*, pages 759–773. Springer, 2012.

Antti Hietanen, Jukka Lankinen, Joni-Kristian KÃ¤mÃ¤rÃ¤inen, Anders Glent Buch, and Norbert KrÃ¼ger. A comparison of feature detectors and descriptors for object class matching. *Neurocomputing*, December 2015. ISSN 09252312. doi : 10.1016/j.neucom.2015.08.106.

Radu Horaud and Olivier Monga. *Vision par ordinateur : outils fondamentaux*. Editions Hermès, 1995.

Jing Huang, S Ravi Kumar, Mandar Mitra, Wei-Jing Zhu, and Ramin Zabih. Image indexing using color correlograms. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 762–768. IEEE, 1997.

Min Huang, Haoxue Liu, Guihua Cui, M Ronnier Luo, and Manuel Melgosa. Evaluation of threshold color differences using printed samples. *JOSA A*, 29(6) :883–891, 2012.

Keigo Ishikura, Naoto Kurita, Damon M Chandler, and Gosuke Ohashi. Saliency detection based on multiscale extrema of local perceptual color differences. *IEEE Transactions on Image Processing*, 27 (2) :703–717, 2018.

Bernd Jähne. *Digital image processing : 6th revised and extended edition*. Springer, 2005.

Alex Pappachen James. Edge detection for pattern recognition : a survey. *International Journal of Applied Pattern Recognition*, 3(1) :1–21, 2016.

Carlos M Jarque and Anil K Bera. A test for normality of observations and regression residuals. *International Statistical Review/Revue Internationale de Statistique*, pages 163–172, 1987.

Hela Jebali, Noel Richard, Hermine Chatoux, and Mohamed Naouai. Relocated colour contrast occurrence matrix and adapted similarity measure for colour texture retrieval. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, pages 609–619. Springer, 2018.

Lianghai Jin, Hong Liu, Xiangyang Xu, and Enmin Song. Improved direction estimation for di zenzo's multichannel image gradient operator. *Pattern Recognition*, 45(12) :4300–4311, 2012.

Lianghai Jin, Enmin Song, Lei Li, and Xiang Li. A quaternion gradient operator for color image edge detection. In *Image Processing (ICIP), 2013 20th IEEE International Conference on*, pages 3040–3044. IEEE, 2013.

Bela Julesz. Visual pattern discrimination. *IRE transactions on Information Theory*, 8(2) :84–92, 1962.

Bela Julesz. Texture and visual perception. *Scientific American*, 212(2) :38–49, 1965.

Bela Julesz and James R Bergen. Human factors and behavioral science : Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal*, 62(6) : 1619–1645, 1983.

Bela Julész, Edgar N Gilbert, and Jonathan D Victor. Visual discrimination of textures with identical third-order statistics. *Biological Cybernetics*, 31(3) :137–140, 1978.

E.G. Karakasis, A. Amanatiadis, A. Gasteratos, and S.A. Chatzichristofis. Image moment invariants as local features for content based image retrieval using the Bag-of-Visual-Words model. *Pattern Recognition Letters*, 55 :22–27, April 2015. ISSN 01678655. doi : 10.1016/j.patrec.2015.01.005.

Yan Ke and Rahul Sukthankar. Pca-sift : A more distinctive representation for local image descriptors. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 2, pages II–506. IEEE, 2004.

Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2) :81–93, 1938.

Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Michael Felsberg, and Jorma Laaksonen. Compact color-texture description for texture classification. *Pattern Recognition Letters*, 51 :16–22, 2015.

Haris Ahmad Khan, Sofiane Mihoubi, Benjamin Mathon, Jean-Baptiste Thomas, and Jon Yngve Hardeberg. Hytexila : High resolution visible and near infrared hyperspectral texture images. *Sensors*, 18(7) :2045, 2018. ISSN 1424-8220. doi : 10.3390/s18072045. URL http://www.mdpi.com/1424-8220/18/7/2045.

NA Khanina, EV Semeikina, and DV Yurin. Scale-space color blob and ridge detection. *Pattern Recognition and Image Analysis*, 22(1) :221–227, 2012.

Ron Kimmel, Nir Sochen, and Ravi Malladi. From high energy physics to low level vision. *Scale-Space Theory in Computer Vision*, pages 236–247, 1997.

