



HAL
open science

Modélisation statistique et classification par apprentissage pour la qualité des images

Christophe Charrier

► **To cite this version:**

Christophe Charrier. Modélisation statistique et classification par apprentissage pour la qualité des images. Traitement des images [eess.IV]. Université de Caen, 2011. tel-01073869

HAL Id: tel-01073869

<https://theses.hal.science/tel-01073869>

Submitted on 10 Oct 2014

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

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

UNIVERSITÉ DE CAEN BASSE NORMANDIE
U.F.R. de Sciences
ÉCOLE DOCTORALE SIMEM

Mémoire présenté par

M. Christophe CHARRIER

et soutenu le

21 novembre 2011

en vue de l'obtention de l'

Habilitation à Diriger les Recherches

Titre :

**MODÉLISATION STATISTIQUE ET CLASSIFICATION PAR
APPRENTISSAGE POUR LA QUALITÉ DES IMAGES**

MEMBRES du JURY :

M. Christophe ROSENBERGER	Professeur	ENSICAEN	<i>(Directeur de HDR)</i>
Mme Christine FERNANDEZ-MALOIGNE	Professeur	Université de Poitiers	<i>(Rapporteur)</i>
M. Patrick LE CALLET	Professeur	École polytechnique de Nantes	<i>(Rapporteur)</i>
M. Azzedine BEGHDAI	Professeur	Université de Paris 13	<i>(Rapporteur)</i>
Mme Marinette REVENU	Professeur Émérite	ENSICAEN	<i>(Examineur)</i>
M. Laurence T. MALONEY	Professeur	Université de New-York	<i>(Examineur)</i>

Remerciements

Ce manuscrit n'aurait pas pu voir le jour sans les collaborations scientifiques développées depuis le début de mes activités de recherche avec mes collègues, qu'ils soient enseignants-chercheurs, vacataires ou doctorants.

Je tiens en premier lieu à remercier les Pr. Christine Fernandez-Maloigne, Patrick Le Callet et Azzedine Beghdadi pour l'honneur qu'ils m'ont fait en acceptant de rapporter ce mémoire. Je tenais également à remercier Christine Fernandez-Maloigne pour son accueil chaleureux au sein du laboratoire XLIM département SIC dans le cadre d'une délégation CNRS obtenue au titre de l'année 2010-2011.

J'adresse également mes remerciements à Larry pour avoir accepté d'être membre de ce jury ainsi que pour les différentes conversations fructueuses que nous avons eues tout au long de ces années écoulées.

Je remercie également Marinette Revenu pour avoir accepté d'être membre de ce jury et m'avoir manifesté ses encouragements.

Mes remerciements s'adressent également à Christophe Rosenberger pour m'avoir convaincu de passer cette habilitation et pour avoir endossé le rôle de directeur de recherche. Sa jovialité et son dynamisme m'ont porté durant la phase d'écriture : « Quand est-ce que je reçois la première version de ton manuscrit ? ». Vous ne pouvez pas savoir comment cette phrase interrogative d'apparence anodine peut se révéler être un véritable supplice chinois lorsqu'elle tourne en boucle !! :o) Du coup, pour ne plus l'entendre vous avez deux solutions : vous partez loin, très loin ou vous passez à l'étape de rédaction. J'ai bien tenté la première approche en partant au Canada, mais sans succès, puisque j'ai quand même fini par appliquer la seconde solution en écrivant ce manuscrit. Quel plaisir à mes oreilles de ne plus entendre ce « doux » refrain...

Que la direction de l'IUT Cherbourg-Manche et plus particulièrement Philippe Makany, directeur de l'institut, trouve ici toute ma gratitude pour avoir mis à la disposition des enseignants chercheurs du département SRC de l'IUT Cherbourg Manche (qui se situe sur le site délocalisé de Saint-Lô) les moyens matériels nécessaires à la continuité et au développement de nos activités de recherche. Sans cet appui, mes activités et les collaborations développées n'auraient pu être viables puisque que le laboratoire de recherche GREYC auquel je suis rattaché depuis janvier 2008 se situe à plus de 60 km de Saint-Lô où je dispense la quasi totalité de mes enseignements. Il a également accepté d'appuyer ma demande locale de CRCT pour l'année 2008-2009 ainsi que ma demande de délégation CNRS obtenue auprès du laboratoire XLIM-SIC de l'Université de Poitiers au titre de l'année 2010-2011.

Je tenais également à remercier mes collègues Gilles Lebrun, Michèle Saad ainsi que Mo-hamad El Abed (qui ont été ou qui sont encore au doctorat) sans qui de nombreux travaux

n'auraient pu voir le jour et seraient encore et toujours remisés au fin fond d'une boîte à idées. Un grand merci également à Baptiste Hémerly et Romain Giot (de l'équipe des biomen) pour leur force rouge et verte et leur bonne humeur.

Que mes collègues de l'Université de Sherbrooke (en autre Pierre-Marc, Maxime, Olivier, ...) soient également remerciés pour avoir accepté de m'accueillir pendant deux ans comme professeur invité au sein du département informatique de la faculté des sciences. Ce qui m'a permis de redécouvrir les joies de se sucrer le bec, de pelleter l'entrée avant que la charrue ne passe (et c'est tout un art!! si si!), la bataille avec les moustiques et autres bibites sur la terrasse le soir venu, la visite d'une famille de marmotte à tous les soirs au pied de la terrasse...

Je tenais également à remercier mes collègues de l'IUT Cherbourg Manche sur le site de Saint-Lô, et notamment Olivier Lézoray (breton naturalisé normand : fais gaffe à la double nationalité. C'est mal vu par les temps qui courent!), Jérôme Clouet (digne descendant de Gaston Lagaffe : merci pour tes histoires vécues!!), Anthony Paul (le lucky luke de l'info : il voit les pannes du réseau avant vous, trop fort!), Patrice Schiller (non non l'épistémologie n'est pas une maladie incurable, quoique!), ...

Je ne saurais terminé cette séance de dédicace sans remercier mon épouse Nathalie ainsi que mes trois enfants qui ont accepté contraints et quasiment forcés une disponibilité de ma part plus proche de zéro que de toute autre valeur lors de la préparation de ce manuscrit. Pfff, ce qui m'ennuie, c'est que maintenant mes enfants sont devenus plus forts que moi sur la Wii et me mettent une pilée sur Mario Galaxy!!! Nintendo c'est plus fort que l'HdR!!!

Dankie, Vielen Dank, tak, gracias, tänan teid, grazie, thanks...

À Nathalie, Colleen, Ruben et Nolan

Table des matières

1	Un problème de qualité	1
1.1	Approche philosophique	1
1.2	La qualité est-elle fidélité ?	2
1.3	Mesures Objectives vs. Mesures Subjectives	3
1.3.1	Évaluation subjective de la qualité	3
1.3.1.1	Les tests comparatifs binaires	4
1.3.1.2	Les tests comparatifs n-naires	5
1.3.1.3	Les tests de mesure absolue	5
1.3.1.4	L'être humain est-il un observateur fiable ?	7
1.3.1.5	Les valeurs différentielles des scores d'opinion : DMOS	8
1.3.2	Évaluation objective de la qualité	8
1.3.3	Les métriques avec référence	9
1.3.3.1	Modèles spécifiques du Système Visuel Humain	10
1.3.3.2	Statistiques des scènes naturelles	12
1.3.4	Les métriques avec référence réduite	12
1.3.5	Les métriques sans référence	13
1.4	Les bases d'images	14
1.5	La performance vue comme une histoire de corrélation	15
1.5.1	Indicateur de précision : coefficient de corrélation de Pearson	16
1.5.2	Indicateur de monotonie : coefficient de corrélation de Spearman	16
1.5.3	Indicateur de monotonie : coefficient de corrélation de Kendall	16
1.5.4	Indicateur de cohérence : ratio des données aberrantes	17
1.6	Point de vue personnel	17

2	Vers une évaluation agnostique de la qualité sans référence	19
2.1	La vision et les statistiques de scènes naturelles	19
2.2	Le choix du domaine transformé	23
2.3	L'évaluation de la qualité sans référence basée sur les SSN	24
2.3.1	Le modèle développé	24
2.3.2	La sélection des caractéristiques basée sur des propriétés du SVH	25
2.3.3	La modélisation des propriétés retenues	25
2.3.3.1	La mesure du contraste	26
2.3.3.2	La mesure de distorsion aux structures	26
2.3.3.3	Les mesures d'anisotropie	27
2.3.4	L'approche multirésolution	28
2.3.5	Le modèle de prédiction probabiliste	30
2.3.6	Résultats	31
2.4	Application en biométrie	32
2.4.1	Méthode d'évaluation de la qualité des données biométriques	33
2.4.2	Mesure de performances	35
2.4.2.1	Apparatus	35
2.4.2.2	Mesure de performance de la méthode développée dans l'estimation de la qualité	36
2.4.2.3	Relation qualité-performance d'un système de vérification biométrique	37
2.5	G-BLIINDS2 : un modèle SSN généralisé pour une métrique agnostique	39
2.5.1	Le choix du modèle probabiliste généralisé	40
2.5.2	Les paramètres induits par le modèle probabiliste	41
2.5.2.1	Le paramètre de forme γ	41
2.5.2.2	Le coefficient de variation fréquentielle ζ	42
2.5.2.3	Le ratio d'énergie des sous-bandes ρ	42
2.5.2.4	La variance anisotropique ξ	44
2.5.3	Le modèle de prédiction	44

2.5.4	Évaluation de la performance	44
2.5.4.1	Apparatus	46
2.5.4.2	Corrélation des paramètres calculés avec le jugement humain	46
2.5.4.3	Performance de la métrique	46
2.6	Conclusion	50
3	Estimation des dégradations d'une image avec référence	51
3.1	Création et manipulation du vecteur de qualité : le choix des attributs	53
3.2	Le processus de classification du vecteur de qualité	55
3.2.1	La sélection du modèle	55
3.2.2	Les règles de combinaison	56
3.2.3	Résultats	57
3.3	De l'incertitude sur la classification	58
3.4	Combinaison des classificateurs sous raisonnement incertain	59
3.4.1	Éléments de théorie de l'information	59
3.4.2	Construction des fonctions de masse	60
3.4.3	Mesure des performances	61
3.5	Schéma de régression	61
3.6	Mesure des performances du schéma général	62
3.6.1	Apparatus	62
3.6.2	Résultats	63
3.6.3	Discussion	67
3.7	Application en segmentation d'images cellulaires	67
3.7.1	Création et fusion des cartes de segmentation	69
3.7.1.1	Segmentation par classification des pixels	69
3.7.1.2	Traitement des pixels incohérents	69
3.7.2	Résultats expérimentaux	70
3.7.2.1	Évaluation de la qualité de la segmentation	70
3.7.2.2	Résultats	71

3.8	Estimation de la combinaison de dégradation	73
3.8.1	L'approche retenue	73
3.8.2	Estimation de la distorsion vue comme un problème d'optimisation	75
3.8.3	De l'espace 2D à l'espace 3D	78
3.8.3.1	Optimisation basée sur l'utilisation de patch	79
3.8.3.2	Algorithme de Newton-Raphson–NR	80
3.8.3.3	Utilisation de patches raffinés	80
3.8.4	Comparaison des méthodes	81
3.9	Conclusion	85
4	Optimisation de la corrélation par échelles de différence	87
4.1	Échelle de différence par maximum de vraisemblance	89
4.2	Procédure d'optimisation d'algorithmes d'évaluation de la qualité.	91
4.2.1	L'algorithme d'évaluation de la qualité	91
4.2.2	Évolution de la corrélation en fonction du niveau de compression	91
4.2.2.1	Apparatus	91
4.2.2.2	Résultats	93
4.2.3	Optimisation par algorithme génétique	96
4.2.3.1	Stratégie de recherche	96
4.2.3.2	Résultats de l'optimisation	97
4.2.4	Validation du protocole	99
4.3	Conclusion	100
5	Conclusion générale et perspectives	103
5.1	Synthèse	103
5.2	Perspectives	105
	Bibliographie	109
	Annexes	116
A	Estimation des dégradations	117

B Curriculum vitæ	119
Etat Civil	119
Postes occupés	119
Formation	120
Responsabilités administratives	121
Activités de Recherche	121
Activités d'Enseignement	123
Activités de Recherche	127
Publications	132
 Liste de figures	 139
 Liste des tables	 145

Un problème de qualité

Sommaire

1.1	Approche philosophique	1
1.2	La qualité est-elle fidélité ?	2
1.3	Mesures Objectives vs. Mesures Subjectives	3
1.4	Les bases d'images	14
1.5	La performance vue comme une histoire de corrélation	15
1.6	Point de vue personnel	17

1.1 Approche philosophique

Au premier coup d'œil posé sur un objet (au sens sémantique du terme), l'être humain est capable de dire si sa vue lui est plaisante ou non. Il ne fait alors ni plus ni moins qu'une classification sémantique de la perception de cet objet en fonction du sentiment procuré et ressenti en, par exemple, deux catégories : « c'est beau » ou « c'est quelconque ». Il est très rare (voire exceptionnel) à ce stade de voir une personne attribuer, de son propre chef, une note quantifiant la qualité de ce qu'il regarde !¹

Même s'il est logique d'imaginer que plusieurs critères puissent influencer la prise de décision finale, une telle aptitude à classer les sensations visuelles est indiscutablement à mettre en relation avec la conscience inhérente à chaque être humain. La conscience est liée à ce que Freud appelle « le système perception-conscience » [1]. Il s'agit d'une fonction périphérique de l'appareil psychique qui reçoit les informations du monde extérieur et celles venant des souvenirs et des sensations internes de plaisir ou de déplaisir. Le caractère immédiat de cette fonction perceptive entraîne une impossibilité pour la conscience de garder une trace durable de ces informations. Elle les communique au pré-conscient, lieu d'une première mise en mémoire. La conscience perçoit et transmet des qualités sensibles. Freud emploie des formules comme « indice de perception, de qualité, de réalité » pour décrire la teneur des opérations du système

1. À moins d'être tombé dedans petit

perception-conscience. Ainsi, la perception d'un objet par un observateur humain est à considérer comme l'une des différentes échelles internes à un processus menant à une appréciation globale de la qualité d'un objet ou en l'occurrence d'une image. À ce stade, il convient de différencier deux expressions indifféremment utilisées pour faire référence à cette notion de qualité : la « qualité d'une image » et la « fidélité d'une image ». En effet, il est à noter que par abus de langage, il est souvent fait un amalgame entre ces deux formulations. La notion de qualité pourrait être à l'Artiste ce que la notion de fidélité serait au faussaire. L'Artiste travaille généralement à partir de concepts, d'impressions liés à son environnement social et/ou professionnel et se place dans un courant artistique existant (relation maître-élève) ou dans un nouveau courant qu'il crée de toutes pièces. Les œuvres réalisées sont ainsi considérées comme des originaux, et les experts parlent de la qualité de l'œuvre. Derrière cette approche, on se rend compte qu'au terme de qualité est associée la notion d'originalité. Qui ne s'est jamais retrouvé face à une œuvre qui l'a laissé perplexe tandis que son voisin s'en est émerveillé ? Il suffit de déambuler dans les musées pour constater ce phénomène. Ainsi, on qualifie la qualité d'une œuvre en fonction de sa conscience et de sa sensibilité personnelle prédéfinie de par son environnement économique et social. Le faussaire travaille généralement à partir d'un modèle et essaie de le reproduire avec la plus grande fidélité possible. Dans ce cas, le faussaire se doit de fournir une pièce irréprochable et il n'est pas rare qu'il utilise les mêmes techniques employées par l'auteur plusieurs siècles auparavant (combinaison de plusieurs pigments pour réaliser la couleur, utilisation d'une toile de la même époque, etc.). Dans ce cas, la copie doit être fidèle à l'original. D'aucuns pourraient prétendre que, dans certains cas, la copie peut dépasser la qualité de l'original. Pour anecdote, l'un des plus célèbres scandales dans l'histoire du marché de l'art a commencé par une banale petite annonce dans un journal anglais : « Fausses copies authentiques (XIXe et XXe siècles) pour 150 livres. » John Myatt (<http://www.johnmyatt.com/>), professeur d'arts plastiques aux abois, avait trouvé ainsi le moyen de s'enrichir. Un certain John Drewe, escroc génial et bon connaisseur des milieux artistiques, flaira l'affaire du siècle et s'associa sur le champ avec l'auteur de l'annonce. Le duo réussit à sévir durant presque dix ans. Drewe et Myatt trompèrent les plus grands experts, les maisons de vente Sotheby's et Christie's, et des musées de renom comme la Tate Gallery ou le Victoria and Albert Museum. Copies magistrales de Giacometti ou de Dubuffet, faux certificats d'authenticité, le système était parfait, jusqu'à ce que Scotland Yard se penche sur le problème !

1.2 La qualité est-elle fidélité ?

D'un point de vue plus pragmatique, la qualité d'une image est un des concepts sur lequel la recherche en traitement d'image prend une part de plus en plus prépondérante. Tout le problème consiste à caractériser la qualité d'une image, tel que peut le faire l'être humain. Dès lors, il nous faut dissocier les deux types de mesures :

1. Les mesures de fidélité

2. Les mesures de qualité

La mesure de fidélité, comme son nom l'indique, permet principalement de savoir si la reproduction de l'image est fidèle ou non à l'originale. Dans ce cas, on met en place une mesure qui calcule une distance entre les deux images. Cette distance symbolise numériquement l'écart qu'il peut y avoir entre les deux reproductions de l'image. L'objectif ultime étant de valider cette mesure d'écart par la sensation visuelle que ledit écart procure sur un observateur humain.

La mesure de qualité est à rapprocher de ce que fait naturellement et instinctivement l'être humain devant toute nouvelle œuvre : il lui donne une appréciation en fonction de sa conscience. Dès lors, l'être humain ne peut être dissocié de la mesure de la qualité. Ainsi, l'étude des mécanismes permettant d'appréhender les échelles internes utilisées lors de l'évaluation de la qualité par un être humain est devenue un domaine de recherche à part entière. Par exemple, dès 1860, Gustav Theodor Fechner proposa de mesurer des événements physiques déclenchés intentionnellement par l'expérimentateur, et des réponses manifestes des observateurs, réponses obtenues selon des modèles spécifiés.

1.3 Mesures Objectives vs. Mesures Subjectives

Que l'on cherche à développer des métriques de qualité (dans le cas où l'on ne dispose pas de la source originale) ou de fidélité (si l'on a accès à la source), la mesure de la performance du modèle algorithmique développé doit être conforme avec les résultats obtenus lors de l'évaluation de la qualité par un être humain dans le cadre de tests psychovisuels [2, 3]. Ainsi, il convient de mettre en place une « mesure de conformité ». Étant donné que les êtres humains sont très majoritairement les observateurs finaux des images et qu'ils prennent des décisions en fonction de ce qu'ils observent sur une image, ils restent encore actuellement les seuls juges fiables de la qualité des images. On prend conscience immédiatement des conséquences désastreuses que pourrait avoir une image de mauvaise qualité sur le diagnostic formulé par un médecin, ou sur l'identification d'une personne suspectée de voie de fait, etc. Dans ce cas, les tests psychovisuels demeurent un outil prépondérant pour l'évaluation des performances des algorithmes de traitement d'images générant des dégradations de la source. Cependant, la compilation des données obtenues via des observateurs humains est fastidieuse à mettre en place et gourmande en temps et moyens humains [4]. Elle permet néanmoins de calculer le score moyen des observateurs (MOS – *Mean Opinion Score*) pour chacune des images présentées lors du test. Les valeurs MOS deviendront ainsi les vérités terrains que toute mesure de qualité algorithmique devra être en mesure de reproduire le plus rigoureusement possible.

1.3.1 Évaluation subjective de la qualité

Les procédures ainsi que les standards d'évaluation subjective de la qualité des images ont été formalisés dans les normes ITU-R Rec. BT.500 [5] et ITU-T Rec. P.910 [4]. Ces recomman-

dations (initialement définies pour des outils d'affichage) fixent plusieurs contraintes pour mener à bien les tests psychovisuels comme la distance entre l'écran de visualisation et le fond de la salle, la réflexion des revêtements, la correction gamma des moniteurs, ainsi que les méthodes d'analyse des résultats. Les paramètres expérimentaux diffèrent pour les deux recommandations, mais les procédures mises en avant dans chacune d'elle doivent être appliquées dès lors que l'on évalue subjectivement la qualité des images. Parmi le panel de procédures décrites dans ces recommandations, on retrouve :

- *Les tests comparatifs binaires* où l'observateur humain doit effectuer une comparaison entre deux images et sélectionner l'image de meilleure qualité,
- *Les tests comparatifs n-aires* pendant lesquels on doit ordonner les images selon un ordre croissant (ou décroissant) de leur qualité ;
- *Les tests de mesure absolue* durant lesquels une image est présentée à un observateur humain qui doit noter sa qualité sans avoir la possibilité de faire une comparaison avec l'image originale.

En outre, pour chacun de ces tests, deux critères liés au temps de présentation des images à l'observateur existent :

1. *une présentation flash* pendant laquelle les images apparaissent brutalement et disparaissent tout aussi brutalement.
2. *une présentation progressive* où l'image apparaît petit à petit à l'écran, en terme de contraste.