Jan J Koenderink and Andrea J van Doorn. Representation of local geometry in the visual system. *Biological cybernetics*, 55(6) :367–375, 1987.

Andrei N Kolmogorov. Wienersche spiralen und einige andere interessante kurven in hilbertscen raum, cr (doklady). *Acad. Sci. URSS (NS)*, 26 :115–118, 1940.

Andreas Koschan. A comparative study on color edge detection. In *Proceedings of the 2nd Asian Conference on Computer Vision*, volume 3, pages 574–578, 1995.

Andreas Koschan and Mongi Abidi. Detection and classification of edges in color images. *IEEE Signal Processing Magazine*, 22(1) :64–73, 2005.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1) :79–86, 1951.

Roland Kwitt and Peter Meerwald. Salzburg texture image database. available online : http ://www.wavelab.at/sources/stex/.

Michael S Landy and Norma Graham. 73 visual perception of texture. *The visual neurosciences*, page 1106, 2004.

Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. A sparse texture representation using affine-invariant regions. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–319. IEEE, 2003.

Audrey Ledoux, Noël Richard, and Anne-Sophie Capelle-Laizé. The fractal estimator : A validation criterion for the colour mathematical morphology. In *Conference on Colour in Graphics, Imaging, and Vision*, pages 206–210. Society for Imaging Science and Technology, 2012.

Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart. BRISK : Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE, 2011.

Gil Levi and Tal Hassner. LATCH : Learned Arrangements of Three Patch Codes. *arXiv preprint arXiv :1501.03719*, 2015.

Olivier Lezoray, Cyril Meurie, and Abderrahim Elmoataz. Graph-based ordering scheme for color image filtering. *International Journal of Image and Graphics*, 8(03) :473–493, 2008.

Tony Lindeberg. Scale-space theory : A basic tool for analyzing structures at different scales. *Journal of applied statistics*, 21(1-2) :225–270, 1994.

Tony Lindeberg. Feature detection with automatic scale selection. *International journal of computer vision*, 30(2) :79–116, 1998.

Ce Liu and Hueng-Yeung Shum. Kullback-leibler boosting. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2003.

Zhen-Bao Liu, Shu-Hui Bu, Kun Zhou, Shu-Ming Gao, Jun-Wei Han, and Jun Wu. A survey on partial retrieval of 3d shapes. *Journal of Computer Science and Technology*, 28(5) :836–851, 2013.

Cédric Loosli, François Lecellier, Stéphanie Jehan-Besson, and Jonas Koko. A color-based selective and interactive filter using weighted tv. In *International Conference on Computer Analysis of Images and Patterns*, pages 315–323. Springer, 2013.

David G. Lowe. Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee, 1999.

David G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004.

L Luccheseyz and SK Mitray. Color image segmentation : A state-of-the-art survey. *Proceedings of the Indian National Science Academy (INSA-A)*, 67(2) :207–221, 2001.

Rastislav Lukac and Konstantinos N Plataniotis. *Color image processing : methods and applications*. CRC press, 2006.

T Mäenpää, Timo Ojala, M Pietikäinen, and Maricor Soriano. Robust texture classification by subsets of local binary patterns. In *Proceedings of the 15th International Conference on Pattern Recognition*, volume 3, pages 939–942, 2000.

Topi Mäenpää and Matti Pietikäinen. Classification with color and texture : jointly or separately ? *Pattern recognition*, 37(8) :1629–1640, 2004.

Baptiste Magnier, Hasan Abdulrahman, and Philippe Montesinos. A review of supervised edge detection evaluation methods and an objective comparison of filtering gradient computations using hysteresis thresholds. *Journal of Imaging*, 4(6) :74, 2018.

Prasanta Chandra Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936.

Elmar Mair, Gregory D. Hager, Darius Burschka, Michael Suppa, and Gerhard Hirzinger. Adaptive and generic corner detection based on the accelerated segment test. In *Computer Vision ECCV 2010*, pages 183–196. Springer, 2010.

Ayoub Jafari Malekabadi, Mehdi Khojastehpour, and Bagher Emadi. A comparative evaluation of combined feature detectors and descriptors in different color spaces for stereo image matching of tree. *Scientia Horticulturae*, 228 :187–195, 2018.

Ahmed Drissi El Maliani, Mohammed El Hassouni, Y. Berthoumieu, and D. Aboutajdine. Color texture classification method based on a statistical multi-model and geodesic distance. *Journal of Visual Communication and Image Representation, Elsevier*, 2014.

Bappaditya Mandal, Zhikai Wang, Liyuan Li, and Ashraf A. Kassim. Performance evaluation of local descriptors and distance measures on benchmarks and first-person-view videos for face identification. *Neurocomputing*, November 2015. ISSN 09252312. doi : 10.1016/j.neucom.2015.07.121.

Benoit B Mandelbrot and John W Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM review*, 10(4) :422–437, 1968.

Kanti V Mardia. Measures of multivariate skewness and kurtosis with applications. *Biometrika*, 57 (3) :519–530, 1970.

Foteini Markatopoulou, Nikiforos Pittaras, Olga Papadopoulou, Vasileios Mezaris, and Ioannis Patras. A study on the use of a binary local descriptor and color extensions of local descriptors for video concept detection. In *International Conference on Multimedia Modeling*, pages 282–293. Springer, 2015.

RA Martínez, Noël Richard, and C Fernandez. Alternative to colour feature classification using colour contrast ocurrence matrix. In *Twelfth International Conference on Quality Control by Artificial Vision 2015*, volume 9534, page 953405. International Society for Optics and Photonics, 2015.

Pedro Martins, Paulo Carvalho, and Carlo Gatta. On the completeness of feature-driven maximally stable extremal regions. *Pattern Recognition Letters*, 74 :9–16, April 2016. ISSN 01678655. doi : 10.1016/j.patrec.2016.01.003.

Jiri Matas, Ondrej Chum, Martin Urban, and TomÃ¡s Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10) :761–767, 2004.

Jiri George Matas, Dimitri Koubaroulis, and Josef Kittler. Colour image retrieval and object recognition using the multimodal neighbourhood signature. In *European Conference on Computer Vision*, pages 48–64. Springer, 2000.

John Reidar Mathiassen, Amund Skavhaug, and Ketil Bø. Texture similarity measure using kullback-leibler divergence between gamma distributions. In *European conference on computer vision*, pages 133–147. Springer, 2002.

William McIlhagga and Kathy T Mullen. Evidence for chromatic edge detectors in human vision using classification images. *Journal of vision*, 18(9) :8–8, 2018.

José Mennesson, Christophe Saint-Jean, and Laurent Mascarilla. Color fourier–mellin descriptors for image recognition. *Pattern Recognition Letters*, 40 :27–35, 2014.

Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10) :1993, 2015.

K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool. A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1-2) : 43–72, November 2005. ISSN 0920-5691, 1573-1405. doi : 10.1007/s11263-005-3848-x.

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.

Krystian Mikolajczyk and Cordelia Schmid. An affine invariant interest point detector. In *Computer Vision ECCV 2002*, pages 128–142. Springer, 2002.

Krystian Mikolajczyk and Cordelia Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1) :63–86, 2004.

Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE transactions on pattern analysis and machine intelligence*, 27(10) :1615–1630, 2005.

Ondrej Miksik and Krystian Mikolajczyk. Evaluation of local detectors and descriptors for fast feature matching. In *Pattern Recognition (ICPR), 2012 21st International Conference on*, pages 2681–2684. IEEE, 2012.

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) :3–27, April 2004. ISSN 1077-3142. doi : 10.1016/j.cviu.2003.10.011.

MIT. Vision texture, vision and modeling group, cambridge. available online : http ://vismod.media.mit.edu/pub/vistex/, 1995.

Philippe Montesinos, Valérie Gouet, and Rachid Deriche. Differential invariants for color images. In *icpr*, page 838. IEEE, 1998.

Hans P Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical report, STANFORD UNIV CA DEPT OF COMPUTER SCIENCE, 1980.

Amit Mukherjee, Miguel Velez-Reyes, and Badrinath Roysam. Interest points for hyperspectral image data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(3) :748–760, 2009.

Huu-Tuan Nguyen and Alice Caplier. Local patterns of gradients for face recognition. *IEEE Transactions on Information Forensics and Security*, 10(8) :1739–1751, 2015.

Vu-Lam Nguyen, Ngoc-Son Vu, Hai-Hong Phan, and Philippe-Henri Gosselin. An integrated descriptor for texture classification. *23rd IEEE International Conference on Pattern Recognition (ICPR)*, 2016.