Quelle que soit la procédure mise en œuvre, l'objectif au final est de pouvoir compiler l'ensemble des valeurs obtenues en une valeur MOS de l'image. Nous allons préciser par la suite chacune de ces procédures.

1.3.1.1 Les tests comparatifs binaires

Dépendamment de la tâche que l'on demande d'effectuer aux observateurs humains, ces tests sont aussi connus sous les acronymes :

1. DSCQS–*Double Stimuli Continuous Quality Scale*–[5] si l'on demande à l'observateur de quantifier la qualité de l'image le long d'une échelle continue,
2. DSIS–*Double Stimulus Impairment Scale*²–s'il est demandé à l'observateur humain de noter la qualité de l'image sur une échelle discrète contenant un nombre restreint de classes.

L'objectif de chacun de ces tests est de noter la qualité de l'image en ayant un accès simultané ou séquentiel³ à l'image de référence.

2. également connu sous l'appellation DCR *Degradation Category Rating*

3. *i.e.* images présentées l'une après l'autre

Qualité	Note	Dégradations
Excellente	5	Défauts imperceptibles
Bonne	4	Défauts tout juste perceptibles
Assez bonne	3	Défauts perceptibles mais pas gênants
Médiocre	2	Défauts perceptibles et gênants
Mauvaise	1	Défauts perceptibles et très gênants

TABLE 1.1: Notes utilisées permettant de quantifier la qualité (à gauche) ou la dégradation (à droite) d'une image selon l'échelle à cinq valeurs définie par l'ITU dans sa recommandation ITU-R BT500.



Même si l'ITU préconise l'utilisation d'une échelle discrète de note à cinq valeurs (tel que présenté dans le tableau 1.1), il est parfois préférable d'utiliser une seconde échelle dont la dynamique est plus grande (jusqu'à 100 notes), ce qui, mathématiquement, représente une transformation bijective de la première échelle vers la seconde. Certains auteurs font référence à cette échelle sous le terme d'*échelle continue*, ce qui, mathématiquement, n'est pas exact.

1.3.1.2 Les tests comparatifs n-naires

Ces tests ont pour principal objectif d'ordonner plusieurs images les unes par rapport aux autres selon un ordre croissant (ou décroissant) des notes de qualité. Dans ce cas, l'observateur est amené à comparer plusieurs versions dégradées d'une même image originale et est autorisé à modifier une note déjà attribuée à une image. Durant l'exécution de ces tests, l'observateur réalise pratiquement des comparaisons d'intervalles perceptuels de couple d'images en fixant lui-même les bornes inférieure et supérieure de chaque intervalle par l'attribution de notes de qualité. La figure 1.1 présente un exemple d'un test comparatif n-aire. Dans ce test, neuf images dégradées suite à l'application de la compression JPEG2000 selon neuf taux de compression différents sont affichées. L'observateur doit alors effectuer le classement de ces images selon un ordre décroissant de qualité.

1.3.1.3 Les tests de mesure absolue

L'utilisation de ces tests permet à un observateur humain de coter numériquement la qualité d'une image présentée seule selon une échelle de notation discrète de plus grande dynamique contenant 100 valeurs. L'avantage d'utiliser une plus grande dynamique de notes est d'assurer une gamme de choix plus importante et de faire en sorte que des images de qualité différente (mais proche) auront des notes proches mais distinctes (ce qui pourrait ne pas être tout le temps possible avec une échelle de dynamique plus faible). La figure 1.2 présente un exemple

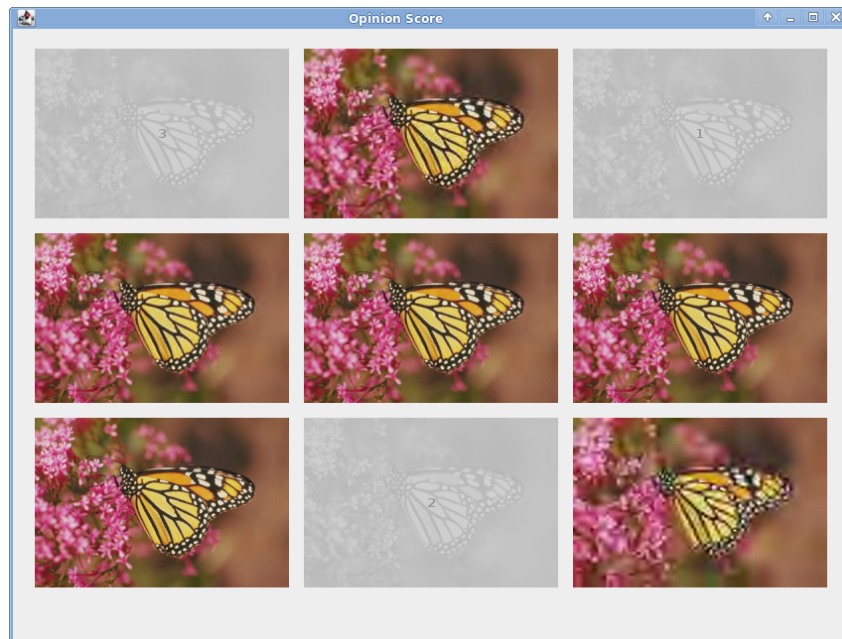


FIGURE 1.1: Exemple de test comparatif n -naire. L'observateur doit classer la qualité des images selon un ordre croissant de qualité.

de test de mesure absolue de la qualité. Lorsqu'une image est affichée, l'observateur positionne le curseur sur la valeur souhaitée et valide son choix en cliquant sur le bouton.

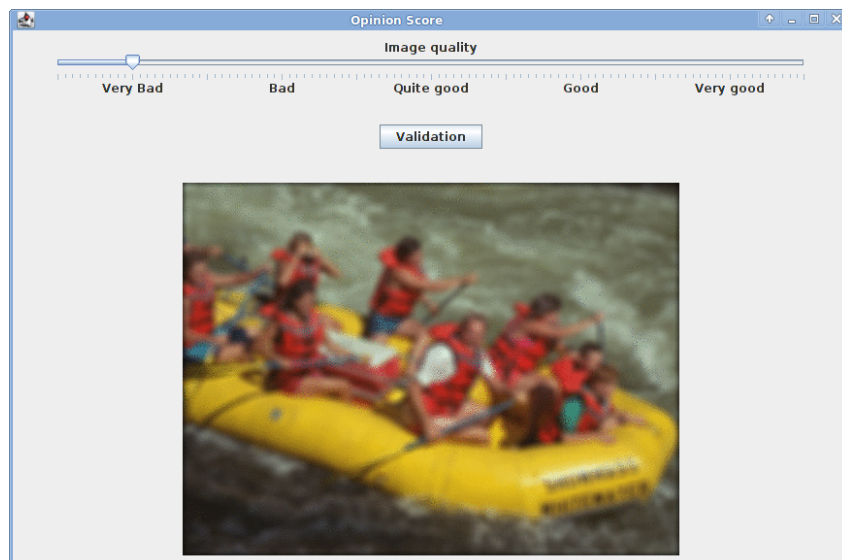


FIGURE 1.2: Exemple de test de mesure absolue de la qualité. L'observateur doit noter la qualité de l'image présentée en faisant varier le curseur et valide son choix en cliquant sur le bouton.

1.3.1.4 L'être humain est-il un observateur fiable ?

Comme mentionné auparavant, tout être humain prend une décision selon un système perception-conscience qu'il s'est forgé tout au long de sa vie. De nombreux critères interfèrent alors dans sa prise de décision finale. Ceci sous-entend qu'il convient de développer des protocoles rigoureux pour assurer une fiabilité et la reproductibilité des résultats obtenus.

Dans une première étape, il convient de calculer la note moyenne (MOS) obtenue pour chacune des images et pour l'ensemble des observateurs :

$$MOS_{jk} = \frac{1}{N} \sum_{i=1}^N u_{ijk}, \quad (1.1)$$

où N est le nombre d'observateurs et u_{ijk} la note attribuée par l'observateur i à l'image originale j ayant subie une dégradation k .

Afin d'évaluer au mieux la fiabilité des résultats ainsi obtenus, il est nécessaire d'associer à chaque valeur MOS un intervalle de confiance à 95% $[MOS_{jk} - \delta_{jk}, MOS_{jk} + \delta_{jk}]$ sous l'hypothèse que la variable aléatoire (v.a.) MOS_{jk} suive une loi gaussienne.

En dernier ressort, il convient de rechercher une inconsistance dans les résultats fournis par un même observateur. L'idée finale étant de ne pas tenir compte de ses réponses lors du calcul de la valeur MOS des images. Dans ce cas, le kurtosis de la distribution de la v.a. MOS_{jk} est calculé afin de vérifier le caractère mésokurtique de la distribution (ce qui signifie que la v.a. suit une loi normale). Finalement, deux quantités sont calculées pour chaque observateur j :

1. une mesure d'écartement par rapport à la moyenne définie par

$$R_1^j = \frac{P_j + Q_j}{\text{card}(Img)} \quad (1.2)$$

2. une mesure définie par

$$R_2^j = \frac{P_j - Q_j}{P_j + Q_j} \quad (1.3)$$

où $\text{card}(Img)$ représente le nombre total d'images utilisés dans le test. Les variables P_j et Q_j sont définies comme suit :

$$\begin{aligned} P_j &= P_j + 1 & \text{si } u_{ijk} &\geq MOS_{jk} + K_p \sigma_{jk}, \\ Q_j &= Q_j + 1 & \text{si } u_{ijk} &\geq MOS_{jk} - K_p \sigma_{jk}. \end{aligned}$$

La valeur de K_p dépend du caractère normal de la distribution de la v.a. MOS_{jk} : $K_p = 2$ pour une distribution normale, $K_p = \sqrt{20}$ dans le cas contraire.

Le critère final de rejet de l'observateur j est défini par la règle suivante :

$$R_1^j > 0.05 \quad \text{et} \quad R_2^j < 0.3 \quad (1.4)$$

Ce critère permet d'évincer les observateurs ayant fourni des réponses incohérentes. Cependant, les observateurs n'ont pas nécessairement utilisé la gamme complète de l'échelle de notation et encore moins le même sous-intervalle. Plusieurs solutions existent pour contrer cet état de fait. La solution la plus répandue consiste à appliquer la transformée en Z-scores. Cette technique permet ainsi de minimiser les variations entre les notes individuelles.

Comme on vient de le remarquer, l'observateur humain n'est peut-être pas un observateur idéal puisqu'il est nécessaire d'effectuer une harmonisation des résultats et de rechercher des observateurs incohérents. Néanmoins, il reste le dernier juge de la qualité des images, et ces résultats constituent une vérité terrain qui sera utilisée pour quantifier les performances des solutions algorithmiques de mesure de qualité.

1.3.1.5 Les valeurs différentielles des scores d'opinion : DMOS

Afin de calculer les valeurs DMOS (*Differential Mean Opinion Score*), les scores bruts sont convertis en scores de différence de qualité :

$$d_{ijk} = u_{ij0} - u_{ijk} \quad (1.5)$$

où u_{ij0} représente la note attribuée par l'observateur i à l'image originale j n'ayant pas subi de dégradation ($k = 0$). Les valeurs d_{ijk} sont ensuite converties en Z-scores après avoir appliqué le protocole de rejet des observateurs et des valeurs aberrantes :

$$z_{ijk} = \frac{d_{ijk} - \bar{d}_i}{\sigma_i} \quad (1.6)$$

où \bar{d}_i correspond à la moyenne des différences des scores pour l'observateur i calculée sur toutes les images, et σ_i est l'écart-type associé. Les Z-scores sont ensuite moyennés pour tous les observateurs de manière à calculer la valeur \hat{z}_{jk} pour l'image originale j ayant subi la dégradation k .

Un processus de réaligement des Z-scores et des valeurs d_{ijk} est effectué. Finalement, les valeurs DMOS sont obtenues par régression linéaire :

$$\text{DMOS}(\bar{z}) = p_1 \bar{z} + p_2 \quad (1.7)$$

où les constantes p_1 et p_2 sont déterminées par minimisation de l'erreur entre les scores $\text{DMOS}(\bar{z}_{jk})$ et DMOS_{jk} , avec DMOS_{jk} les valeurs obtenues après application du procédé de réaligement.

1.3.2 Évaluation objective de la qualité

La possibilité de développer des solutions algorithmiques conformes aux données subjectives (obtenues par observation humaine) devient un échappatoire providentiel aux mesures

subjectives de qualité, puisqu'elles permettent d'économiser ressources et temps. Les objectifs à atteindre lors du développement de telles mesures de qualité peuvent être résumés comme suit :

1. *Interprétabilité* de la valeur du MOS pour une application donnée : les personnes doivent être en mesure d'interpréter toute valeur du MOS en terme de qualité,
2. *Fiabilité* des valeurs MOS prédites : les valeurs prédites doivent être très proches des valeurs obtenues lors des tests psychovisuels,
3. *Reproductibilité* des résultats : toute personne qui utilise la métrique sur un même jeu de données doit obtenir des résultats identiques.

Force est de constater que parmi les multiples métriques algorithmiques développées jusqu'à présent⁴, aucune ne répond simultanément aux trois critères énoncés ci-dessus, tandis qu'elles répondent quasiment toutes aux deux derniers critères [7].

L'ensemble des métriques algorithmiques existantes peut être clairement regroupé en trois grandes familles :

1. les métriques avec référence (*FR–Full Reference*),
2. Les métriques avec référence réduite (*RR–Reduced Reference*),
3. les métriques sans référence (*NR–No Reference*).

1.3.3 Les métriques avec référence

Au lieu d'utiliser le terme de métrique de qualité, il conviendrait d'utiliser le terme de mesure de similarité, puisque l'objectif est finalement de noter le degré de conformité de l'image dégradée à l'image originale.

Chaque méthode développée aspire à délivrer une note de qualité à l'instar de ce que fait un observateur humain. Dans ce cas, il est nécessaire de pallier les inconvénients des mesures simples telles que l'Erreur Quadratique Moyenne (EQM) ou tout autre norme L_p [8]. En effet, l'utilisation de telles mesures impliquent les hypothèses suivantes :

- La qualité d'une image ne dépend pas des relations spatiales entre les pixels (ou entre les zones de pixels) de l'image. Ainsi, le fait de changer l'organisation spatiale des pixels d'une image n'influence pas le résultat final,
- La qualité d'une image ne dépend pas des relations entre le signal d'origine et le signal dégradé. Ainsi, pour un même signal d'erreur, la mesure de distorsion doit fournir le même résultat, et ce indépendamment du signal source,
- La qualité d'une image dépend exclusivement de l'amplitude du signal d'erreur. Ainsi, le simple fait de modifier le signe du signal d'erreur ne doit pas modifier la valeur finale,
- Tous les pixels de l'image ont la même importance visuelle.

4. En 2009, Pederson avait déjà recensé plus de 111 métriques existantes [6]

Cependant, aucune de ces hypothèses n'est valable lorsque l'on cherche à utiliser ces métriques de qualité dédiées aux images ou vidéo dans le seul domaine spatial. On obtient alors des valeurs qui ne reflètent pas ce que l'on perçoit. Néanmoins, lorsque ces métriques sont appliquées dans des domaines autres, comme par exemple celui des ondelettes multi-échelles et multirésolution, les effets sont atténués [9].

Depuis lors, les mesures de qualité avec référence n'ont eu de cesse de proliférer afin de pallier les défauts inhérents à l'utilisation des normes L_p . La multitude d'algorithmes sous-jacents a été classée de multiples façons, certains préférant utiliser une approche descendante vs. une approche ascendante, d'autres une approche basée perception vs. une approche orientée application, etc.

La classification qui est proposée ici se veut un consensus entre les différentes classifications existantes. Les deux catégories retenues, non nécessairement disjointes, sont :

1. approches utilisant des modèles spécifiques du Système Visuel Humain,
2. approches basées sur une modélisation des scènes naturelles.

1.3.3.1 Modèles spécifiques du Système Visuel Humain

Dans la première catégorie, l'objectif est d'évaluer la qualité des images dégradées d'une manière qui émule le traitement bas-niveau du cerveau humain, afin de pouvoir extraire des caractéristiques de distorsions qui correspondent au traitement effectué au niveau de la perception de la qualité. Parmi les caractéristiques de vision bas-niveau extraites, on observe

les fonctions de sensibilité au contraste (*CSF–Contrast Sensitivity Function*) qui modélisent la capacité du système visuel à détecter des différences de luminance, donc de déterminer l'existence de « bords » entre surfaces homogènes.

le phénomène d'adaptation à la lumière, également connu comme étant la loi de Weber-Fechner. Cette loi modélise la propriété du SVH à encoder le contraste d'un stimulus visuel plutôt que l'intensité absolue de luminance.

l'effet de masquage qui est un concept général faisant référence à une baisse de visibilité d'une composante de l'image (le signal) due à la présence d'une autre composante (le masqueur). La force du masquage est mesurée par la variation de la visibilité du signal lors de la présence et de l'absence du masqueur. En général, la force du masquage est la plus élevée lorsque le signal et le masqueur sont similaires de par leur localisation spatiale, de leur contenu fréquentiel et de leur orientation.

la fovéation induite par la distribution non uniforme des cônes et des cellules ganglionnaires dans la rétine [10]. Les effets d'une telle distribution s'observent lorsqu'un être humain fixe un point quelconque d'une image (ou d'une scène réelle) : une image de résolution spatiale variable est transmise au cerveau. La région autour du point de fixation (région de la fovéa contenant la densité maximale de cônes) est échantillonnée avec la résolution

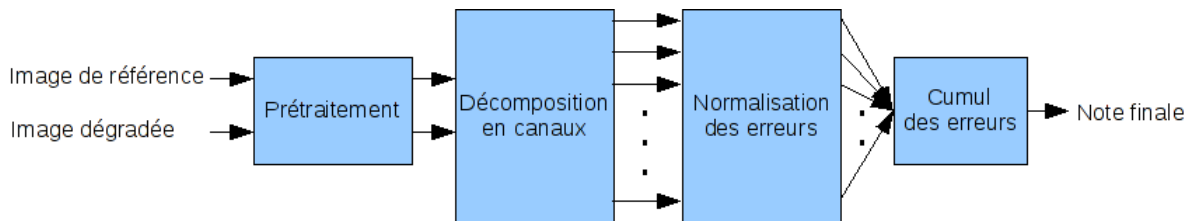


FIGURE 1.3: Schéma conducteur utilisé lors du développement d'une métrique de qualité avec référence.

spatiale maximale, tandis que cette résolution spatiale décroît lorsque l'on s'en éloigne. Ainsi, il n'y a quasiment aucune différence entre le centre de l'image simulant une vision fovéale et l'image originale, alors que cette différence s'accroît avec l'éloignement au centre de l'image de vision fovéale. Cette caractéristique est modélisée par une opération de convolution entre l'image originale et une fonction d'étalement du point (*PSF—Point Spread Function*). Cette fonction peut être définie selon des modèles théoriques [11], ou mesurée directement [12].

La grande majorité des méthodes développées dans le cadre de cette approche se basent sur le paradigme des cartes de distorsion (auss appelé paradigme des cartes d'erreur). Une quantification des erreurs (en l'occurrence des différences) entre l'image dégradée et l'image originale est réalisée en intégrant une ou plusieurs des caractéristiques présentées ci-dessus. La figure 1.3 illustre le schéma conducteur usuellement suivi lors de la construction d'une métrique de qualité avec référence.

Les opérations couramment utilisées lors de l'étape du pré-traitement regroupent un recalage spatial des images, une transformation colorimétrique (très souvent dans un espace opposition de couleur de type YC_bC_r), une non-linéarité par morceaux, un filtrage de type PSF, un filtrage de type CSF (très souvent modélisé par un filtre passe-bas). Chaque implantation peut inclure une ou plusieurs des opérations précédentes.

La deuxième étape concerne la représentation du signal image en plusieurs canaux, afin de respecter le fait que les neurones présents dans l'aire primaire visuelle sont sensibles aux orientations, fréquences et localisations spatiales. On retrouve ainsi la décomposition de Fourier, de Gabor, la Transformée en Cosinus Discrète (TCD) locale par bloc, la transformée en ondelettes séparables, ainsi que la transformée en ondelettes polaires séparables (incluant la transformée cortex et la décomposition en pyramide orientable). Il est communément admis que les décompositions basées sur les ondelettes orientables (de type Gabor ou décomposition en pyramide orientable) simulent mieux les champs réceptifs des neurones de l'aire primaire du cortex, même si aucune réponse claire existe quant à la meilleure transformation à utiliser.

La troisième étape regroupe la construction de cartes d'erreur. Ces cartes sont générées en calculant la différence pondérée entre le signal de l'image originale et de l'image dégradée dans l'espace transformé retenu. Ces différences sont calculées en intégrant deux caractéristiques du

SVH, à savoir la fonction de sensibilité au contraste et les effets de masquage. À noter que si la fonction de sensibilité au contraste a déjà été utilisée dans l'étape de pré-traitement, il n'est pas utile de la prendre en compte à ce niveau. L'effet de masquage est, quant à lui, appliqué selon un mécanisme de gain de contrôle, dans lequel le facteur de normalisation pour un coefficient est déterminé par l'énergie des coefficients voisins dans le signal image origine.