CL Novak and SA Shafer. Color edge detection. In *Proc. DARPA Image Understanding Workshop*, volume 1, pages 35–37, 1987.

Timo Ojala, Matti Pietikainen, and David Harwood. Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A : Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582–585. IEEE, 1994.

Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7) :971–987, 2002.

Ville Ojansivu and Janne Heikkilä. Blur insensitive texture classification using local phase quantization. In *International conference on image and signal processing*, pages 236–243. Springer, 2008.

Karen Panetta, Chen Gao, Sos Agaian, and Shahan Nercessian. A new reference-based edge map quality measure. *IEEE Transactions on Systems, Man, and Cybernetics : Systems*, 46(11) :1505–1517, 2016.

Unsang Park, Jongseung Park, and Anil K Jain. Robust keypoint detection using higher-order scale space derivatives : application to image retrieval. *IEEE Signal Processing Letters*, 21(8) :962–965, 2014.

Karl Pearson. X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302) :157–175, 1900.

Michal Perdoch, Jiri Matas, and Stepan Obdrzalek. Stable affine frames on isophotes. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8. IEEE, 2007.

Pietro Perona and Jitendra Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on pattern analysis and machine intelligence*, 12(7) :629–639, 1990.

Minh-Tan Pham, Grégoire Mercier, and Lionel Bombrun. Color texture image retrieval based on local extrema features and riemannian distance. *Journal of Imaging*, 3(4), 2017.

A. Porebski, N. Vandenbroucke, and D. Hamad. Lbp histogram selection for supervised color texture classification. *ICIP*, pages 3239–3243, 2013.

W.K. Pratt. *Digital image processing, 3rd edition*. Wiley-interscience publication. Wiley, 2001.

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

Edoardo Provenzi, Julie Delon, Yann Gousseau, and Baptiste Mazin. On the second order spatiochromatic structure of natural images. *Vision research*, 120 :22–38, 2016.

AK Qin and David A Clausi. Multivariate image segmentation using semantic region growing with adaptive edge penalty. *IEEE Transactions on Image Processing*, 19(8) :2157–2170, 2010.

Taha H Rassem and Bee Ee Khoo. Performance evaluation of new colour histogram-based interest point detectors. *Multimedia Tools and Applications*, 74(24) :11357–11398, 2015.

Alfréd Rényi. On measures of entropy and information. Technical report, HUNGARIAN ACADEMY OF SCIENCES Budapest Hungary, 1961.

N Richard, M Ivanovici, and A Bony. Toward a metrology for non-uniform surface using the complexity notion. In *4th CIE Expert Symposium on Colour and Visual Appearance, Czech Republic, Prague*, pages 40–50, 2016a.

N. Richard, R.A. Martínez, and C. Fernandez. Colour local pattern : a texture feature for colour images. *Journal of the International Colour Association*, 16 :56–68, 2016b.

Noël Richard, David Helbert, Christian Olivier, and Martin Tamisier. Pseudo-divergence and bidimensional histogram of spectral differences for hyperspectral image processing. *Journal of Imaging Science & Technology*, 60(5), 2016c.

Jean-Francois Rivest, Pierre Soille, and Serge Beucher. Morphological gradients. *Journal of Electronic Imaging*, 2(4) :326–336, 1993.

Lawrence Gilman Roberts. *Machine perception of three-dimensional solids*. PhD thesis, Massachusetts Institute of Technology, 1963.

Ruth Rosenholtz. Texture perception. *Oxford handbook of perceptual organization*, pages 167–186, 2014.

Paul L. Rosin. Measuring corner properties. *Computer Vision and Image Understanding*, 73(2) :291–307, 1999.

Edward Rosten and Tom Drummond. Machine learning for high-speed corner detection. In *Computer Vision ECCV 2006*, pages 430–443. Springer, 2006.

Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB : an efficient alternative to SIFT or SURF. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE, 2011.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. A metric for distributions with applications to image databases. In *Computer Vision, 1998. Sixth International Conference on*, pages 59–66. IEEE, 1998.

Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2) :99–121, 2000.

Mark Ruzon, Carlo Tomasi, et al. Edge, junction, and corner detection using color distributions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11) :1281–1295, 2001.

Aurora Sáez, Carlos S Mendoza, Begoña Acha, and Carmen Serrano. Development and evaluation of perceptually adapted colour gradients. *IET Image Processing*, 7(4) :355–363, 2013.