En dernier lieu, un cumul des erreurs est obtenu selon le sommation de Minkowski

$$E = \left(\sum_m \sum_n |e(m,n)|^\beta \right)^{1/\beta} \quad (1.8)$$

où $e(m,n)$ représente l'erreur normalisée du coefficient n du canal m , et $1 \leq \beta \leq 4$. Il est également possible que le processus de cumul soit appliqué d'abord dans le domaine spatial puis fréquentiel.

1.3.3.2 Statistiques des scènes naturelles

Dans la seconde catégorie, l'hypothèse formulée repose sur le principe que les systèmes perceptifs biologiques ont évolué en fonction des propriétés physiques de l'environnement naturel, le tout modélisable par des propriétés statistiques. Cette approche est également connue sous l'étiquette « statistiques de scènes naturelles » (*NSS-Natural Scene Statistics*), qui est un spectre de recherche à part entière du domaine de la Perception. Ainsi, les images originales (modèles NSS générés à partir de caractéristiques extraites des images originales) sont considérées comme évoluant dans un sous-espace d'un vaste espace vectoriel regroupant toutes les images possibles (versions dégradées des images originales). Toute déviation d'un modèle NSS sera alors interprétée comme une dégradation de l'image originale associée et servira d'évaluateur de la qualité de l'image dégradée. De cette approche, sont nés les indices SSIM [13], MS-SSIM [14], VIF [15], VSNR [16] pour ne citer que ceux-ci.

1.3.4 Les métriques avec référence réduite

Les métriques avec référence réduite font partie des algorithmes d'évaluation de la qualité qui requièrent une information partielle du signal image d'origine de manière à prédire la qualité du signal image dégradé. Les solutions fournies par cette catégorie d'algorithmes se situent à mi-chemin des solutions apportées par les métriques algorithmiques avec référence et sans référence. Elles sont très utilisées dans le domaine du multimédia et plus précisément de la transmission d'images où la présence de l'image originale sur le poste de client l'utilisateur est impossible.

Ce type de système procède par extraction de caractéristiques du côté client et du côté serveur et analyse la qualité plutôt du côté serveur. Les caractéristiques extraites du côté client

sont transmises au serveur par un canal auxiliaire. Ce canal est réputé être idéal (*i.e.*, sans erreur), ce qui n'est pas nécessairement le cas en réalité. Cependant, même si des erreurs de transmission sont probables, des données décodées de manière parcellaire sont néanmoins suffisantes pour évaluer la qualité d'une image [17].

L'objectif d'un tel algorithme de mesure de la qualité est de trouver le meilleur compromis entre le nombre minimum de caractéristiques extraites et la précision de la qualité prédite. Ceci ne signifie nullement que le fait de posséder le plus possible de caractéristiques garantisse une qualité optimale du processus de prédiction de la qualité du signal image.



À ce stade, une très forte synergie existe avec les techniques d'apprentissage statistique où l'objectif est de construire des fonctions de décision simulant un oracle à partir de caractéristiques extraites du signal image lors de la classification.

Les algorithmes d'évaluation de la qualité les plus couramment déployés sont les algorithmes œuvrant dans le domaine des ondelettes. Dans [17], Wang *et al.* modélise les coefficients d'ondelettes par une densité gaussienne généralisée par image. Zepernick *et al.* [18] ont développé un algorithme basé sur l'utilisation d'ondelettes hybrides et des bancs de filtres directionnels. L'idée est de ne conserver qu'un ensemble de coefficients provenant des sous-bandes d'ondelettes pour calculer le degré de similitude entre les coefficients calculés sur l'image originale et l'image dégradée.

1.3.5 Les métriques sans référence

Actuellement, il n'existe pas d'algorithme d'évaluation de la qualité sans référence ayant les mêmes niveaux de performance que les algorithmes de mesure de qualité avec référence. On peut observer en effet que l'indice SSIM (qui est une mesure de qualité avec référence) a été inclus dans la partie 10 AVC du standard H.264/MPEG4 [19] alors que l'utilisation d'une métrique sans référence n'est pour l'heure aucunement préconisée. Le problème majeur de tels algorithmes de mesure de qualité réside dans l'absence du signal image de référence.

Les approches existantes dans le domaine de la mesure de la qualité sans référence sont variées et suivent généralement une des trois tendances décrites ci-après :

1. **une approche spécifique aux distorsions.** Dans ce cas, l'algorithme quantifie une ou plusieurs distorsions spécifiques sans prendre en compte les éventuelles interactions existantes, et délivre une note finale en fonction de ces quantifications. On peut citer, par exemple, les effets de blocs [20, 21, 22, 23, 24], de flou [25, 26, 27, 28, 29], de ringing (appelé également phénomène de Gibbs) [30, 31, 32, 33] ou encore d'épanchement des couleurs. L'épanchement des couleurs est généré par le même processus que celui de l'effet de flou mentionné auparavant mais pour les composantes chromatiques :

Nom base	Date	Nb Distorsions	Nb Img ref	Nb Img Deg.
IRCCyN/IVC	2005	4	10	185
LIVE	2006	5	29	779
Tampere Image Database (TID)	2008	25	17	1700
MICT	2008	2	14	168
CSIQ	2009	6	30	866
WIQ	2009	1	7	80

TABLE 1.2: Caractéristiques des principales bases d'images disponibles.

suppression des hautes fréquences des composantes couleurs. Il se caractérise par un mélange de couleurs entre les zones à forte différence chromatique donnant l'impression à un observateur de percevoir une perte de couleurs.

2. **une approche d'apprentissage statistique.** Selon cette approche, un vecteur de caractéristiques est extrait des images. Un algorithme d'apprentissage (SVM, réseaux de neurones, ...) est ensuite entraîné de manière à prédire le score final de qualité [34, 35, 36, 37, 38, 39].
3. **une approche basée sur les statistiques des scènes naturelles–NSS (*Natural Scene Statistics*).** Cette approche est basée sur l'hypothèse que les images naturelles non dégradées occupent un sous-espace de l'espace de toutes les images possibles. Dans ce cas, on cherche à déterminer la distance entre l'image dégradée (qui est réputée ne pas appartenir au sous-ensemble des images non dégradées) et le sous-espace des images naturelles. Cette approche repose sur les changements observés dans les statistiques de l'image lors de l'application de dégradations [40, 41, 39]

1.4 Les bases d'images

Une fois qu'une solution algorithmique d'évaluation de la qualité est établie, il convient d'en mesurer les erreurs de prédiction, ce qui sous-entend que l'on dispose d'une vérité terrain. Cette vérité terrain est généralement accessible sous la forme des valeurs MOS ou DMOS. Actuellement, il existe quelques bases d'images annotées librement accessibles dont les caractéristiques sont résumées dans le tableau 1.2.

Chacune de ces bases propose son propre jeu de dégradation ainsi que les scores des observateurs humains :

1. les distorsions utilisées dans la base IRCCyN/IVC sont celles générées par la compression JPEG, JPEG2000, LAR [42] ainsi que par l'application d'un flou. 15 observateurs ont participé au test de mesure de qualité.

URL : <http://www2.irccyn.ec-nantes.fr/ivcdb/>

2. les distorsions utilisées dans la base LIVE sont celles générées par la compression JPEG, JPEG2000, l'application d'un flou gaussien, d'un bruit blanc et de la simulation d'erreur dans le canal à évanouissement de Rayleigh. Deux groupes de 27 et 29 observateurs ont participé au test de mesure de la qualité.

URL : <http://live.ece.utexas.edu/research/quality/subjective.htm>

3. les distorsions utilisées dans la base TID sont de loin les plus nombreuses. Elles regroupent les distorsions de compression et de transmission avec erreurs JPEG et JPEG2000, locales de bruit et de flou ainsi que des variations de luminance et de contraste. Au total, 838 observateurs ont participé aux tests.

URL : <http://www.ponomarenko.info/tid2008.htm>

4. La base de données MICT se focalise essentiellement sur les distorsions issues des schémas de compression JPEG et JPEG2000. 16 observateurs ont participé aux tests.

URL : <http://mict.eng.u-toyama.ac.jp/mictdb.html>

5. Les distorsions utilisées dans la base d'images CSIQ se concentrent sur celles obtenues par compression JPEG, JPEG2000, ainsi que celles issues de variation de contraste globale, de bruit additif gaussien et de flou gaussien. 35 observateurs ont participé aux tests psychovisuels.

URL : <http://vision.okstate.edu/index.php?loc=csiq>

6. La base d'images WIQ contient des images dégradées suite à l'application du schéma de compression JPEG en niveau de gris et d'une simulation de transmission sur un lien sans-fil. 30 observateurs ont annoté la qualité des images.

URL : <http://www.bth.se/tek/rcg.nsf/pages/wiq-db>

1.5 La performance vue comme une histoire de corrélation

Ainsi, les scores (MOSp) obtenus par les solutions algorithmiques sont comparées aux scores subjectifs (MOS) selon une mesure de conformité afin de quantifier la performance de chacune des métriques algorithmiques testées. Cette mesure de conformité doit s'effectuer selon le plan de test de VQEG (*Video Quality Expert Group*) [43] qui préconise l'utilisation de plusieurs indicateurs. Ces indicateurs permettent de délivrer des informations quantitatives de précision, de monotonie et de cohérence de la prédiction. En outre, étant donné que les algorithmes d'évaluation de la qualité fournissent des scores qui ne sont pas nécessairement inscrits dans la même dynamique que celle des valeurs MOS, il est nécessaire d'utiliser une transformation non-linéaire afin de corriger cet inconvénient :

$$f(x) = \frac{\tau_1 - \tau_2}{1 + \exp\left(\frac{x - \tau_3}{\tau_4}\right)} + \tau_2 \quad (1.9)$$

où τ_1 , τ_2 , τ_3 et τ_4 sont calculés de manière à minimiser l'erreur quadratique moyenne (EQM) entre les valeurs MOS et les valeurs MOSp prédites par les algorithmes.

1.5.1 Indicateur de précision : coefficient de corrélation de Pearson

Le coefficient de corrélation de Pearson (encore appelé coefficient de corrélation linéaire) LCC quantifie la dépendance existante entre les valeurs MOS et MOSp. il est défini par :

$$LCC = \frac{\sum(MOS_i - \overline{MOS})(MOSp_i - \overline{MOSp})}{\sqrt{\sum(MOS_i - \overline{MOS})^2} \sqrt{\sum(MOSp_i - \overline{MOSp})^2}} \quad (1.10)$$

où \overline{MOS} et \overline{MOSp} sont les moyennes respectives des valeurs MOS et MOSp, et i représente l'image dégradée i . Plus le coefficient est proche des valeurs extrêmes -1 et 1, plus les deux variables sont fortement corrélées. À l'inverse, un coefficient égal à 0 signifie que les variables sont linéairement indépendantes.

1.5.2 Indicateur de monotonie : coefficient de corrélation de Spearman

La *monotonie* mesure quant à elle si l'accroissement d'une variable est associé à l'accroissement d'une autre variable. Idéalement, les variations de la valeur de la métrique pour différentes images devraient toujours avoir le même signe que les variations du jugement subjectif correspondant. Le degré de monotonie peut être évalué par la corrélation de rang de Spearman SROCC (*Spearman Rank Order Correlation Coefficient*). Avant de calculer ce coefficient, il convient de classer par ordre croissant (ou décroissant) les valeurs MOS et MOSp. La formulation de ce coefficient de corrélation est donnée par :

$$\begin{aligned} SROCC &= \frac{\sum(\chi_i - \bar{\chi})(\psi_i - \bar{\psi})}{\sqrt{\sum(\chi_i - \bar{\chi})^2} \sqrt{\sum(\psi_i - \bar{\psi})^2}} \\ &= 1 - \frac{6(\psi_i - \psi)^2}{N(N^2 - 1)} \end{aligned} \quad (1.11)$$

où χ_i est le rang de MOS_i , ψ_i le rang de $MOSp_i$ et $\bar{\chi}$ et $\bar{\psi}$ sont respectivement les rangs médians. La corrélation de rang de Spearman est non-paramétrique, *i.e.*, aucune hypothèse sur la forme du rapport entre les MOS_i et les $MOSp_i$ n'est formulée.

1.5.3 Indicateur de monotonie : coefficient de corrélation de Kendall

Le coefficient de corrélation de rang de Kendall KROCC (*Kendall Rank Order Correlation Coefficient*) mesure l'association existante entre deux mesures. Plus spécifiquement, il mesure la similarité de classement des valeurs. Ainsi, les paires $(MOS_i, MOSp_i)$ et $(MOS_j, MOSp_j)$ sont réputées concordantes si les rangs pour chacun des deux éléments sont identiques, *i.e.* si $(MOS_i > MOS_j) \wedge (MOSp_i > MOSp_j)$ ou $(MOS_i < MOS_j) \wedge (MOSp_i < MOSp_j)$. Dans le cas contraire, les paires sont réputées discordantes.

Le coefficient de Kendall est défini par :

$$KROCC = \frac{(\# \text{ de paires concordantes}) - (\# \text{ de paires discordantes})}{\frac{1}{2}n(n-1)} \quad (1.12)$$

où n représente le nombre de paires $(MOS_i, MOSp_i)$.

1.5.4 Indicateur de cohérence : ratio des données aberrantes

À l'instar de la procédure suivie pour rejeter un observateur incohérent (*i.e.*, pour lequel on observe une inconsistance dans les réponses), il est possible de définir un indicateur de cohérence en calculant le ratio des images « mal notées » sur le nombre total d'images. Une image i sera considérée comme mal notée si le score prédit $MOSp_i$ se situe à l'extérieur de l'intervalle de confiance à 95% de la valeur cible de l'image MOS_i .

1.6 Point de vue personnel

Une présentation des principales techniques d'évaluation de la qualité des images a été effectuée dans ce chapitre. Dès lors, pour chaque algorithme de qualité développé, il est nécessaire de mesurer la corrélation existante avec les notes acquises lors de séances de tests psychovisuels impliquant des observateurs humains. Une procédure de mesure de fiabilité des observateurs doit être réalisée afin de ne pas biaiser les résultats en conservant des données aberrantes.

Le fait d'avoir recours à des observateurs humains est un processus long et fastidieux, car il se doit d'être conforme au protocole défini par l'ITU dans sa recommandation BT-500. Ceci se traduit par l'existence de peu de bases d'images dégradées pour lesquelles les scores moyens des observateurs humains sont disponibles. Finalement, ces scores moyens font office de vérité terrain et toute technique algorithmique d'évaluation de la qualité se doit d'être fiable et consistante au regard de ces scores. Parmi les techniques d'évaluation de la qualité, on trouve trois grandes familles : les algorithmes requérant un signal image de référence (qui peuvent être considérées comme des mesures de similarité à l'instar de la norme L_2), les algorithmes basés sur une comparaison de deux vecteurs de caractéristiques respectivement associés au signal image de référence et dégradé (on parle de la présence de référence réduite) et finalement les techniques d'évaluation de la qualité pour lesquelles l'accès au signal source image est inexistant. Outre leur catégorie, l'objectif de chacun des algorithmes développés est de prédire le score de qualité, tel que le ferait un oracle (*i.e.*, l'observateur humain). La qualité de la prédiction est mesurée à l'aide d'indicateurs statistiques de cohérence et de régularité.

Le besoin de développer des algorithmes d'évaluation de la qualité, tandis que le signal image source n'est pas accessible, est de plus en plus prépondérant dans notre société où

l'ubiquité du message visuel est de mise. Dès lors, plusieurs réflexions ouvertes peuvent être formulées :

- Si l'on admet que les fonctionnalités du SVH ont évolué depuis le début de l'existence de l'être humain de manière à s'adapter aux propriétés statistiques du monde environnant, il est alors possible d'utiliser ces statistiques pour juger la qualité d'une image. Se pose alors le problème de la modélisation des dégradations de l'image dont l'origine est inconnue.
- Il n'est pas naturel pour un observateur humain de noter directement la qualité des images : il procède par analyse sémantique de la qualité, ce qui induit de sa part l'utilisation d'un processus de classification. Le fait de vouloir attribuer un scalaire représentant la qualité est purement pragmatique et est à mettre en lien avec la relation d'ordre total qui existe dans \mathbb{R} (qui est un corps totalement ordonné archimédien), alors qu'il n'existe plus pour $\mathbb{R}^n, n \geq 2$. Dès lors, le classement par ordre croissant (ou décroissant) de la qualité des images n'en est que plus facile.
Si l'on souhaite procéder à l'instar ce qui est effectué par le SVH, peut-être serait-il souhaitable d'intégrer un processus de classification de la qualité en fonction de classes sémantiques prédéterminées dans le calcul de la note finale de qualité.
- Afin d'établir la conformité d'une technique algorithmique d'évaluation de la qualité en fonction d'une vérité terrain, il est d'usage de faire appel à des mesures de corrélation. Or, les informations délivrées par ces mesures sont globales, *i.e.*, aucune informations quant aux scores prédits qui se situent aux extrémités de l'intervalle de confiance à 95% lors de la recherche de données aberrantes n'est accessible. Cependant, ils ont une influence indéniable sur la valeur finale de corrélation. Par conséquent, avoir une méthode qui garantisse un écart-type sur la valeur de corrélation le plus faible possible présenterait un avantage certain : il y aurait une garantie de très forte corrélation sur toute la dynamique des dégradations, ce qui n'est pas nécessairement le cas actuellement.

Ces points sont abordés dans les chapitres suivants et constituent mes activités de recherche menées depuis une dizaine d'années.

Vers une évaluation agnostique de la qualité sans référence

La problématique de l'évaluation de la qualité sans image de référence est abordée dans ce chapitre. Dans une première approche, la distribution des statistiques de scènes naturelles, calculées dans le domaine DCT, est directement modélisée afin de prédire la note de qualité de l'image. Ceci sera utilisé en biométrie afin de quantifier la qualité de données biométriques morphologiques. Une seconde approche se base sur l'utilisation des paramètres d'un modèle généralisé des statistiques de scènes naturelles afin de définir de nouvelles caractéristiques, dont la distribution sera modélisée par un modèle probabiliste généralisé quantifiant la qualité de l'image.

Sommaire

2.1 La vision et les statistiques de scènes naturelles	19
2.2 Le choix du domaine transformé	23
2.3 L'évaluation de la qualité sans référence basée sur les SSN	24
2.4 Application en biométrie	32
2.5 G-BLIINDS2 : un modèle SSN généralisé pour une métrique agnostique . . .	39
2.6 Conclusion	50

DANS ce chapitre, vont être abordées les notions de statistiques de scènes naturelles qui seront appliquées pour définir une métrique agnostique de la qualité des images sans référence. Ces travaux ont été réalisés en collaboration avec mes collègues de l'Université du Texas à Austin, le Pr. Alan C. Bovik et l'étudiante au doctorat Michèle Saad.

Une application de ces travaux en biométrie impliquant mes collègues du GREYC (Christophe Rosenbeger et l'étudiant au doctorat Mohamad El Abed) est également présentée.

2.1 La vision et les statistiques de scènes naturelles

La vue fait partie des cinq sens dont sont pourvus les êtres humains¹. Ce sens permet d'apprécier l'apparence superficielle de notre monde, *i.e.*, que ce que nous observons n'est que la réflexion de la lumière sur un objet, qui selon la ou les longueurs d'ondes réfléchies par

1. même si une légende qui remonte à des temps immémoriaux veut que la gente féminine possède un sixième sens...

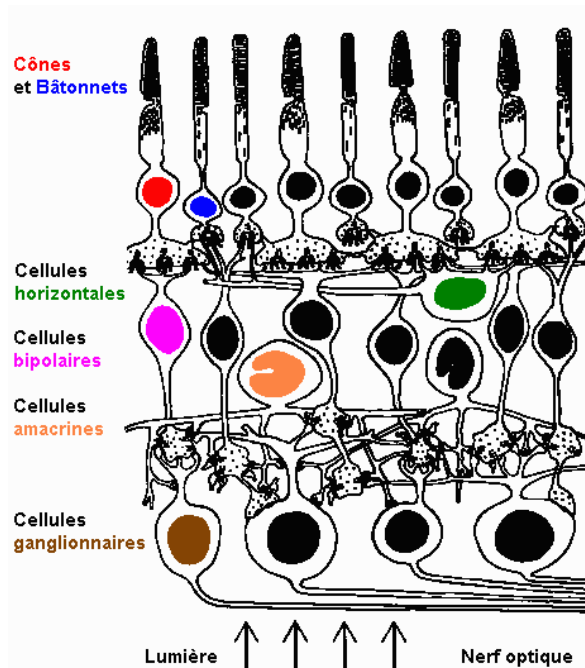


FIGURE 2.1: Schéma des différentes couches de cellules constituant la rétine.

l'objet aura une couleur particulière. Ainsi, le système visuel humain, et plus particulièrement la rétine, organe en première ligne, contient des cellules sensibles à diverses longueurs d'ondes réfléchies par les objets.

De manière plus fine, la rétine est constituée de plusieurs couches qui semblent être disposées à l'envers. Alors que l'on s'attend à avoir en premier lieu les cellules sensibles aux diverses longueurs d'ondes, paradoxalement, elle se situe le plus profondément enfouies sous la surface rétinienne. Les cellules ganglionnaires dont les axones constituent le nerf optique sont, quant à elles, situées à la surface interne de la rétine (Fig. 2.1). Entre ces couches, se situe une succession de cellules permettant de raffiner une partie de l'information visuelle et de cellules de soutien. La couche de la rétine la plus éloignée de la lumière contient des cellules pigmentées contenant beaucoup de mélanine (substance noire qui donne à la peau son teint bronzé ou noir). Ces cellules permettent d'absorber les rayons lumineux qui n'ont pas été captés par les cellules sensibles. Cela permet de prévenir les effets de réflexion sur le globe oculaire, ce qui pourrait s'avérer désastreux sur la fiabilité du système visuel humain.

La couche cellulaire sensible à la lumière est constituée de photorécepteurs de deux types : les cônes et les bâtonnets. Les cônes sont de trois types définis par les longueurs d'ondes absorbées :

1. Type S (*Short*) pour les courtes longueurs d'onde (420 nm),
2. Type M (*Medium*) pour les longueurs d'onde moyennes (530 nm),
3. Type L (*Long*) pour les longueurs d'onde élevées (565 nm)

Ces trois types de détecteurs, qui avaient été en quelque sorte annoncés par Young dans sa théorie trichromatique, vont traduire en trois signaux tout élément de distribution spectrale donnée (au sens de l'analyse newtonienne de la lumière polychromatique).

Lorsque l'énergie lumineuse est absorbée par la substance photosensible, cette dernière se décompose et entraîne une série de réactions chimiques. Ces réactions chimiques bloquent les courants sodiques (de sodium ; Na⁺) dans la cellule, ce qui crée un courant négatif. Le signal négatif (hyperpolarisation), généré par les cônes et les bâtonnets, est ensuite simultanément transmis aux cellules bipolaires et aux cellules horizontales. Les cellules bipolaires convertissent le signal négatif en une excitation des cellules ganglionnaires dont les axones constituent le nerf optique. Les cellules horizontales, qui sont excitées par le signal négatif des cônes et des bâtonnets, projettent latéralement leur axone et inhibent les cellules bipolaires avoisinantes. Ceci permet d'augmenter le contraste entre les zones très éclairées et celles qui le sont moins. Le rôle des cellules amacrines, est moins bien connu. Elles ne sont activées que pendant une fraction de seconde lors de changement d'intensité lumineuse, et pourraient participer à une phase rapide d'adaptation ou de transition d'une luminosité à une autre. Les cellules ganglionnaires sont les cellules dont les axones constituent le nerf optique. Elles peuvent être stimulées lors de l'excitation lumineuse, ou inhibées par l'inhibition latérale. Ce sont ces cellules qui transmettent l'information lumineuse au cerveau.

C'est à partir des cellules ganglionnaires que le codage trichromatique serait remplacé par un codage faisant intervenir les couples antagonistes de Hering (rouge-vert, jaune-bleu, blanc-noir).

Malgré la complexité du système visuel humain, ce que nous voyons est très limité par rapport à tout ce qu'il y aurait à voir. Par exemple, les serpents voient surtout dans le spectre de l'infrarouge ; ils distingueront une pierre d'une éponge si elles sont à des températures différentes, tandis que l'être humain les distinguera si elles sont de couleurs ou de textures différentes, sans avoir aucune idée de leurs températures. Notre vue est limitée au spectre du visible.

Un principe qui prévaut dans les sciences de la vision veut que les caractéristiques du SVH telles qu'elles ont été brièvement décrites précédemment aient été façonnées par leur grande facilité d'adaptation au monde dans lequel l'être humain évolue. Par exemple, Mollon [44] ainsi que Párraga *et al.* [45] ont souligné que le nombre et le pic de la longueur d'onde des photorécepteurs de la rétine est le produit d'un processus de sélection naturelle. Cette hypothèse est également valide pour tout système d'acquisition haute qualité des images ou des vidéos du monde réel dans le spectre des longueurs d'onde visibles par le SVH. Cette hypothèse ne serait en revanche plus valable pour des données textuelles, des images de synthèse, des dessins, des peintures ou dès lors que le système d'acquisition s'écarte du domaine visible tels les images radar, les radiographies par rayons X, les sonars, etc.

La philosophie première des statistiques de scènes naturelles repose sur l'idée que les images naturelles sont des éléments d'un vaste espace vectoriel appelé l'espace des images. L'espace qu'occupent les images naturelles est extrêmement petit au regard de toutes les

scènes possibles [46]. Une manière d'interpréter les statistiques de scènes naturelles est d'isoler ce sous-ensemble (appelé variété d'images) et d'imposer un modèle probabiliste simple (uniforme ou gaussien, par exemple). Une autre manière d'aborder l'interprétation est de dériver des modèles probabilistes sur un espace vectoriel de dimension supérieure et d'appliquer toute probabilité significative sur la variété adéquate.

La recherche de descripteurs statistiques des images naturelles impose que l'image soit perçue comme une réalisation d'un processus stochastique spatial défini sur un domaine dans \mathbb{R}^2 . Ce domaine est supposé être soit une région rectangulaire continue ou une grille uniforme et finie. L'hypothèse la plus communément formulée lors de la modélisation des images est la stationnarité du processus de construction de l'image, *i.e.* les probabilités de l'image sont invariantes aux translations dans le plan image.

Une des modélisations classiques consiste à supposer que les images sont des processus de second ordre. Cependant, les observations tendent à infirmer cette hypothèse. Certaines statistiques d'ordre supérieur des images ont montré des propriétés intéressantes [47] et plusieurs recherches ont été menées sur ces moments d'ordre supérieur. Ceci induit que les statistiques ne suivent pas une distribution gaussienne, hypothèse très largement utilisée en traitement d'image. Par exemple, lors de la décomposition d'une image sur une base d'ondelettes, on constate que les histogrammes des coefficients d'ondelettes affichent une distribution leptokurtique (queue épaisse, un cône étroit à la valeur médiane et des corrélations importantes à travers les échelles). Ce comportement non-gaussien des images a été démontré en premier lieu par Field [46] et a été souvent modélisé par la suite [48, 49, 50, 51]. Huang [52] a montré que les images, lorsqu'elles sont filtrées par un filtre moyen de taille 8×8 ont une distribution avec un fort kurtosis, un cône étroit à zéro et une longue queue exponentielle. Il a également montré que les décompositions linéaires appliquées sur des images naturelles tendent à maximiser le kurtosis, ou toute autre mesure de non-uniformité. Ceci a induit le développement de nombreux modèles permettant la décomposition du signal sur des bases spatialement orientées d'approximativement un octave de largeur de bande [53], ce qui est très similaire aux nombreuses décomposition multi-échelles existantes. En outre, des résultats similaires sont obtenus en cherchant à minimiser les dépendances existantes entre les coefficients, et ce de manière linéaire, par analyse en composante indépendante [54]. Tous ces résultats justifient l'utilisation très large des ondelettes orthonormales dans les applications d'analyse d'images. L'utilisation des ondelettes de Gabor est également justifiée par le fait que les champs récepteurs des cellules simples du cortex visuel des animaux sont modélisables de manière efficace et fiable par des fonctions de Gabor [55].

Ces statistiques de scènes naturelles ont été explicitement utilisées dans diverses applications d'analyse et de traitement d'images, comme par exemple dans le domaine de la compression [56, 57, 51], de la modélisation des images [58, 59], etc.

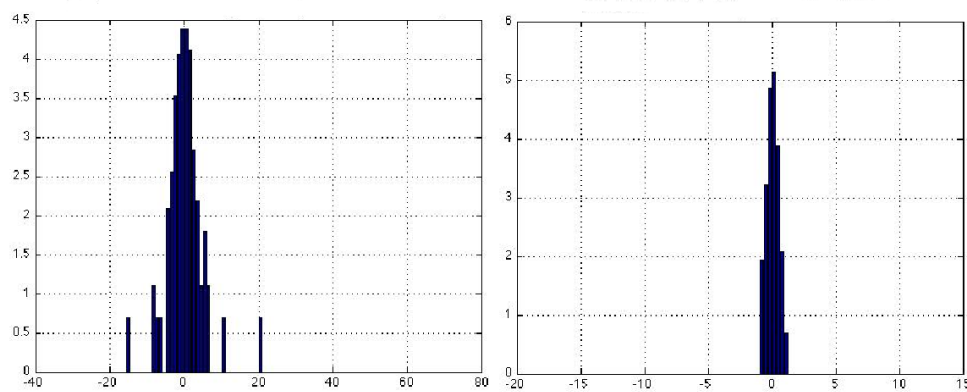
2.2 Le choix du domaine transformé

L'objectif intrinsèque de tout algorithme d'évaluation de la qualité d'une image est d'aboutir à un schéma de notation objectif pour lequel l'erreur de prédiction est minimale. Cette erreur de prédiction, qui est une mesure de performance, peut être considérée comme une fonction de représentativité naturelle des caractéristiques utilisées pour prédire la qualité.

Wang *et al.* [14] ont démontré que les images naturelles comportent de très fortes dépendances structurelles spatiales. Qui plus est, il a été vu précédemment que le SVH est adapté aux statistiques des images dans leur environnement naturel. Dès lors, l'idée retenue est de pouvoir définir des caractéristiques représentatives des structures de l'image dont les statistiques évoluent en fonction du type de dégradation et de la puissance de cette dernière.

L'information structurelle des images est localisée sur les contours visibles. Ces contours correspondent à des fréquences spatiales qui interfèrent de manière positive ou négative avec les fréquences restantes afin de produire des structures spatiales de l'image.

De plus, de nombreuses applications existantes de transmission des images ou des vidéos sont basées sur des transformations en blocs DCT. On trouve ainsi les standards de compression JPEG, MPEG2, H.263 et H.264. Une autre motivation à choisir le domaine transformé DCT repose sur le fait que l'histogramme des coefficients DCT d'une image est modifié en fonction du niveau de dégradation appliqué. La figure 2.2 illustre une telle observation entre l'histogramme obtenu pour une image originale et sa version dégradée par du flou gaussien.



(a) Histogramme des coefficients DCT dans le cas d'une image originale non dégradée. (b) Histogramme des coefficients DCT dans le cas de l'image originale utilisée en (a) dégradée par un flou gaussien.

FIGURE 2.2: Modification d'un histogramme des coefficients DCT dans le cas d'une image originale non dégradée (a) relativement à celui de la même image originale dégradée par flou gaussien (b).

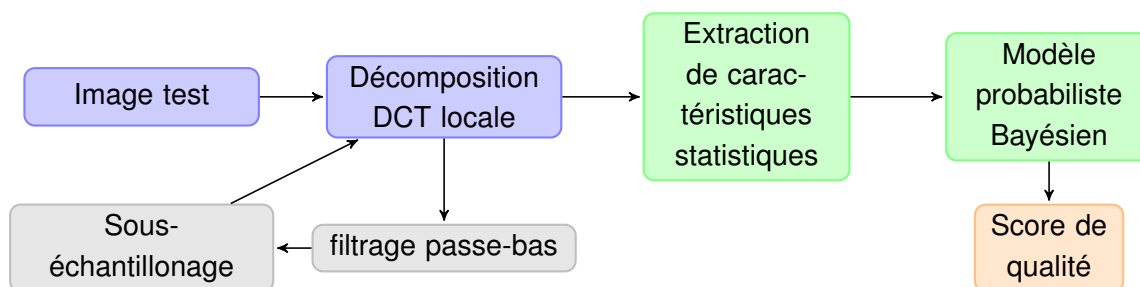


FIGURE 2.3: *Synopsis de la métrique de qualité sans référence agnostique basée sur une extraction de caractéristiques statistiques des scènes naturelles.*

2.3 L'évaluation de la qualité sans référence basée sur les SSN

L'efficacité d'un modèle est intrinsèque à la qualité des caractéristiques utilisées. Cela induit la sélection de caractéristiques suffisamment pertinentes pour mener à un outil fiable d'évaluation de la qualité des images. Selon les justifications formulées précédemment, ces caractéristiques sont développées dans le cadre de travail des statistiques de scènes naturelles alors que l'image d'origine n'est pas disponible.

2.3.1 Le modèle développé

La figure 2.3 présente le synopsis de la méthode développée. Après avoir transformé l'image test dans le domaine DCT en utilisant des blocs de décomposition de taille 17×17 , une extraction de caractéristiques basée sur le calcul d'attributs statistiques est opérée. Un modèle de prédiction bayésien est ensuite appliqué sur le vecteur de caractéristiques ainsi construit de manière à prédire le score de qualité de l'image test.

La largeur et la hauteur des blocs de décomposition locale a été fixée à la valeur 17. Le choix de cette valeur est justifié par le fait que la taille d'un bloc de nombre impair supérieur à 16 peut permettre d'éviter les problèmes de bord qui apparaissent quelquefois lors de l'utilisation des schémas de codage JPEG, MPEG-2 ou H.264.

Les diverses étapes du modèle développé sont présentées dans les sections suivantes, de même que les étapes utilisées lors de la phase d'apprentissage du modèle de prédiction des notes de qualité des images.

2.3.2 La sélection des caractéristiques basée sur des propriétés du SVH

Le processus de sélection des caractéristiques repose sur l'hypothèse fondamentale qu'une image naturelle est fortement structurée de par les fortes dépendances existantes entre les pixels. Ces dépendances sont alors interprétées comme des vecteurs d'information de la scène visuelle. Le SVH possède plusieurs propriétés qu'il convient de prendre en compte dans le choix des caractéristiques à développer. On retrouve ainsi :

- la forte sensibilité du SVH au contraste puisque tout être humain peut discerner plusieurs niveaux de luminance dans une image. Cette sensibilité varie en fonction de l'âge des individus, pour atteindre un maximum vers l'âge de 20 ans avec des fréquences spatiales de l'ordre de 2 à 5 cycles par degré. En outre, la sensibilité au contraste peut-être affectée non seulement par l'âge mais aussi par des pathologies de type cataracte, glaucome, rétinopathie diabétique conséquence d'une hyperglycémie chronique.
- la sensibilité du SVH aux distorsions des structures de l'image. Cette caractéristique vient de l'hypothèse formulée sous la théorie des statistiques de scènes naturelles qui stipule que le SVH est très fortement adapté à son environnement naturel et est capable d'extraire les informations structurelles de son champ visuel,
- la sensibilité du SVH aux objets nets (ce qui est à relier à la vision fovéale) due à la distribution non uniforme des photorécepteurs coniques et des cellules ganglionnaires de la rétine. Ainsi, la région autour du point de fixation est échantillonnée avec la plus forte résolution spatiale. À mesure que la distance au point de fixation augmente, la résolution diminue, ce qui induit une introduction de flou dans la vision périphérique.
- la sensibilité du SVH aux orientations des structures, également appelée sensibilité anisotropique, qui est un des aspects fondamentaux de la vision humaine. Alors que la sensibilité à l'orientation des champs récepteurs de la rétine est démontrée depuis de nombreuses années, les données issues d'IRM fonctionnelle montrent également que l'aire V1 du cortex visuel primaire joue un rôle prépondérant dans la reconnaissance des orientations cardinales et obliques [60].

Il est à noter que la troisième propriété énoncée ci-dessus dépend fortement du contenu de l'image. Dès lors, l'approche retenue ne constituera pas à quantifier le niveau de flou (ou de son dual : la netteté) en tant que telle, mais bien d'exploiter les variations des statistiques calculées dans le domaine spatio-fréquentiel mesurées entre l'image originale et l'image dégradée.

2.3.3 La modélisation des propriétés retenues

Une étude des histogrammes des coefficients AC révèle un cône pointu sur la valeur médiane ainsi qu'une large queue. Ceci est révélateur d'une distribution leptokurtique des coefficients DCT, à l'instar de ce qui a été formulé précédemment avec les coefficients d'ondelettes. Dès lors, la modélisation des quatre propriétés énoncées précédemment s'effectuera dans le domaine de la DCT et non dans le domaine des ondelettes.

2.3.3.1 La mesure du contraste

Parmi les nombreuses mesures de contraste existantes dans la littérature, la mesure globale du contraste de Michelson et celle de Peli sont probablement les plus fréquemment utilisées. Alors que la première n'est pas fortement corrélée avec la perception visuelle humaine et que la seconde est relativement complexe à déployer, nous proposons de calculer la valeur du contraste dans le domaine de la DCT. Une valeur de contraste est calculée pour chacun des blocs générés à partir de l'image :

$$c^k(x) = \frac{1}{N} \sum_{i=1}^N \frac{x_{AC}^i}{x_{DC}} \quad (2.1)$$

où N représente la taille du bloc DCT, x_{DC} représente la composante continue de la DCT et l'ensemble des valeurs x_{AC}^i fait référence aux coefficients AC de la DCT.

La mesure de contraste local est alors définie comme

$$v_1 = \frac{1}{M} \sum_{i=1}^M c^i(x) \quad (2.2)$$

où M représente le nombre de blocs de l'image.

La figure 2.4(c) présente les cartes de distorsion respectivement obtenues lors du calcul du contraste pour les trois images tests. L'image de droite de la figure 2.4(a) est une image exempte de dégradation, l'image du centre correspond à une image dégradée par application d'un bruit blanc et l'image de droite est une image compressée par le standard JPEG.

2.3.3.2 La mesure de distorsion aux structures

Les caractéristiques associées à la mesure de distorsion aux structures des images sont dérivées localement des coefficients DCT. Du fait de la forte amplitude de la valeur continue (DC) obtenue suite à l'application de la DCT, seuls les coefficients correspondant aux composantes AC sont considérés.

Afin de mesurer la distorsion des structures, l'histogramme des valeurs AC est considéré. En effet, il a été observé, dans ce cas, que l'amplitude du cône est inversement proportionnelle à la force de la dégradation appliquée [61]. De même, une diminution de la largeur de la queue proportionnelle à la force de la dégradation a été constatée. Ainsi, pour mesurer ces variations, le kurtosis est calculé sur l'histogramme des coefficients DCT :

$$\kappa(x) = \frac{E(x - \mu)^4}{\sigma^4} \quad (2.3)$$

où μ représente la moyenne des coefficients DCT et σ l'écart-type associé.

Finalement, le kurtosis sur chaque bloc DCT est calculé. Seules les valeurs situées en dessous du 10è percentile sont prises en compte pour calculer la moyenne finale, quantifiant la distorsion de structure.

La figure 2.4(b) présente les cartes de distorsion respectivement obtenues lors du calcul du Kurtosis pour les trois images tests affichées en 2.4(a).

2.3.3.3 Les mesures d'anisotropie

Dès lors que l'on applique une dégradation à une image originale, cela induit une perte d'information dans l'image. Or, l'entropie (telle que définie par Shannon) est une mesure d'information disponible dans une source. Il est tentant d'utiliser cette mesure de l'information dans un processus de mesure de la qualité d'une image. Cependant, Gabarda *et al.* [40] ont souligné que si analytiquement l'entropie augmente avec la netteté de l'image, pratiquement, on observe une faible corrélation entre les valeurs entropiques et les images dégradées par un bruit de type quelconque.

Précédemment, il a été fait mention de la variation de sensibilité du SVH aux orientations des structures de l'image. De ce fait, il est proposé d'utiliser la notion d'anisotropie afin de mesurer la perte d'information selon les diverses orientations auxquelles le SVH est sensible. Dans ce cas, l'entropie sera utilisée pour mesurer la quantité d'information selon les différentes orientations, afin d'obtenir une mesure d'entropie directionnelle.

La mesure d'entropie a été initialement proposée en 1948 par C. Shannon [62] et N. Wiener [63] comme une mesure de l'information portée par un symbole, provenant d'une source d'information stochastique. Cette mesure est très utilisée dans les schémas de compression (huffman, shanon-fano, etc.). En 1960, A. Rényi a repris cette notion et a développé une mesure d'entropie généralisée, connue comme l'entropie de Rényi [64]. L'entropie de Rényi d'ordre α calculée dans un espace spatio-frequentiel 2D est définie par :

$$R_\alpha = \frac{1}{1-\alpha} \log_2 \left(\sum_i P_i^\alpha \right), \quad \forall \alpha \geq 0, \alpha \neq 1. \quad (2.4)$$

où P_i représente la probabilité de chaque symbole i de l'espace spatio-fréquentiel 2D. Comme rapporté par Flandrin *et al.* [65], il est intéressant de noter que l'expression de l'entropie de Shannon H est obtenue lorsque $H = \lim_{\alpha \rightarrow 1} R_\alpha$. De même, les auteurs préconisent l'utilisation de valeurs de $\alpha \geq 2$ pour l'utilisation des distributions spatio-frequentielles.