Faten Sandid and Ali Douik. Robust color texture descriptor for material recognition. *Pattern Recognition Letters*, 80 :15–23, 2016.

Guillermo Sapiro. Color snakes. *Computer Vision and Image Understanding*, 68(2) :247–253, 1997.

Dietmar Saupe. Algorithms for random fractals. In *The science of fractal images*, pages 71–136. Springer, 1988.

Frederik Schaffalitzky and Andrew Zisserman. Multi-view matching for unordered image sets, or how do i organize my holiday snaps ? In *Computer Vision ECCV 2002*, pages 414–431. Springer, 2002.

Cordelia Schmid. *Appariement d'images par invariants locaux de niveaux de gris. Application à l'indexation d'une base d'objets*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 1996.

H. Schulz-Mirbach. Constructing invariant features by averaging techniques. pages 387–390, Jerusalem, Israel, 25-28 octobre 1994.

Nicu Sebe, Theo Gevers, Joost Van de Weijer, and Sietse Dijkstra. Corner detectors for affine invariant salient regions : is color important ? In *International Conference on Image and Video Retrieval*, pages 61–71. Springer, 2006.

Jianbo Shi and Carlo Tomasi. Good features to track. Technical report, Cornell University, 1993.

Huang-Chia Shih and En-Rui Liu. Automatic reference color selection for adaptive mathematical morphology and application in image segmentation. *IEEE Transactions on Image Processing*, 25(10) : 4665–4676, 2016.

P Shivakumara, DS Guru, and HT Basavaraju. Color and gradient features for text segmentation from video frames. In *Multimedia Processing, Communication and Computing Applications*, pages 267–278. Springer, 2013.

Raju Shrestha. Simulating colour vision deficiency from a spectral image. *Studies in health technology and informatics*, 229 :392–401, 2016.

S.Mangijao Singh and K. Hemachandran. Content-based image retrieval using color moment and gabor texture feature. In *IJCSI International Journal of Computer Scienc*, volume 9, pages 299–309, September 2012.

John R Smith and Shih-Fu Chang. Tools and techniques for color image retrieval. In *Storage and Retrieval for Still Image and Video Databases IV*, volume 2670, pages 426–438. International Society for Optics and Photonics, 1996.

Stephen M. Smith and J. Michael Brady. SUSAN : a new approach to low level image processing. *International journal of computer vision*, 23(1) :45–78, 1997.

Irwin Sobel and Gary Feldman. A 3x3 isotropic gradient operator for image processing. *the Stanford Artificial Intelligence Project*, 1968.

Pierre Soille. *Morphological image analysis : principles and applications*. Springer, 2002.

Yang Song, Qing Li, Dagan Feng, Ju Jia Zou, and Weidong Ca. Texture image classification with discriminative neural networks. *Ann. Math. Statist*, 2(4) :367–377, 2016.

Raphaël Soulard, Philippe Carré, and Christine Fernandez-Maloigne. Vector extension of monogenic wavelets for geometric representation of color images. *IEEE Transactions on Image Processing*, 22(3) : 1070–1083, 2013.

Christoph Strecha, Wolfgang Von Hansen, Luc Van Gool, Pascal Fua, and Ulrich Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. Ieee, 2008.

Smaranika Subhasini and Manglem Singh. Color image edge detection : A survey. 2017.

Michael J Swain and Dana H Ballard. Color indexing. *International journal of computer vision*, 7(1) : 11–32, 1991.

Saman Tahery and Mark S Drew. A novel colour hessian and its applications. *Electronic Imaging*, 2017 (18) :171–176, 2017.

Xiaoou Tang. Texture information in run-length matrices. *IEEE transactions on image processing*, 7(11) : 1602–1609, 1998.

Yuliya Tarabalka, Jocelyn Chanussot, and Jon Atli Benediktsson. Segmentation and classification of hyperspectral images using watershed transformation. *Pattern Recognition*, 43(7) :2367–2379, 2010.

Kasim Terzić, João MF Rodrigues, and JM Hans du Buf. Bimp : A real-time biological model of multi-scale keypoint detection in v1. *Neurocomputing*, 150 :227–237, 2015.

Tian Tian, Yun Zhang, Hao Dou, and Hengjian Tong. Land-use classification with biologically inspired color descriptor and sparse coding spatial pyramid matching. *Multimedia Tools and Applications*, 76(21) :22943–22958, 2017.