Afin d'aboutir à une mesure d'anisotropie, l'entropie de Rényi est calculée sur l'ensemble des orientations $\Theta = \theta_1, \theta_2, \dots, \theta_m$. Pour ce faire, à un pixel de l'image sont associées m fenêtres 1D d'orientation θ_i . Une transformation DCT 1D est ensuite appliquée sur chacun de ces m fenêtres. Sous ces conditions, l'équation (2.4) s'écrit alors

$$R_{\theta,\alpha}[n] = \frac{1}{1-\alpha} \log_2 \left(\sum_n \sum_k \tilde{P}_\theta^\alpha[n,k] \right), \quad \forall \alpha \geq 0, \alpha \neq 1. \quad (2.5)$$

où θ est une des orientations de l'ensemble Θ , $\tilde{P}[n, k]$ représente le coefficient DCT normalisé situé à l'indice k de la fenêtre orientée 1-D associée et n représente l'indice du bloc DCT 1D. Un processus de normalisation est effectivement nécessaire pour satisfaire la condition $\sum_k \tilde{P}_\theta[n, k] = 1, \forall n : 1 \leq n \leq M$ où M est la taille des données et $-N/2 \leq N \leq N/2$; elle est formulée comme suit

$$\tilde{P}_\theta[n, k] = \frac{P_\theta[n, k]^2}{\sum_k P_\theta[n, k]^2} \quad (2.6)$$



Cette normalisation est inspirée par la mécanique quantique où la fonction d'onde Ψ représente l'amplitude de la probabilité pour trouver une particule à un point fixé dans l'espace à un moment donné. La probabilité de trouver une particule est donnée par le produit de la fonction d'onde avec son complexe conjugué Ψ^* de manière à satisfaire la condition $\int \Psi \Psi^* dr = 1$.

Finalement, deux mesures d'anisotropie H_1 et H_2 sont développées comme suit :

$$H_1 = \sigma^2 [\mathbb{E}(R_{\theta, \alpha}[n])], \quad (2.7)$$

et

$$H_2 = \max [\mathbb{E}(R_{\theta, \alpha}[n])], \quad (2.8)$$

où $\mathbb{E}[\cdot]$ représente l'espérance mathématique de l'entropie de Rényi $R_{\theta, \alpha}[n]$ calculées sur les blocs DCT 1D de même orientation θ . L'entropie de Rényi d'ordre 3 a été utilisée dans la suite de nos travaux.

La figure 2.4(d) présente les cartes de distorsion respectivement obtenues lors du calcul du maximum d'entropie de Rényi pour les trois images tests affichées en 2.4(a). La figure 2.4(e) présente, quant à elle, les cartes de distorsion obtenues lors du calcul du maximum d'entropie de Rényi.

2.3.4 L'approche multirésolution

La perception des défauts dépend de nombreux facteurs, dont notamment la résolution des images, la distance de visualisation et l'acuité du système visuel. Pour prendre en compte la variabilité de ces facteurs, une approche multirésolution sur n niveaux est développée.

L'extraction des caractéristiques est itérée de la façon suivante : après application d'un filtre passe bas de type filtre à noyau gaussien rotationnel symétrique, l'image est sous-échantillonnée par un facteur 2. Les blocs ne sont pas exclusifs dans le sens où un chevauchement de deux pixels existe entre deux blocs consécutifs. Ceci permet de prendre en compte les interactions existantes entre deux pixels voisins mais qui ne sont pas situés dans deux blocs consécutifs distincts.

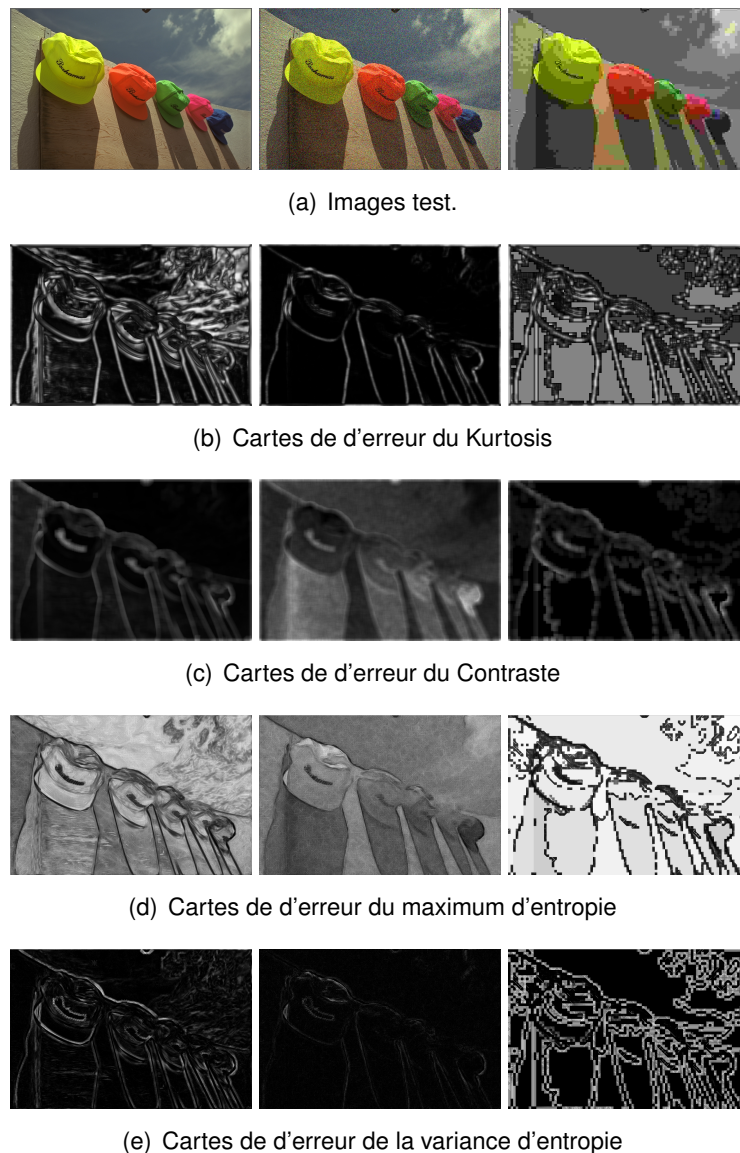


FIGURE 2.4: *Exemple des cartes de dégradations obtenues sur trois images tests (a) dont une est exempte de dégradation (à gauche). L'image du centre a été dégradée par application d'un bruit blanc tandis que l'image de droite a subi une compression JPEG à 0.15312 bpp. Les sous-figures (b) correspondent à la carte d'erreur du Kurtosis tandis que les sous-figures (c) correspondent à la carte d'erreur du contraste. Les sous-figures (d) présentent les cartes de distorsion liées au maximum d'entropie, alors que les sous-figures (e) présentent les cartes de distorsion liées à la variance d'entropie.*

2.3.5 Le modèle de prédiction probabiliste

Soit $X_i = \{x_1, x_2, \dots, x_n\}$ l'ensemble des caractéristiques calculées sur l'image candidate i . Disposant d'une vérité terrain y_i pour chacune des images i , (sous la forme du score subjectif moyen), l'objectif est alors de modéliser la distribution du couple (X_i, y_i) .

Deux modèles probabilistes ont été utilisés pour réaliser une telle modélisation. Il s'agit de la distribution gaussienne multivariable et de la distribution laplacienne multivariable. L'objectif étant dans ce cas de déterminer les paramètres de chacun de ces modèles. La distribution gaussienne multivariable est définie par :

$$P_X(x) = \frac{1}{(2\pi)^{d/2} \sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}q(x)\right) \quad (2.9)$$

où $q(x) = (x - \mu)^t \Sigma^{-1} (x - \mu)$ et Σ est la matrice de covariance interne de la variable X .

Dans le cas de la distribution gaussienne multivariable, la moyenne empirique μ et la matrice de covariance Σ des données sont utilisées pour estimer les paramètres de la distribution. La distribution laplacienne multivariable est définie par [66] :

$$P_X(x) = \frac{1}{(2\pi)^{d/2}} \frac{2}{\lambda} \frac{K_{(d/2)-1} \sqrt{\frac{\lambda}{2} q(x)}}{\sqrt{\frac{\lambda}{2} q(y)}^{(d/2)-1}} \quad (2.10)$$

où $q(x)$ et Σ sont définis précédemment et $K_m(x)$ la fonction de Bessel modifiée de deuxième espèce d'ordre m évaluée en x .

Les paramètres de la distribution laplacienne multivariable μ, Σ, λ sont estimés à partir des données empiriques. Dans ce cas, μ représente la moyenne de l'échantillon. En notant C la matrice de covariance de l'échantillon, on obtient [66]

$$\lambda = \det(C)^{1/d} \quad \text{et} \quad \Sigma = \frac{1}{\lambda} C \quad (2.11)$$

Le poids des critères de qualité n'est pas identique, dans le sens où chacune des caractéristiques n'a pas nécessairement la même influence sur la perception globale de la qualité de la scène. Ceci est à mettre en relation avec les échelles internes mises en œuvre lorsqu'un observateur humain prend une décision. Dès lors, pour simuler cette différence d'influence entre les attributs, un facteur de pondération α_i est calculé de manière à satisfaire la condition $\sum_{i=1}^N \alpha_i = 1$. La valeur de ces exposants est déterminée par le coefficient de corrélation de Spearman $SROCC_i$ calculé entre la valeur de chacun des attributs i et les valeurs DMOS. Ceci revient à calculer la confiance que l'on accorde aux attributs à prédire correctement la qualité des images. Finalement, les coefficients α_i sont normalisés :

$$\alpha_i = \frac{SROCC_i}{\sum_{k=1}^n SROCC_k}, \forall 1 \leq i \leq n \quad (2.12)$$

Corrélation SROCC pour le modèle gaussien					
Dégradation	1 niveau	2 niveaux	3 niveaux	4 niveaux	5 niveaux
JPEG2000	0.951	0.953	0.951	0.950	0.949
JPEG	0.939	0.940	0.942	0.944	0.943
Bruit blanc	0.975	0.977	0.978	0.977	0.978
flou gaussien	0.934	0.941	0.944	0.944	0.941
fast fading	0.904	0.916	0.927	0.927	0.926
base entière	0.825	0.908	0.912	0.911	0.910
Corrélation SROCC pour le modèle laplacien					
Dégradation	1 niveau	2 niveaux	3 niveaux	4 niveaux	5 niveaux
JPEG2000	0.950	0.951	0.949	0.948	0.948
JPEG	0.937	0.942	0.935	0.936	0.938
Bruit blanc	0.977	0.968	0.977	0.980	0.978
flou gaussien	0.921	0.938	0.945	0.945	0.943
fast fading	0.910	0.926	0.931	0.928	0.930
base entière	0.819	0.901	0.909	0.907	0.908

TABLE 2.1: Coefficients de corrélation de la métrique appliquée sur la base test LIVE pour divers niveaux de résolution.

2.3.6 Résultats

Cette approche a été testée sur la base d'images LIVE [67] ainsi que sur la base TID2008 [68].

Une base d'apprentissage a été construite à partir d'une partie des images de la base LIVE de manière à respecter la distribution des valeurs DMOS. Parmi les 29 images de référence et leurs versions dégradées disponibles dans la base LIVE, 15 images de référence (avec leurs versions dégradées) ont été choisies afin de définir la base d'apprentissage sur laquelle les paramètres des modèles ainsi que les valeurs des exposants ont été calculés. De plus, une technique de bootstrap à 999 répliques a été utilisée afin de renforcer la validité des résultats.

Le tableau 2.1 présente les taux de corrélation obtenus en fonction du nombre de niveaux de résolution utilisés. Une stabilisation du niveau de corrélation SROCC est observée à compter du troisième niveau de décomposition, ce qui signifie que seuls trois niveaux de résolution sont utiles pour obtenir un haut niveau de corrélation. Cette remarque est valable quel que soit le modèle utilisé, à savoir le modèle gaussien multivariable et le modèle laplacien multivariable.

Le tableau 2.2 contient les valeurs du coefficient de corrélation de Spearman (SROCC) obtenues pour la méthode proposée selon un modèle gaussien (G-BLIINDS), un modèle laplacien (L-BLIINDS) et la mesure du PSNR. Alors que le PSNR nécessite l'utilisation de l'image de référence, les résultats obtenus montrent que la méthode développée permet de prédire le score de qualité des images de manière plus fiable pour quasiment tous les types de dégradation présents dans la base d'images LIVE. Les valeurs de corrélation sont meilleures pour

Dégradation LIVE	G-BLIINDS	L-BLIINDS	PSNR
JP2K	0.951	0.949	0.8658
JPEG	0.942	0.935	0.8889
Bruit blanc	0.978	0.977	0.9791
Flou gaussien	0.944	0.945	0.7887
Fast Fading	0.927	0.931	0.8986
Base entière	0.912	0.909	0.8645

TABLE 2.2: *Corrélation de Spearman (SROCC) entre les scores prédits par la méthode proposée selon un modèle gaussien (G-BLIINDS), un modèle laplacien (L-BLIINDS) et la mesure du PSNR sur la base de test.*

une dégradation par bruit blanc lorsque la prédiction de la qualité est effectuée par le PSNR. Cependant, cette différence n'est pas statistiquement significative. Lorsque la base de test extraite de la base LIVE est considérée, les valeurs de corrélation sont plus élevées pour l'indice G-BLIINDS et L-BLIINDS que pour le PSNR.

Ceci tend à montrer que l'approche retenue permet d'obtenir de très bon taux de corrélation avec les scores fournis par les observateurs humains, et est supérieure en terme de qualité de prédiction à la mesure du PSNR, qui est encore très couramment utilisée pour juger de la qualité d'une image. De plus, la technique développée est une approche sans référence au contraire du PSNR.

La figure 2.5 présente les résultats obtenus lors de la mesure de la qualité des images dégradées. Les quatre images présentées sont extraites de la base TID2008 et ont été respectivement dégradées (de gauche à droite) par un bruit de haute fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit du à une insertion de blocs de couleur et d'intensité différente. Les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellente). Si on compare les scores de la deuxième et de la quatrième image, on constate que les scores donnés par la métrique G-BLIINDS sont en adéquation avec les scores subjectifs (MOS), ce qui n'est pas le cas pour les métriques L-BLIINDS et le PSNR. Les trois métriques échouent à prédire correctement le score de la première image, puisque les valeurs obtenues sont les plus basses (ce qui correspond à des images de mauvaise qualité) alors que la valeur du MOS est la plus élevée.

2.4 Application en biométrie

La technique d'évaluation de la qualité BLIINDS présentée précédemment a été utilisée avec succès pour mesurer la qualité de données biométriques morphologiques. Ces travaux ont été réalisés avec mes collègues Christophe Rosenberger (professeur à l'ENSICAEN) et

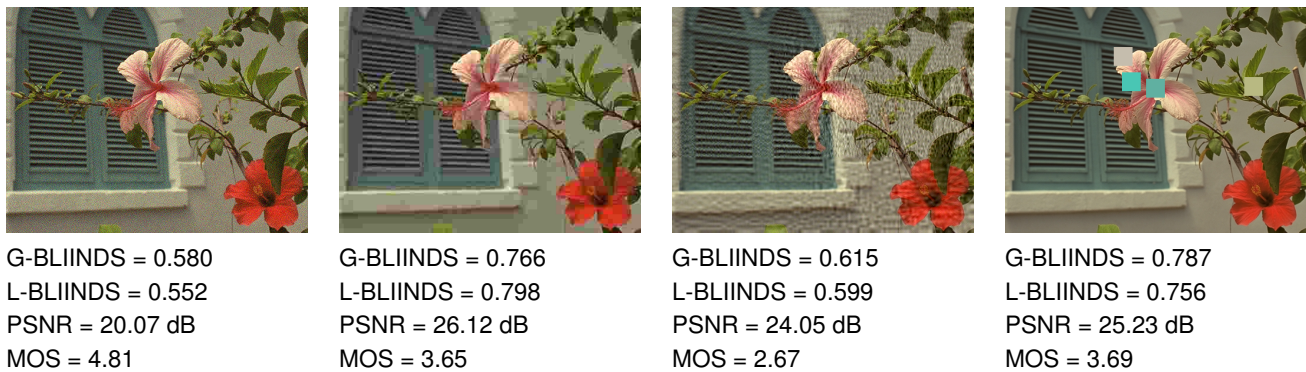


FIGURE 2.5: Scores obtenus avec la méthode de mesure de qualité sans référence G-BLIINDS et L-BLIINDS ainsi qu'avec le PSNR. Les images sont extraites de la base TID2008. De gauche à droite les dégradations subies sont un bruit haute-fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit du à une insertion de blocs de couleur et d'intensité différente. les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellent)

Mohamad El Abed (doctorant au GREYC). Une partie des recherches menées dans la thèse de Mohamad concerne le développement d'une méthode d'évaluation de la qualité des données biométriques.

Vouloir quantifier la qualité d'une donnée biométrique ne peut se faire uniquement en se basant sur la qualité de l'image *stricto sensu*. En effet, la qualité des données biométriques s'appuie sur l'information portée par une caractéristique biométrique de type empreintes digitales, visages, etc. Par exemple, une image d'empreinte digitale peut être de bonne qualité mais cela ne signifie par pour autant que les crêtes des lignes soient parfaitement exploitables, car elles peuvent être biologiquement dégradées.

2.4.1 Méthode d'évaluation de la qualité des données biométriques

La méthode développée repose sur l'exploitation des résultats de mesure de qualité des images et de l'information biométrique représentée, tel que décrit dans la figure 2.6.

À partir d'une image originale, le calcul de la note de qualité sans référence est réalisé selon l'algorithme G-BLIINDS précédemment décrit. Parallèlement à ce calcul, une extraction des points d'intérêt sur l'image est effectuée. Ce choix est motivé par le fait que ces points permettent d'avoir une description des zones de l'image dans lesquelles l'information est importante, et ce afin d'obtenir une représentation indépendante de la modalité biométrique. La technique des points d'intérêt a été utilisée avec succès dans des problématiques de reconnaissance de forme ou de texture, de localisation de robot, de forage de données vidéo et de construction de panorama [69]. L'algorithme utilisé dans notre étude est l'algorithme SIFT

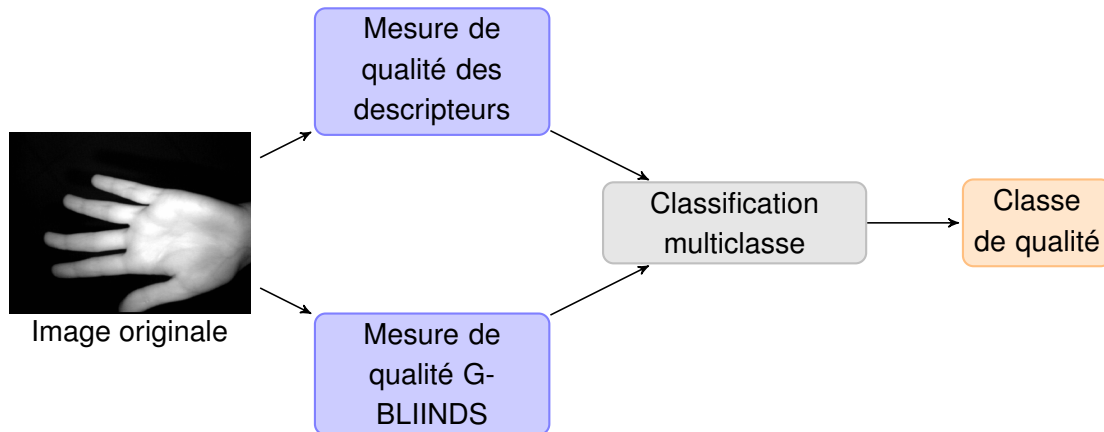


FIGURE 2.6: Synopsis de la méthode d'évaluation de la qualité des données biométriques

(*Scale Invariant Feature Transformation*) [70] qui a montré sa robustesse et sa supériorité dans le processus d'extraction des points d'intérêt vis-à-vis des techniques d'extractions de points locaux d'intérêt telles que les filtres orientables, PCA-SIFT, les invariants différentiels, les filtres complexes, etc. [69].

Une image I est alors caractérisée par l'ensemble $Y(I)$ des points d'intérêt défini par $Y(I) = \{k_i = (x_i, y_i, e_i, o_i, v_i) | \forall i \in [1, N(I)]\}$ où $N(I)$ fait référence au nombre de points d'intérêt obtenus pour l'image I . (x_i, y_i) sont les coordonnées cartésiennes du point d'intérêt i ; les paramètres e_i et o_i font respectivement référence à l'échelle et à l'orientation du point i . v_i est un vecteur descripteur du point i contenant 128 valeurs. Ces valeurs sont obtenues à partir d'un ensemble d'histogrammes d'orientation sur des voisinages du pixel i de taille 4×4 . Le gradient est calculé sur chacun de ces 16 points de la grille. Il est ensuite quantifié par un histogramme à huit orientations. Dès lors, le vecteur descripteur SIFT d'un point d'intérêt contient $4 \times 4 \times 8 = 128$ éléments.

À partir de ces données, quatre attributs sont extraits (leurs intérêts ont été mis en évidence par analyse statistique de leur comportement) :

1. le nombre de points d'intérêt détectés dans l'image I ,
2. le coefficient DC issue de la TCD de la matrice de taille $N(I) \times 128$ contenant les vecteurs $(v_k)_{1 \leq k \leq N(I)}$ des $N(I)$ points d'intérêt de l'image I ,
3. la moyenne des valeurs d'échelle utilisées par les $N(I)$ points d'intérêt,
4. l'écart-type des valeur d'échelle utilisées par les $N(I)$ points d'intérêt.