E. Tola, V. Lepetit, and P. Fua. DAISY : An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5) :815–830, May 2010. ISSN 0162-8828. doi : 10.1109/TPAMI.2009.77.

A Trémeau, A au, C Fernandez-Maloigne, and P Bonton. *éditeurs. Image numérique couleur–De l'acquisition au Traitement*. Dunod, janvier, 2004.

Tinne Tuytelaars and Luc Van Gool. Matching widely separated views based on affine invariant regions. *International journal of computer vision*, 59(1) :61–85, 2004.

Tinne Tuytelaars, Krystian Mikolajczyk, et al. Local invariant feature detectors : a survey. *Foundations and trends® in computer graphics and vision*, 3(3) :177–280, 2008.

Ranjith Unnikrishnan and Martial Hebert. Extracting scale and illuminant invariant regions through color. 2006.

Michael Unser. Sum and difference histograms for texture classification. *IEEE transactions on pattern analysis and machine intelligence*, (1) :118–125, 1986.

Koen van de Sande, Theo Gevers, and Cees Snoek. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(9) :1582–1596, 2010.

Koen EA van de Sande, Theo Gevers, and Cees GM Snoek. Color descriptors for object category recognition. In *Conference on Colour in Graphics, Imaging, and Vision*, volume 2008, pages 378–381. Society for Imaging Science and Technology, 2008.

Joost van de Weijer and Th Gevers. Tensor based feature detection for color images. In *Color and Imaging Conference*, volume 2004, pages 100–105. Society for Imaging Science and Technology, 2004.

Joost van de Weijer and Cordelia Schmid. Coloring local feature extraction. In *European conference on computer vision*, pages 334–348. Springer, 2006.

Luc van Gool, Theo Moons, and Dorin Ungureanu. Affine/photometric invariants for planar intensity patterns. In *European Conference on Computer Vision*, pages 642–651. Springer, 1996.

Richard F Voss. Fractals in nature : from characterization to simulation. In *The science of fractal images*, pages 21–70. Springer, 1988.

Fu-Ping Wang and Peng-Lang Shui. Noise-robust color edge detector using gradient matrix and anisotropic gaussian directional derivative matrix. *Pattern Recognition*, 52 :346–357, 2016.

Qing Wang, Sanjeev R Kulkarni, and Sergio Verdú. Divergence estimation for multidimensional densities via k-nearest-neighbor distances. *IEEE Transactions on Information Theory*, 55(5) :2392–2405, 2009.

Zhenhua Wang, Bin Fan, and Fuchao Wu. Local intensity order pattern for feature description. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 603–610. IEEE, 2011.

Rolf P Würtz and Tino Lourens. Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image and vision computing*, 18(6-7) :531–541, 2000.

S. Xiaoyan, C. Shao-Hui, L. Jiang, and M. Frederic. Automatic diagnosis for prostate cancer using run-length matrix method. *Medical Imaging. Procceding of SPIE*, 7260, 2009.

Changda Xing, Zhisheng Wang, and Chong Dong. Fusion of infrared and visible images with propagation filtering. *Infrared Physics & Technology*, 94 :232 – 243, 2018. ISSN 1350-4495. doi : https://doi.org/10.1016/j.infrared.2018.09.016. URL http://www.sciencedirect.com/science/article/pii/S1350449518300513.

P-T Yap and R Paramesran. Content-based image retrieval using legendre chromaticity distribution moments. *IEE Proceedings-Vision, Image and Signal Processing*, 153(1) :17–24, 2006.

Zhenqiang Ying, Ge Li, Sixin Wen, and Guozhen Tan. Orgb : Offset correction in rgb color space for illumination-robust image processing. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 1557–1561. IEEE, 2017.

Oulun Yliopisto. University of oulu, outex texture database. http ://www.outex.oulu.fi/index.php ?page=classification.

Lin Zhang, Zhiqiang Zhou, and Hongyu Li. Binary gabor pattern : An efficient and robust descriptor for texture classification. In *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pages 81–84. Ieee, 2012.

Andrew Ziegler, Eric Christiansen, David Kriegman, and Serge J. Belongie. Locally uniform comparison image descriptor. In *Advances in Neural Information Processing Systems*, pages 1–9, 2012.