Au final, le vecteur contient cinq attributs : les quatre valeurs précédemment décrites auxquelles s'ajoute le score de qualité d'image. Au lieu de procéder au calcul d'une valeur scalaire reflétant la qualité des données biométriques, un classificateur multiclassée a été utilisé. Le classificateur retenu est le séparateur à vaste marge (SVM) qui est décrit dans la section 3.2. La règle de combinaison utilisée est celle du vote majoritaire.

Catégorie	Dégradations	Sémantique
I	Image originale	Bonne
II	$mvt(l_1), bacq(l_1), dacq(l_1)$	Moyenne
III	$mvt(l_2), bacq(l_2), dacq(l_2)$	Mauvaise
IV	$mvt(l_3), bacq(l_3), dacq(l_3)$	Très mauvaise

TABLE 2.3: *Catégories de qualité retenues dans le cadre de l'évaluation de la qualité des données biométriques.*

Le choix d'utiliser une méthodologie de classification en lieu et place du calcul d'un score de qualité est principalement motivé par le fait qu'il n'existe pas de base de données biométriques contenant des informations afférentes à la qualité de la donnée.

2.4.2 Mesure de performances

2.4.2.1 Apparatus

Cinq bases de données ont été utilisées pour quantifier les performances de la méthodologie précédemment proposée. Parmi ces cinq ensembles, quatre contiennent des données de visage (FACES94 [71], ENSIB [72], FERET [73] et AR [74]) et la dernière contient des données d'empreintes digitales (FVC2002 DB₂ [75]).

Pour chacune de ces bases, trois types de dégradation ont été appliqués. Ils permettent de simuler trois artéfacts d'acquisition qui surviennent habituellement lors de l'utilisation du système par un utilisateur. La simulation de la dégradation liée au mouvement ($mvt(.)$) est réalisée par l'application d'un flou gaussien. Le bruit d'acquisition ($bacq(.)$) est simulé par un bruit gaussien tandis que la distance d'acquisition ($dacq(.)$) est simulée par une opération de sous-échantillonnage de l'image.

Pour chacun de ces artéfacts, trois niveaux de distorsion ($l_k)_{1 \leq k \leq 3}$ ont été appliqués sur chacune des bases test. Pour chacune de ces bases (contenant les images de référence et les versions dégradées associées), une base d'apprentissage ($TrainS_i)_{1 \leq i \leq 5}$ et une base de test ($TestS_i)_{1 \leq i \leq 5}$ ont été construites de manière équilibrée (*i.e.*, chacun des deux ensembles contient le même nombre d'exemples par classe). Par ailleurs, un sixième ensemble d'apprentissage ($TrainS)_{all}$ a été construit en sélectionnant des images parmi les quatre bases de visage. Les images non sélectionnées constituent la base de test ($TestS)_{all}$.

L'objectif de la méthode étant de réaliser une classification de la qualité de la donnée biométrique, un nombre total de quatre classes a été retenu, tel que présenté dans le tableau 2.3.

Attributs	mvt(.)	bacq(.)	dacq(.)
Nombre de points d'intérêt	-0.573	0.390	-0.488
Coefficient DC	-0.616	0.567	-0.525
Moyenne des échelles	0.793	-0.563	-0.396
Écart-type des échelles	0.347	-0.347	-0.473
G-BLIINDS	0.632	-0.801	-0.102

TABLE 2.4: Valeurs du coefficient LCC calculé entre les cinq attributs et les altérations sur les bases de visages. Les valeurs mises en valeur correspondent à de fortes corrélations.

2.4.2.2 Mesure de performance de la méthode développée dans l'estimation de la qualité

Comme précédemment mentionné, cinq attributs sont calculés pour permettre l'évaluation de la qualité des données biométriques. La capacité de chacun de ces attributs à distinguer les altérations utilisées est étudiée. Pour ce faire, le coefficient de corrélation linéaire de Pearson (LCC) est mesuré entre chacun des cinq critères et pour chaque type de dégradation, et ce pour les quatre bases de visage. Le tableau 2.4 présente les résultats obtenus après calcul du coefficient LCC. Le coefficient DC est l'attribut le plus pertinent pour détecter les trois types d'altération. L'attribut lié au score de qualité des images G-BLIINDS est fortement corrélé avec deux types de dégradations : mvt(.) et bacq(.) respectivement simulés par un flou et un bruit gaussien. En revanche, la dernière dégradation dacq(.) n'est pas détectée par cet attribut. Ceci n'est pas étonnant puisque cette dégradation (simulée par un processus de sous-échantillonnage) n'altère pas le contenu de l'image originale. La moyenne des échelles ainsi que le nombre de points d'intérêt se révèle également pertinent pour détecter les types de dégradation. L'écart-type des échelles est l'attribut qui présente la plus faible corrélation avec les types de dégradation utilisés.

Le tableau 2.5 présente les résultats obtenus lors de l'application de la méthode d'évaluation de la qualité des images de visage proposée sur chacune des cinq bases de test $(\text{TestS}_i)_{1 \leq i \leq 5}$ ainsi que sur l'ensemble $(\text{TestS})_{\text{all}}$. Quelque soit l'ensemble de test utilisé, les taux de reconnaissance sont supérieurs à 81.2%, ce qui permet d'en déduire que la méthode d'évaluation de la qualité des données biométriques est pertinente. De plus, les taux de reconnaissance sont globalement plus élevés pour les ensembles $(\text{TestS}_i)_{1 \leq i \leq 5}$ que pour l'ensemble $(\text{TestS})_{\text{all}}$, à l'exception de la base ENSIB. Néanmoins, ces différences ne se révèlent pas statistiquement significatives. Ces résultats ne sont pas surprenant étant donné que les images de référence présentes dans les quatre bases de visages ne sont pas drastiquement différentes les unes des autres, et sont de même complexité.

La figure 2.7 présente un exemple des classes prédites par la méthode développée pour une image de visage issue de la base FACE94 et une image d'empreinte digitale issue de la base DB₂. Chacune des images a subi trois altérations : 1) application d'un bruit gaussien de niveau 1 ($\text{bacq}(I_1)$), un redimensionnement de niveau 3 ($\text{dacq}(I_3)$) et un flou gaussien de niveau 3 ($\text{mvt}(I_3)$). Les prédictions contiennent non seulement les classes de qualité (telles que définies

Bases	Précision sur $(\text{TestS}_i)_{1 \leq i \leq 5}$	Précision sur $(\text{TestS})_{\text{all}}$
FACE94	0.867	0.853
ENSIB	0.898	0.911
FERET	0.812	0.812
AR	0.891	0.889
DB ₂	0.837	--

TABLE 2.5: Précision des modèles SVM multiclassés sur les ensembles de test $(\text{TestS}_i)_{1 \leq i \leq 5}$ et sur l'ensemble $(\text{TestS})_{\text{all}}$. La base DB₂ contient des images d'empreintes digitales tandis que les quatre premières contiennent des images de visage.

dans le tableau 2.3), mais aussi les dégradations associées.









Dégradation	Image Originale	Bruit gaussien $\text{bacq}(l_1)$	Redim. $\text{dacq}(l_3)$	Flou gaussien $\text{mvt}(l_3)$
				
Classe prédite	IV ($\text{mvt}(l_3)$)	II ($\text{bacq}(l_1)$)	IV ($\text{dacq}(l_3)$)	IV ($\text{mvt}(l_3)$)
Classe réelle	I	II	IV	IV
				
Classe prédite	I	II $\text{bacq}(l_1)$	IV ($\text{dacq}(l_3)$)	IV ($\text{mvt}(l_3)$)
Classe réelle	I	II	IV	IV

FIGURE 2.7: Classes prédites et réelles par la méthode développée pour une image de visage et une image d'empreinte digitale originales ayant subies une dégradation par bruit gaussien, une opération de redimensionnement de l'image et un flou gaussien. Les classes sont définies dans le tableau 2.3

2.4.2.3 Relation qualité-performance d'un système de vérification biométrique

Afin de quantifier l'efficacité de la méthode d'évaluation de la qualité des données biométriques, il est intéressant de voir dans quelle mesure il est possible de prédire la performance d'un système de vérification biométrique compte tenu de la qualité des données biométriques. Le système retenu est GREYC-Face, système de reconnaissance faciale développé au GREYC réalisant un calcul de similarité entre l'image du visage de l'individu et une nouvelle capture

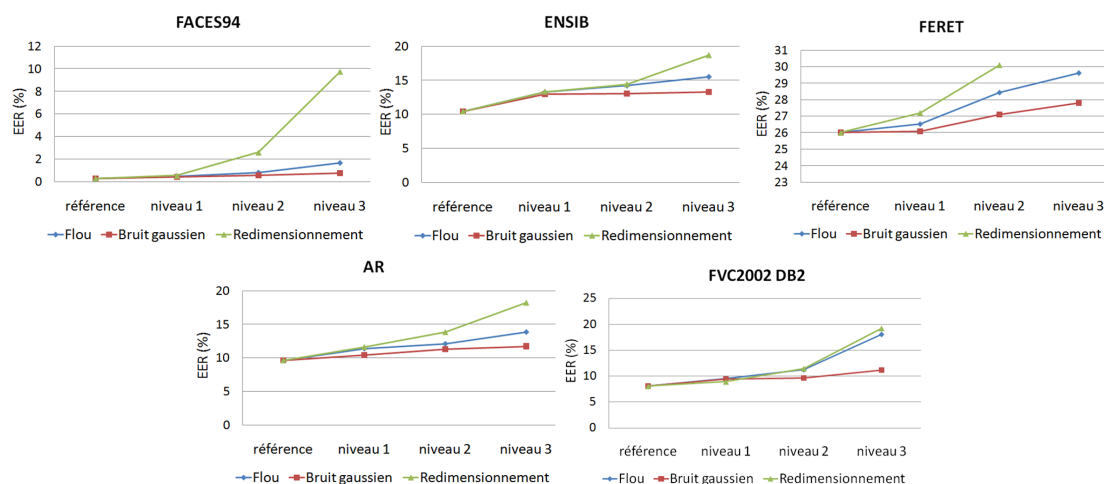


FIGURE 2.8: Valeurs de l'EER (en %) sur chacune des bases testées permettant de mesurer l'effet des dégradations sur la performance globale du système biométrique utilisé.

(phase d'enrôlement uniquement) [76]. Le taux d'égale erreur (*EER*–*Equal Error Rate*) est utilisé pour mesurer la performance du système biométrique utilisé. La valeur EER correspond au point d'intersection entre la courbe ROC et la droite d'équation $y = x$. Cette valeur constitue un bon indicateur de la précision d'un dispositif de vérification biométrique : plus la valeur de l'EER est faible, plus le système est performant. À noter que ce taux d'erreur est couramment utilisé pour illustrer la performance des systèmes biométriques.

La figure 2.8 présente l'impact du type et du niveau de dégradation de l'image sur la performance globale du système testé. Ces EER sont calculés en utilisant l'image de référence pour la phase d'enrôlement (*i.e.*, phase d'apprentissage), tandis que les images restantes servent pour la phase de test. Toutes les altérations ont un impact majeur sur la performance du système biométrique GREYC-Face (ceci simule un usage opérationnel du système où l'on ne garantit pas la qualité de l'image acquise). La performance décroît linéairement (*i.e.*, augmentation de la valeur EER) avec les niveaux de dégradation appliqués, et ce quel que soit le type de dégradation considéré.

Dans une dernière étape d'évaluation de la méthode proposée, une comparaison des performances obtenues est réalisée avec les résultats obtenus en utilisant la métrique de qualité NFIQ dédiée aux empreintes digitales et proposée par le *National Institute of Standards and Technology* (NIST) [77]. Afin de mesurer la différence des deux distributions de scores des utilisateurs légitimes (scores intra) et des imposteurs (scores inter), le test de Kolmogorov-Smirnov (KS) [78] est appliqué. Si la valeur du test est proche de 0, alors les deux distributions sont homogènes. Dans le cas contraire (*i.e.*, valeur proche de 1), les deux distributions sont indépendantes. Le tableau 2.6 décrit les valeurs statistiques du test KS pour chacun des catégories de qualité définies dans le tableau 2.3. La méthode proposée permet une meilleure séparation des deux distributions des utilisateurs légitimes et des imposteurs. Concernant la classe « Très

	Bonne qualité	Qualité moyenne	Mauvaise qualité	Très mauvaise qualité
Méthode proposée	0.869	0.828	0.797	0.624
NFIQ	0.820	0.697	0.632	0.640

TABLE 2.6: Comparaison entre la méthode d'évaluation de la qualité proposée et la métrique NFIQ avec des valeurs du test de KS pour un intervalle de confiance égal à 95%

mauvaise qualité », la méthode NFIQ est légèrement meilleure que la méthode proposée dans la capacité à séparer les deux distributions. Néanmoins, cette différence n'est pas statistiquement significative.

L'intérêt de cette méthode est d'être générique (*i.e.*, pouvant être appliquée sur tout type de données biométriques représentée par une image) et elle se révèle plus fiable dans le cas des empreintes digitales.

2.5 G-BLIINDS2 : un modèle SSN généralisé pour une métrique agnostique

La métrique développée dans la section 2.3 a permis de valider l'utilisation des SSN dans le domaine transformé de la DCT. Au lieu de modéliser le comportement de caractéristiques calculées dans le domaine DCT, l'idée est de dériver un modèle généralisé des SSN utilisant les coefficients DCT locaux, et de transformer ensuite les paramètres du modèle en caractéristiques permettant de prédire la qualité des images. A l'instar de ce qui a été réalisé précédemment, un modèle probabiliste généralisé est ensuite obtenu à partir de ces caractéristiques, ce qui permettra de quantifier la qualité des images.

Cette approche repose sur la capacité de l'algorithme d'évaluation de la qualité des images à apprendre la façon dont le modèle SSN varie le long des différents niveaux perceptuels des distorsions des images. L'algorithme est directement entraîné sur le modèle statistique paramétrique généralisé des coefficients DCT d'une image à divers niveaux de résolution.

Contrairement à ce qui a été réalisé en amont, quasiment aucune utilisation directe des modélisations perceptuelles spécifiques existantes (telles que les décompositions corticales de l'aire V1 du cortex, les effets de masquage, etc.) ne sera effectuée. Néanmoins, l'approche proposée se veut perceptuellement cohérente puisque les modèles des SSN se veulent le reflet des propriétés statistiques du monde environnant influençant l'évolution des fonctions perceptuelles du SVH.

La figure 2.9 détaille le synopsis global du processus d'évaluation sans référence et agnostique de la qualité des images. Une image candidate est tout d'abord décomposée en blocs de taille $n \times n$ sur lesquels une décomposition DCT bidimensionnelle est appliquée. Cette décomposition est opérée sur plusieurs niveaux de résolution. Ensuite, pour chaque bloc pris dans son

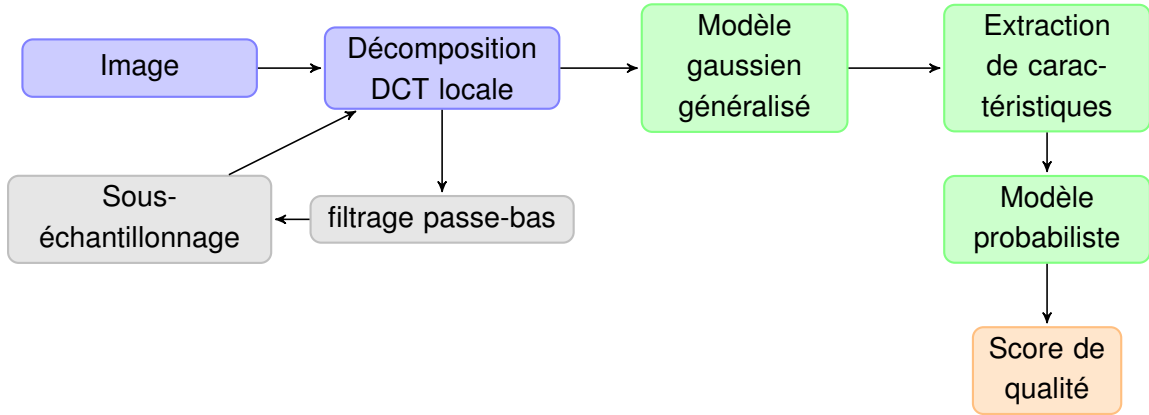


FIGURE 2.9: Synopsis de la métrique de qualité sans référence agnostique G-BLIINDS2 basée sur une modélisation statistique des scènes naturelles.

ensemble, un modèle de densité gaussienne généralisée est estimé. À ce stade, plusieurs partitions spécifiques du bloc sont également prises en compte afin de définir plusieurs modèles de densité. Les paramètres issus de ces diverses modélisations sont ensuite utilisés en tant que caractéristiques afin de prédire le score de qualité de l'image candidate. Cette prédiction est finalement effectuée à l'aide d'un modèle bayésien maximisant la probabilité que l'image candidate ait un certain score de qualité étant données les caractéristiques utilisées.

2.5.1 Le choix du modèle probabiliste généralisé

Du fait de sa particularité à inclure un large éventail du comportement observé des coefficients DCT, le modèle probabiliste généralisé s'impose. L'équation (2.9) s'écrit dans ce cas

$$P_X(x) = \alpha \exp(-\beta q(x))^\gamma \quad (2.13)$$

où α et β sont des constantes de normalisation et d'échelle définies par

$$\alpha = \frac{\beta \gamma}{2\Gamma(1/\gamma)} \quad (2.14)$$

$$\beta = \frac{1}{\sigma} \sqrt{\frac{\Gamma(3/\gamma)}{\Gamma(1/\gamma)}} \quad (2.15)$$

avec $\Gamma(\cdot)$ la fonction gamma définie par

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt \quad (2.16)$$

β est un paramètre d'échelle et γ est un paramètre de forme.

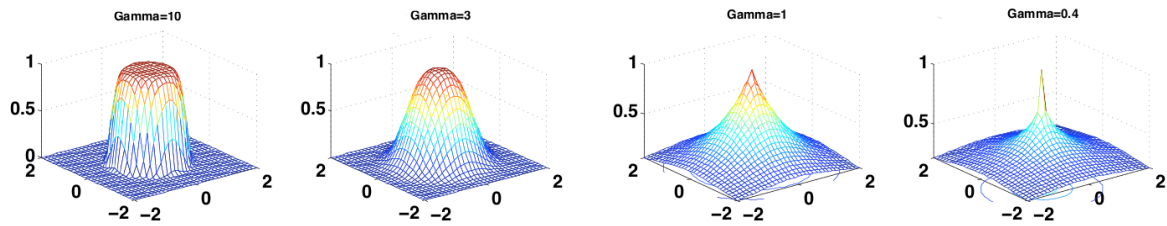


FIGURE 2.10: Illustration de l'évolution des distributions d'une gaussienne généralisée 2D en fonction des valeurs du paramètre γ

2.5.2 Les paramètres induits par le modèle probabiliste

Les paramètres qui serviront à mesurer la qualité d'une image de manière agnostique sont issus du modèle probabiliste décrit précédemment. Afin de valider un tel choix, la performance de chacun de ces paramètres sera étudiée et analysée indépendamment les uns des autres. Par ailleurs, du fait de la manipulation de données fréquentielles, un moyennage des résultats obtenus ne sera pas effectué sur la totalité des blocs, mais seules les valeurs du 10^è percentile seront utilisées pour le calcul du paramètre global. La raison est que les fait saillants (par exemple, les dégradations des contours) sont concentrés sur quelques valeurs locales, et que ces défauts influent très fortement l'impression subjective de la qualité d'un observateur humain. La sensibilité de l'observateur humain aux dégradations d'une image n'est pas une fonction linéaire du degré de distorsion appliquée. Afin de valider cette approche, une comparaison avec un moyennage classique sera réalisée pour chaque calcul des paramètres globaux.

2.5.2.1 Le paramètre de forme γ

La figure 2.10 montre l'évolution des distributions d'une gaussienne généralisée 2D en fonction des valeurs du paramètre γ . Ce paramètre rejoint l'hypothèse validée empiriquement lors du développement de l'algorithme d'évaluation de la qualité précédent, à savoir que la forme des histogrammes (pics, queues, etc.) évolue en fonction du degré de dégradation appliquée sur l'image.

Pour chaque bloc de taille $n \times n$, la densité gaussienne généralisée (Eq. 2.13) est estimée sur les coefficients DCT en excluant la composante continue DC.

L'estimation du paramètre suit la procédure suivante :

1. Calcul de la moyenne μ et de la variance σ^2 ,
2. Calcul de l'espérance mathématique $\mathbb{E}(|X|) = \frac{1}{M} \sum_{i=1}^M M|x_i - \mu|$,
3. Calcul du ratio $r = \sigma^2 / \mathbb{E}^2(|X|)$,

4. Recherche de la solution à l'équation $\gamma = \phi^{-1}(r)$ où $\phi(\cdot)$ est la fonction ratio gaussienne généralisée définie par

$$\phi(\gamma) = \frac{\Gamma(1/\gamma)\Gamma(3/\gamma)}{\Gamma^2(2/\gamma)} \quad (2.17)$$

Le critère global, basé sur le paramètre de forme γ de la densité de probabilité gaussienne, est obtenu par moyennage des valeurs locales du 10^è percentile. La figure 2.12(b) présente les cartes de distorsion respectivement obtenues lors du calcul du paramètre de forme γ pour les trois images affichées en 2.12(a).

2.5.2.2 Le coefficient de variation fréquentielle ζ

Dans notre étude, le coefficient de variation fréquentielle est défini par

$$\zeta = \frac{\sigma_{|X|}}{\mu_{|X|}} \quad (2.18)$$

où $\mu_{|X|}$ mesure le centre de la distribution des amplitudes des coefficients DCT, alors que $\sigma_{|X|}$ mesure la dispersion des amplitudes des coefficients DCT. Ce coefficient peut-être vu comme un processus de normalisation, interprétable par analogie avec le comportement physiologique des neurones. Dans la littérature en sciences de la vision, il a été montré qu'une opération de normalisation du contrôle de gain local est approprié et efficace pour prendre en compte les réponses neuronales des systèmes de vision biologique [79, 80]. En outre, les mécanismes de contrôle de gain non-linéaires sont construits sur des modèles de transformation linéaire où chaque réponse neuronale est normalisée par l'énergie associée à son voisinage.

Le paramètre ζ est calculé pour chacun des blocs de l'image. La valeur globale de variation fréquentielle est calculée comme la moyenne du 10^è percentile le plus élevé. La figure 2.12(c) présente les cartes de distorsion obtenues lors du calcul de variation fréquentielle ζ pour les trois images affichées en 2.12(a).

2.5.2.3 Le ratio d'énergie des sous-bandes ρ

Les distorsions appliquées à des images modifient les signatures spectrales locales d'une image de manière non naturelle. Afin de prendre en compte ces modifications, le ratio de l'énergie locale des sous-bandes DCT est mesuré et constitue une caractéristique pertinente.

Dans un bloc DCT, les coefficients de plus basses fréquences sont usuellement placés dans le coin supérieur gauche, tandis qu'à mesure que l'on se déplace diagonalement vers le coin inférieur droit, les coefficients sont de plus hautes fréquences. La figure 2.11(a) représente l'agencement des fréquences en trois sous-bandes fréquentielles.

Soit Δ_n l'ensemble de coefficients DCT appartenant à la sous-bande fréquentielle $n, \forall n \in [1, 2, 3]$. Le modèle de densité gaussienne généralisée est ensuite appliqué sur l'histogramme

DC	C ₁₂	C ₁₃	C ₁₄	C ₁₅	C ₁₆	C ₁₇
C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅	C ₂₆	C ₂₇
C ₃₁	C ₃₂	C ₃₃	C ₃₄	C ₃₅	C ₃₆	C ₃₇
C ₄₁	C ₄₂	C ₄₃	C ₄₄	C ₄₅	C ₄₆	C ₄₇
C ₅₁	C ₅₂	C ₅₃	C ₅₄	C ₅₅	C ₅₆	C ₅₇
C ₆₁	C ₆₂	C ₆₃	C ₆₄	C ₆₅	C ₆₆	C ₆₇
C ₇₁	C ₇₂	C ₇₃	C ₇₄	C ₇₅	C ₇₆	C ₇₇

(a) Sous-bandes fréquentielles des coefficients DCT

DC	C ₁₂	C ₁₃	C ₁₄	C ₁₅	C ₁₆	C ₁₇
C ₂₁	C ₂₂	C ₂₃	C ₂₄	C ₂₅	C ₂₆	C ₂₇
C ₃₁	C ₃₂	C ₃₃	C ₃₄	C ₃₅	C ₃₆	C ₃₇
C ₄₁	C ₄₂	C ₄₃	C ₄₄	C ₄₅	C ₄₆	C ₄₇
C ₅₁	C ₅₂	C ₅₃	C ₅₄	C ₅₅	C ₅₆	C ₅₇
C ₆₁	C ₆₂	C ₆₃	C ₆₄	C ₆₅	C ₆₆	C ₆₇
C ₇₁	C ₇₂	C ₇₃	C ₇₄	C ₇₅	C ₇₆	C ₇₇

(b) Sous-bandes des coefficients DCT selon les trois directions privilégiées

FIGURE 2.11: Disposition des coefficients DCT d'un bloc de taille 7×7 en sous-bandes fréquentielles (a) et sous-bandes orientées (b)

des coefficients DCT Δ_n . La variance σ_n^2 associée est alors considérée comme la moyenne de l'énergie de la sous-bande n .

Le ratio d'énergie R_n des sous-bandes est alors défini comme suit

$$R_n = \frac{|\sigma_n^2 - \frac{1}{n-1} \sum_{j < n} \sigma_j^2|}{|\sigma_n^2 + \frac{1}{n-1} \sum_{j < n} \sigma_j^2|}, \forall n \in [2, 3] \quad (2.19)$$

Une valeur de ratio importante correspond à une grande disparité des énergies fréquentielles entre la sous-bande locale n et les sous-bandes inférieures. Cette caractéristique mesure ainsi la distribution relative des énergies dans les sous-bandes de basses fréquences et de hautes fréquences, qui peuvent être affectées par les distorsions appliquées à l'image. En outre, le lien entre le contraste dans le domaine fréquentiel et le domaine spatial a été établi par Blake et Sekuler [81] lors de l'étude approfondie de la fonction de sensibilité au contraste dans laquelle les différents seuils de contraste visible pour différentes fréquences spatiales ont été mesurés.

Les valeurs de R_2 et R_3 sont calculées pour tous les blocs de l'image, et un moyennage est effectué sur le 10^è percentile le plus élevé afin d'obtenir le ratio d'énergie global des sous-bandes. Finalement, le ratio d'énergie des sous-bandes ρ est défini comme la moyenne arithmétique de R_2 et R_3 :

$$\rho = \frac{R_2 + R_3}{2} \quad (2.20)$$

La figure 2.12(d) présente les cartes de distorsion générées lors du calcul du ratio d'énergie des sous-bandes ρ pour les trois images affichées en 2.12(a).

2.5.2.4 La variance anisotropique ξ

À l'instar de la modification des signatures spectrales locales lors de la dégradation d'une image, les distorsions appliquées à des images modifient également les énergies locales des orientations d'une image de manière non naturelle.

Afin de capturer l'information directionnelle présente dans une image, le contenu des blocs DCT sera analysé selon trois orientations (0, 45 et 90 degrés), telles qu'illustrées dans la figure 2.11(b). Un modèle gaussien généralisé est ensuite estimé pour chacune de ces orientations, et le paramètre associé ζ est déterminé à partir de l'histogramme obtenu.

La variance du paramètre ζ est calculé le long des trois orientations retenues pour chacun des blocs de l'image. La moyenne du 10^e percentile le plus élevé est finalement calculé. Le résultat obtenu est une mesure de variance anisotropique de l'image notée ξ .

La figure 2.12(e) présente les cartes de distorsion lors du calcul de la variance anisotropique ξ des trois images affichées en 2.12(a).

2.5.3 Le modèle de prédiction

Dans la section 2.3, il a été mis en évidence que le modèle gaussien généralisé multivariable permettait une prédiction de la qualité des images sans différence statistiquement significative avec le modèle laplacien généralisé multivariable. Le modèle gaussien est retenu dans notre cas et nécessite alors de calculer la moyenne et la covariance des données afin d'avoir une estimation de ses paramètres.

Soit l'ensemble $X_i = \{x_1, x_2, \dots, x_m\}$ le vecteur contenant les caractéristiques calculées à plusieurs niveaux de résolution. Trois niveaux de décomposition ont été sélectionnés, afin d'assurer une cohérence avec les résultats obtenus dans la section 2.3.6. L'idée est alors de modéliser la distribution du couple (X_i, DMOS_i) où DMOS_i correspond à la mesure de qualité fournie par les observateurs humains. Le modèle probabiliste est ensuite utilisé pour prédire le score de qualité des images en maximisant la probabilité $p(\text{DMOS}_i/X_i)$, ce qui revient à maximiser la distribution jointe de X et de DMOS $p(X, \text{DMOS}) = p(\text{DMOS}/X)p(X)$.

2.5.4 Évaluation de la performance

Les attributs modélisés précédemment l'ont été selon le précepte des statistiques de scènes naturelles qui stipule que les fonctions de vision d'une être humain sont modélisables par des distributions statistiques. Il convient maintenant de vérifier que tel est le cas, en calculant les coefficients de corrélation entre les valeurs retournées par chacun des attributs et les scores DMOS obtenus pour les observateurs humains.

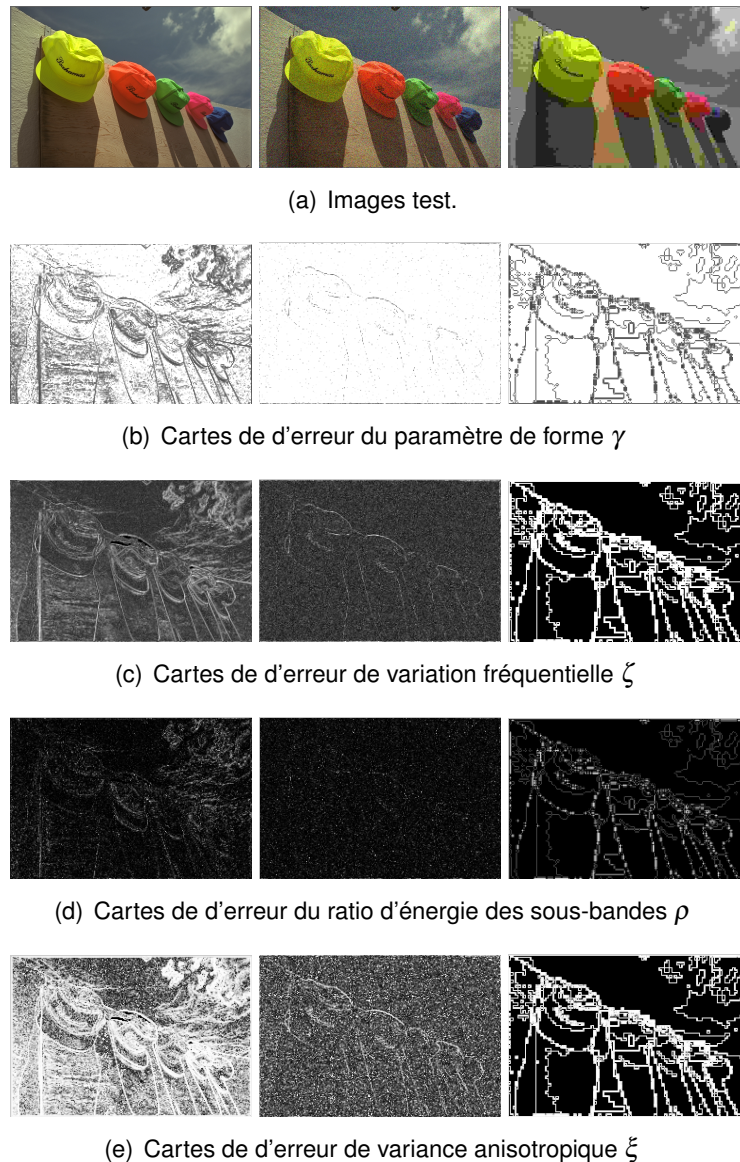


FIGURE 2.12: Exemple des cartes de dégradations obtenues sur trois images tests (a) dont une est exempte de dégradation (à gauche). L'image du centre a été dégradée par application d'un bruit blanc tandis que l'image de droite a subi une compression JPEG à 0.15312 bpp. Les sous-figures (b) correspondent à la carte d'erreur du coefficient γ tandis que les sous-figures (c) correspondent à la carte d'erreur du coefficient ζ . Les sous-figures (d) présentent les cartes de distorsion liées au coefficient ρ , alors que les sous-figures (e) présentent les cartes de distorsion liées au coefficient ξ .

2.5.4.1 Apparatus

À l'instar du protocole suivi dans la section 2.3.6, une base d'apprentissage a été construite à partir d'une partie des images de la base LIVE de manière à respecter la distribution des valeurs DMOS. Parmi les 29 images de référence et leurs versions dégradées disponibles dans la base LIVE, 15 images de référence (avec leurs versions dégradées) ont été choisies afin de définir la base d'apprentissage sur laquelle les paramètres des modèles ainsi que les valeurs des exposants ont été calculés. De plus, une technique de bootstrap à 999 répliques a été utilisée afin de renforcer la validité des résultats. La base TID2008 a également été utilisée comme base de test pour juger la performance de l'algorithme d'évaluation de la qualité agnostique.

Les scores prédits par la méthode développée G-BLIINDS2 sont comparés à ceux obtenus par la métrique sans référence BIQI [39] et deux métriques avec référence MS-SSIM [14] et le PSNR. Cette comparaison est réalisée à l'aide des coefficients de corrélation SROCC et LCC.

2.5.4.2 Corrélation des paramètres calculés avec le jugement humain

Le tableau 2.7 présente les valeurs de corrélation pour chacun des quatre paramètres pris indépendamment et les valeurs DMOS subjectives pour la base de test issue de la base d'image LIVE. Pour chaque paramètre, la moyenne de l'ensemble des valeurs calculées ainsi que la moyenne sur le 10^è percentile sont affichées.

Pour l'ensemble des dégradations présentes dans la base de test LIVE, les meilleurs taux de corrélation SROCC sont ceux obtenus avec la moyenne calculée sur le 10^è percentile des valeurs pour le paramètre de forme γ et le ratio d'énergie ρ . Concernant le paramètre de variation fréquentielle ζ et la dégradation « fast fading », la corrélation SROCC calculée sur l'ensemble des valeurs est meilleure que la corrélation SROCC calculée sur le 10^è percentile des valeurs. Cependant, la différence de corrélation n'est pas significative. En revanche, la variance anisotropique ξ présente une différence de corrélation SROCC significative lorsqu'elle est calculée sur l'ensemble des valeurs. Lorsque l'ensemble des dégradations est considérée, la majorité des meilleures corrélations est obtenue lorsque le calcul de la moyenne du 10^è percentile est appliqué.

Dès lors, ceci justifie le calcul des coefficients comme la moyenne du 10^è percentile de toutes les valeurs calculées.

2.5.4.3 Performance de la métrique

Le tableau 2.8 présente les valeurs de corrélation calculées entre les scores prédits par la métrique développée G-BLIINDS2, l'indice BIQI et les métriques avec référence MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test LIVE. Si l'on compare la

	paramètre de forme γ				paramètre variation fréquentielle ζ			
	Valeurs LCC		Valeurs SROCC		Valeurs LCC		Valeurs SROCC	
Dégradation	10è perc.	moyenne	10è perc.	moyenne	10è perc.	moyenne	10è perc.	moyenne
JPEG2000	0.804	0.745	0.921	0.733	0.944	0.923	0.933	0.913
JPEG	0.717	0.801	0.779	0.730	0.840	0.533	0.807	0.407
Bruit blanc	0.966	0.930	0.958	0.923	0.913	0.836	0.961	0.937
flou gaussien	0.807	0.318	0.901	0.330	0.918	0.849	0.925	0.861
fast fading	0.773	0.296	0.827	0.628	0.840	0.831	0.831	0.841
	ratio d'énergie ρ				variance anisotropique ξ			
	Valeurs LCC		Valeurs SROCC		Valeurs LCC		Valeurs SROCC	
Dégradation	10è perc.	moyenne	10è perc.	moyenne	10è perc.	moyenne	10è perc.	moyenne
JPEG2000	0.937	0.882	0.931	0.875	0.800	0.789	0.881	0.810
JPEG	0.802	0.308	0.949	0.460	0.156	0.779	0.210	0.823
Bruit blanc	0.948	0.875	0.975	0.961	0.923	0.845	0.921	0.917
flou gaussien	0.892	0.680	0.885	0.581	0.601	0.690	0.696	0.886
fast fading	0.851	0.735	0.860	0.756	0.701	0.725	0.755	0.826

TABLE 2.7: Coefficients de corrélation LCC et SROCC entre les scores prédits par chacun des paramètres développés et les valeurs DMOS subjectives pour la base de test issue de la base d'image LIVE.

	Corrélation SROCC				Corrélation LCC			
	NR-IQA		FR-IQA		NR-IQA		FR-IQA	
Dégradation	G-BLIINDS2	BIQI	MS-SSIM	PSNR	G-BLIINDS2	BIQI	MS-SSIM	PSNR
JPEG2000	0.951	0.799	0.950	0.866	0.963	0.809	0.940	0.864
JPEG	0.942	0.891	0.966	0.889	0.979	0.901	0.942	0.886
Bruit blanc	0.978	0.951	0.964	0.979	0.985	0.954	0.979	0.979
flou gaussien	0.944	0.846	0.932	0.789	0.948	0.829	0.891	0.782
fast fading	0.927	0.707	0.942	0.899	0.944	0.733	0.943	0.888
base entière	0.901	0.819	0.911	0.865	0.911	0.821	0.926	0.863

TABLE 2.8: Coefficients de corrélation calculés entre les scores prédits par les métriques G-BLIINDS2, BIQI, MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test LIVE.

qualité de prédiction des scores par l'approche G-BLIINDS2 et l'indice BIQI, on constate que la corrélation SROCC est meilleure pour G-BLIINDS2 et que cette différence est statistiquement significative. En outre, lorsque G-BLIINDS2 est comparé avec l'indice MS-SSIM, on constate que les valeurs de corrélation sont plus faibles. Ceci n'est pas surprenant étant donné que l'indice MS-SSIM est une métrique de qualité avec référence et qu'elle montre des taux de corrélation élevés avec le jugement humain. Néanmoins, les valeurs de corrélation obtenues avec l'approche G-BLIINDS2 sont très satisfaisants. Cette approche surclasse la prédiction fournie par le PSNR. Cette constatation est aussi valable pour le coefficient de corrélation LCC.

	Corrélation SROCC				Corrélation LCC			
	NR-IQA		FR-IQA		NR-IQA		FR-IQA	
Dégradation	G-BLIINDS2	BIQI	MS-SSIM	PSNR	G-BLIINDS2	BIQI	MS-SSIM	PSNR
Bruit gaussien	0.802	0.800	0.809	0.510	0.765	0.713	0.744	0.913
Bruit couleur	0.799	0.791	0.806	0.687	0.781	0.764	0.778	0.897
Bruit corrélé	0.801	0.791	0.822	0.523	0.716	0.738	0.759	0.938
Bruit masqué	0.786	0.765	0.810	0.621	0.789	0.755	0.785	0.789
Bruit haute freq.	0.865	0.852	0.869	0.640	0.823	0.832	0.818	0.917
Bruit impulsif	0.701	0.687	0.691	0.746	0.644	0.617	0.623	0.715
Bruit quantification	0.841	0.866	0.856	0.715	0.741	0.751	0.755	0.804
Flou gaussien	0.918	0.899	0.956	0.609	0.881	0.859	0.874	0.870
Débruitage	0.931	0.912	0.959	0.802	0.897	0.883	0.917	0.773
JPEG	0.943	0.918	0.933	0.661	0.936	0.919	0.929	0.833
JPEG2000	0.966	0.959	0.970	0.602	0.889	0.873	0.933	0.806
Transmission JPEG	0.861	0.871	0.868	0.809	0.808	0.832	0.815	0.656
Transmission JPEG2000	0.899	0.861	0.861	0.592	0.804	0.799	0.796	0.874
Bruit non centré	0.609	0.542	0.739	0.517	0.619	0.611	0.666	0.651
Distorsions de bloc	0.732	0.698	0.711	0.331	0.745	0.710	0.729	0.550
Décalage d'intensité	0.711	0.681	0.724	0.178	0.631	0.615	0.669	0.146
Chgt. contraste	0.667	0.618	0.635	0.453	0.781	0.699	0.772	0.425
Base entière	0.848	0.819	0.854	0.336	0.785	0.738	0.793	0.221

TABLE 2.9: Coefficients de corrélation calculés entre les scores prédits par les métriques G-BLIINDS2, BIQI, MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test TID2008.

Puisque le modèle a été réglé sur une base d'apprentissage extraite de la base d'images LIVE, et que cela peut influencer les résultats précédents, nous avons appliqué le modèle ainsi appris sur la base d'images TID2008 qui contient des dégradations non apprises. Cela va nous permettre de vérifier le pouvoir généralisateur de notre modèle. Le tableau 2.9 présente les valeurs de corrélation calculées entre les scores prédits par la métrique développée G-BLIINDS2, l'indice BIQI et les métriques avec référence MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test TID2008.

Pour les dégradations liées à un bruit de quantification et une erreur de transmission JPEG, l'indice BIQI permet une meilleure prédiction des scores de qualité. Néanmoins, pour les quinze dégradations restantes, l'approche G-BLIINDS2 fournit de meilleures prédictions. Lorsque la base complète TID est considérée, les scores de qualité sont plus corrélés avec le jugement humain pour l'approche développée G-BLIINDS2 qu'avec l'indice BIQI.

En comparant G-BLIINDS2 avec MS-SSIM, on constate que ce dernier est globalement plus corrélé (selon le coefficient SROCC) avec le jugement humain que G-BLIINDS2. Ceci n'est pas réellement surprenant étant donné que l'indice MS-SSIM est une métrique de qualité avec

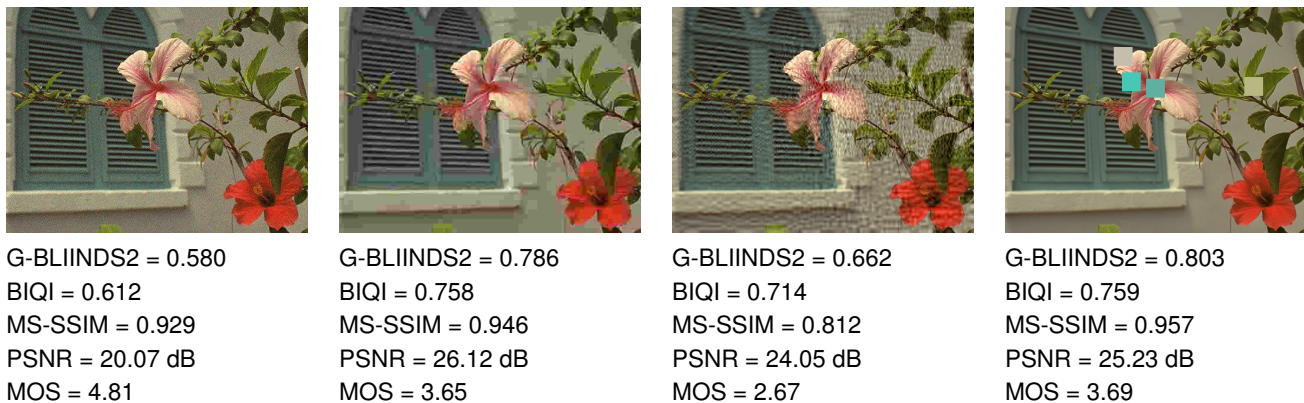


FIGURE 2.13: Scores obtenus avec la méthode de mesure de qualité sans référence G-BLIINDS et BIQI ainsi qu'avec l'indice MS-SSIM et le PSNR, toutes deux nécessitant une image de référence. Les images sont extraites de la base TID2008. De gauche à droite les dégradations subies sont un bruit haute-fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit dû à une insertion de blocs de couleur et d'intensité différente. Les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellente).

référence. La même interprétation est valable pour le coefficient LCC.

Pour cinq types de dégradation (Bruit impulsionnel, JPEG, Erreur de transmission JPEG2000, Distorsions de blocs et Changement de contraste), l'indice G-BLIINDS2 prédit les scores de qualité avec une fiabilité plus grande qu'en utilisant l'indice MS-SSIM, selon le coefficient SROCC.

La figure 2.13 présente les résultats obtenus lors de la mesure de la qualité des images dégradées en utilisant les indices G-BLIINDS2 et BIQI (métriques sans référence) et les métriques MS-SSIM et le PSNR (métriques avec référence). Les quatre images présentées sont extraites de la base TID2008 et ont été respectivement dégradées (de gauche à droite) par un bruit de haute fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit dû à une insertion de blocs de couleur et d'intensité différente. Les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellente). Si on compare les scores de la deuxième et de la quatrième image, on constate que les scores donnés par les métriques G-BLIINDS2, BIQI et MS-SSIM sont en adéquation avec les scores subjectifs (MOS), ce qui n'est pas le cas pour PSNR. Les quatre métriques échouent à prédire correctement le score de la première image, puisque les valeurs obtenues sont les plus basses (ce qui correspond à des images de mauvaise qualité) alors que la valeur du MOS est la plus élevée.

2.6 Conclusion

Dans ce chapitre, une métrique de qualité sans référence a été développée en se basant sur les statistiques de scènes naturelles. Deux approches ont été appliquées. La première consiste à modéliser certaines caractéristiques du SVH dans le domaine DCT. Ces caractéristiques formant un vecteur multidimensionnel, la modélisation de la distribution de ces facteurs selon une loi gaussienne multivariée a été réalisée, ce qui permet de calculer un indice de qualité pour une image donnée. La métrique ainsi développée montre de fort taux de corrélation avec le jugement humain. En outre, ceci a permis de valider l'utilisation des SSN dans le domaine transformé de la DCT. La métrique développée a été utilisée avec succès dans le domaine de la biométrie afin de mesurer la qualité de données biométriques morphologiques.

Dans une seconde approche, l'idée a été de dériver un modèle généralisé des SSN utilisant les coefficients DCT locaux, et de transformer ensuite les paramètres du modèle en caractéristiques permettant de prédire la qualité des images. A l'instar de ce qui a été réalisé précédemment, un modèle probabiliste généralisé est ensuite obtenu à partir de ces caractéristiques, ce qui permet de quantifier la qualité des images.

Estimation des dégradations d'une image avec référence

Dans ce chapitre, le thème de l'évaluation de la qualité avec image de référence est étudiée. Une approche utilisant un processus de classification multiclassés sous raisonnement incertain des images en fonction de classes de qualité est réalisée. Afin d'aboutir à une note finale de qualité, une régression par SVM est opérée. Le processus de classification sera également utilisé en imagerie de microscopie cellulaire. Dans une dernière partie, les caractéristiques usuellement combinées pour mesurer la qualité d'une image seront employées afin d'estimer le niveau de dégradation de l'image pour divers types de bruit et de flou conjointement utilisés dans le processus de distorsion de l'image.

Sommaire

3.1	Création et manipulation du vecteur de qualité : le choix des attributs	53
3.2	Le processus de classification du vecteur de qualité	55
3.3	De l'incertitude sur la classification	58
3.4	Combinaison des classificateurs sous raisonnement incertain	59
3.5	Schéma de régression	61
3.6	Mesure des performances du schéma général	62
3.7	Application en segmentation d'images cellulaires	67
3.8	Estimation de la combinaison de dégradation	73
3.9	Conclusion	85

LES schémas de compression avec perte d'information des images couleur induisent indubitablement des dégradations dont la visibilité est plus ou moins élevée selon leur importance et leur voisinage proche (effet de masquage, ou, son dual, effet de facilitation). L'évaluation de la qualité des images compressées est importante et peut être basée sur l'observation humaine. Les résultats ainsi obtenus, permettent de déterminer des seuils de sensibilité de l'observateur humain. L'un des inconvénients majeurs de ce type de mesure, est une mise en œuvre particulièrement longue ainsi que l'utilisation d'un protocole contraignant [5]. L'objectif est alors de disposer d'une mesure automatique de qualité, plus rapide, et, *a priori*, sans intervention directe de l'être humain dans le processus de prédiction du score de qualité d'une image.

Comme mentionné dans l'introduction de ce manuscrit, de nombreux travaux ont été réalisés au cours des deux dernières décennies pour développer une mesure de qualité fiable et robuste dans le cas où l'accès à l'image d'origine (non dégradée) est possible. Parmi les plus usités,

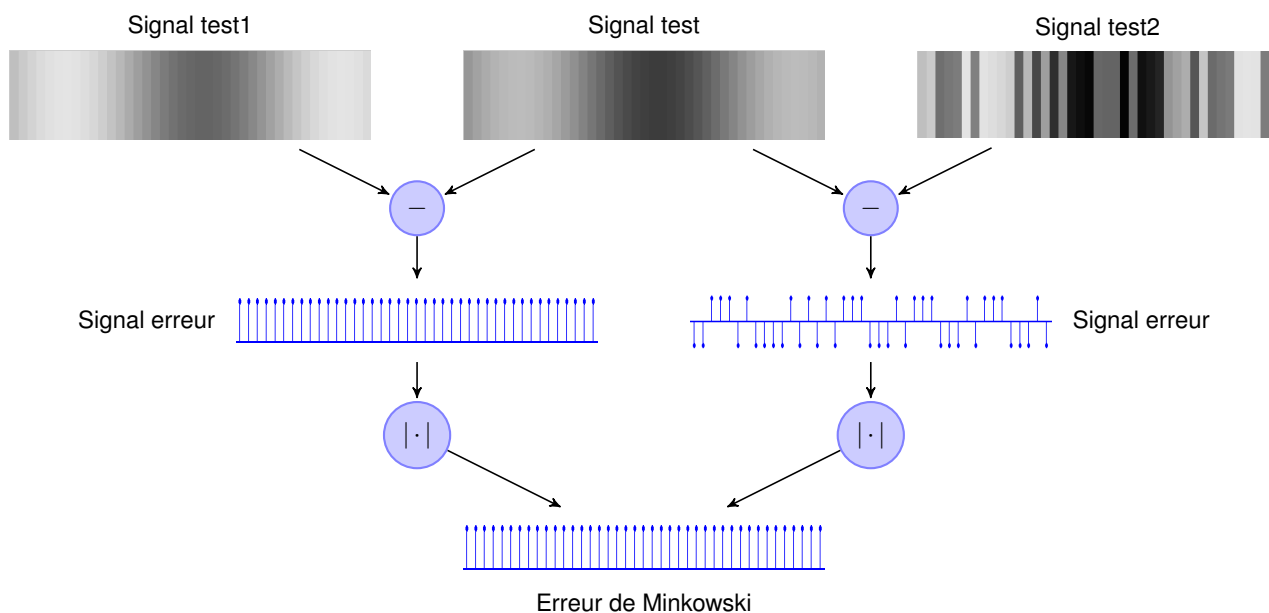


FIGURE 3.1: Illustration des limitations de l'utilisation de la métrique de Minkowski.

on retrouve la mesure du PSNR, de l'Erreur Quadratique Moyenne (EQM), etc. Dernièrement, à mesure de l'étoffement de la connaissance du fonctionnement du SVH, des modélisations de certaines caractéristiques ont été intégrées directement dans les métriques afin de « plagier » le plus fidèlement possible le modèle biologique. Toutes ces métriques délivrent au final une valeur scalaire permettant ainsi de procéder aisément à un ordonnancement des qualités. En revanche, l'un des inconvénients majeurs de toutes ces solutions algorithmiques repose sans conteste sur l'utilisation de la métrique de Minkowski pour obtenir le scalaire final, indépendamment de la norme utilisée (distance de Manhattan, distance euclidienne, etc.). En effet, des études antérieures [82] ont démontré que l'application de cette métrique n'est pas très performante pour le calcul de la note finale de la qualité. Ceci est dû au fait que l'hypothèse émise pour l'application de la métrique de Minkowski est l'indépendance des signaux, ce qui n'est pas vérifié dans le cas des images. La figure 3.1 illustre les limitations de la métrique de Minkowski. Dans cette illustration, le signal de test original est dégradé de façon à générer deux signaux test : les niveaux de gris du signal d'origine ont été modifiés par ajout d'une quantité égale à 50 pour générer le signal « Signal test1 », tandis que le signal « Signal test2 » a été modifié pour ajout ou retrait d'une quantité égale à 50 de manière aléatoire. Alors que la différence visuelle entre chacun des signaux dégradés et le signal d'origine est majeure pour le signal « Signal test2 », l'erreur associée est identique à l'erreur mesurée entre le « Signal test1 » et le signal d'origine.

Si l'on s'intéresse d'un peu plus près au processus interne engagé par un être humain lorsqu'on lui demande de juger la qualité d'une image (avec ou sans connaissance de l'image originale), force est de constater que ce dernier ne donne pas directement de note à l'image,

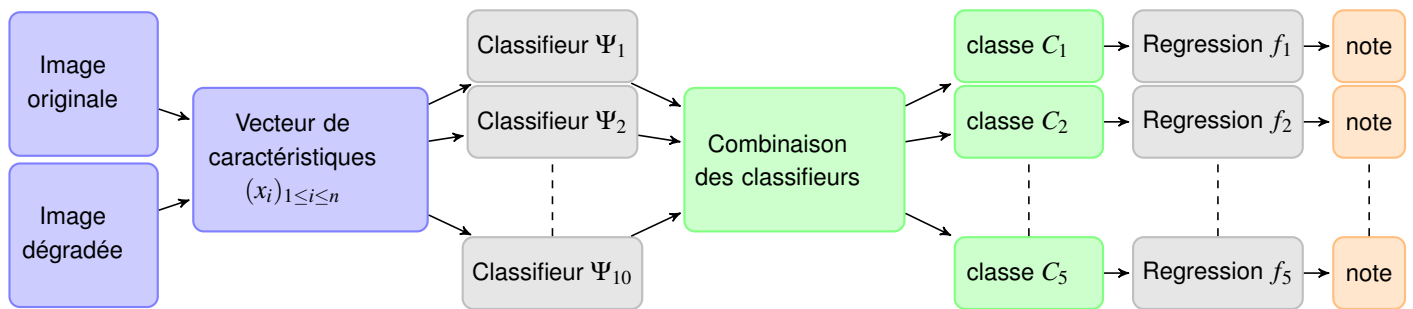


FIGURE 3.2: Synopsis de la méthode d'évaluation de la qualité basée sur une approche d'apprentissage statistique, nommée MLIQM—Machine Learning based Image Quality Metric.

mais opère plutôt par comparaison pour au final aboutir à une classification de la qualité de l'image. En effet, la décision finale est prise en fonction de plusieurs échelles internes [83, 84], ce qui permet de qualifier la démarche globale d'approche multi-critères qui ne semble pas opérer par un simple moyennage des défauts visualisés.

L'idée développée ici est d'utiliser un processus de classification multi-classe de la qualité des images auquel sera joint un processus de régression pour chacune des classes de manière à obtenir une prédiction du score final de qualité, tel que présenté dans la figure 3.2. Tout au long de ce chapitre, il sera fait référence à cette méthode sous l'acronyme MLIQM—Machine Learning based Image Quality Metric.

Quelques travaux d'estimation de la qualité basés sur le développement d'un expert informatique existent [85, 86, 37]. Toutes ces approches reposent sur l'utilisation de réseaux de neurones de type perceptron multicouche. Plus récemment, Narwaria [87] a proposé une méthode d'estimation de la qualité utilisant une régression à base de vecteurs support (SVR—Support Vector Regression). Les vecteurs de caractéristiques utilisés sont ceux obtenus après application d'une décomposition en valeurs singulières de la matrice image sur toute la dynamique de la qualité. Or, la sensibilité du SVH aux dégradations n'évolue pas de manière linéaire. Dès lors, dans la méthodologie développée en commun avec mes collègues Olivier Lezoray et Gilles Lebrun de l'UCBN, l'idée était d'utiliser les classes de qualité telles que définies par la recommandation ITU-R BT500.11 et de définir un schéma de classification multi-classe basé sur les Séparateurs à Vaste Marge (SVM) suivi d'un processus de régression à base de vecteur support pour chacune des classes.

3.1 Création et manipulation du vecteur de qualité : le choix des attributs

Afin de réaliser l'apprentissage de la phase de classification de la qualité des images [88], il est nécessaire de construire un vecteur d'attributs.

La catégorie des critères retenus s'inscrit dans le thème de la mesure de similarité entre une image et sa version dégradée. Dans le chapitre 2, nous avons mentionné l'hypothèse forte que le SVH a évolué au cours du temps pour s'adapter à son environnement notamment en développant une adaptation aux statistiques des scènes naturelles qui l'entoure.

Afin de prendre en compte cette hypothèse, les premiers critères sélectionnés sont ceux utilisés pour définir la métrique MS-SSIM [89] (Cf. section 4.2.1). On retrouve ainsi le critère de distorsion de luminance $l(I, J)$, de contraste $c(I, J)$ et un critère de comparaison de structure $s(I, J)$ entre une image originale I et sa version dégradée J . Contrairement à la définition de l'indice MS-SSIM original dans lequel le critère $l(I, J)$ était uniquement calculé sur le dernier niveau de résolution, chacun des trois critères de distorsion est calculé à travers les cinq échelles de résolution. Chacun des critères ainsi obtenu sera utilisé en tant que caractéristique à part entière dans le vecteur de qualité. Ainsi, à ce stade, le vecteur de qualité est constitué de 15 caractéristiques.

En outre, la sensibilité du SVH aux variations fréquentielles est établie depuis de nombreuses années et a permis de formaliser diverses propriétés fonctionnelles pertinentes du SVH, telles la sensibilité au contraste, les effets de masquage, etc. Plusieurs modèles estimant la visibilité des erreurs en intégrant ces divers formalismes ont été développés depuis lors. Tous ces modèles reposent sur une décomposition multi-canal du signal, chacun des canaux étant sélectif en orientation et en fréquences spatiales de manière à prendre en compte la sensibilité du SVH aux fréquences spatiales. Les diverses décompositions disponibles diffèrent principalement en nombre de canal (bande radiale), en orientation et en largeur de bande [90, 91, 92].

Notre choix s'est porté sur la transformation en pyramide orientée (*steerable pyramid transform*) [93] de manière à mesurer les effets de masquage. Une décomposition en trois niveaux et quatre orientations (0,45,90 et 135 degrés) suivie d'un filtrage passe-bas isotrope a été utilisée. Soit $a(x, y, f, \theta)$ le coefficient obtenu après application de la transformation précédemment décrite à la position (x, y) dans la bande fréquentielle f d'orientation θ . De manière à obtenir une mesure d'énergie locale, le coefficient $a(x, y, f, \theta)$ est normalisé comme suit :

$$r(x, y, f, \theta) = k \frac{(a(x, y, f, \theta))^2}{\sum_{\phi \in \{0, 45, 90, 135\}} (a(x, y, f, \phi))^2 + \sigma^2}, \quad (3.1)$$

Cette étape de normalisation est utile pour discriminer les différences de contraste le long de bande de faible dynamique. Ceci requiert l'utilisation de nombreuses bandes pour déterminer les variations de contraste. L'erreur de contraste entre l'image originale $r_0(x, y, f, \theta)$ et sa version dégradée $r_1(x, y, f, \theta)$ pour chaque bande fréquentielle f d'orientation θ est définie par :

$$\Delta r(f, \theta) = \left\| \sum_{x, y} r_0(x, y, f, \theta) - r_1(x, y, f, \theta) \right\|^2 \quad (3.2)$$

Ainsi, nous disposons de 12 valeurs supplémentaires qui viennent compléter le vecteur des caractéristiques.

Le vecteur final S contient alors 27 attributs $(s_k)_{1 \leq k \leq 27}$ de qualité, tant spatial que fréquentiel.

3.2 Le processus de classification du vecteur de qualité

La théorie de l'Apprentissage Statistique de Vapnik et de Chervonenkis [94] a conduit au développement d'une classe d'algorithmes connus sous le nom de SVM (Support Vector Machines). Ils permettent de réaliser des estimations en classification (et en régression). Une des originalités de la méthode est de produire une fonction de décision qui n'utilise qu'un sous-ensemble de la base d'apprentissage. Les éléments de ce sous-ensemble sont nommés Vecteurs de Support (**SV**). En outre, le choix de cette technique de classification a été motivé par ses fortes capacités en généralisation de données de grande dimension.

Soit une base d'apprentissage $A = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_k, y_k)\}$ composé de k couples (vecteur d'attributs, label) avec $x_i \in \mathcal{R}^n$ et $y_i \in \{-1, +1\}$. L'algorithme des SVM projette les vecteurs x_i dans un espace de travail \mathbf{H} à partir d'une fonction non linéaire $\phi : \mathcal{R}^n \rightarrow \mathbf{H}$. L'hyperplan optimal de séparation des deux classes dans l'espace \mathbf{H} est ensuite recherché. Cet hyperplan (\mathbf{w}, b) matérialise la frontière de séparation entre les deux classes. La classe y d'un nouvel exemple \mathbf{x} est définie par :

$$y = \text{sign}(\mathbf{w} \cdot \phi(\mathbf{x}) + b) \quad (3.3)$$

L'hyperplan est optimal s'il maximise la distance qui le sépare des exemples dont il est le plus proche. Cette distance est usuellement appelée marge du classificateur. Il a été démontré [94] que maximiser cette marge correspond à maximiser le pouvoir généralisateur du classificateur. En choisissant une fonction noyau $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$, il a également été montré que la résolution du problème dual correspondant produit une fonction de décision de la forme

$$f(x) = \sum_{x_i \in SV} \alpha_i y_i K(x_i, x) + b \quad (3.4)$$

où $\{\alpha_i\}$ et b correspondent à la solution optimale du problème dual.

3.2.1 La sélection du modèle

Le choix de la fonction noyau est critique dans la construction d'un expert informatique. La fonction RBF (*Radial Basis Function*) est très souvent utilisée comme fonction noyau ; la principale raison étant qu'elle fonctionne comme une mesure de similarité entre deux exemples à classifier. Puisque aucune connaissance *a priori* existe sur l'importance d'un attribut s par rapport aux autres, la fonction classique RBF a été modifiée de manière à modéliser les modulations d'importance des attributs (s_k) du vecteur d'attributs S telle que :

$$K_\beta(s_i, s_j) = \exp\left(-\sum_{k=1}^n \beta_k (s_i^k - s_j^k)^2 / r^2\right) \quad (3.5)$$

avec $s_{k,i}$ le $k^{\text{ième}}$ attribut de l'image i . β_k , qui est valeur booléenne, permet de modéliser la prise en compte ou non de la caractéristique k . Lorsque les valeurs de β_k ne sont pas fixées par des

a priori, elles sont déterminées en utilisant le paradigme de sélection d'attributs. La qualité d'un sous-ensemble d'attributs utilisé pour l'apprentissage des SVM binaires est mesurée par le taux de reconnaissance.

3.2.2 Les règles de combinaison

Les SVM étant des classificateurs binaires, ils ne peuvent pas être directement appliqués pour définir un schéma de classification multiclassés. Il convient d'effectuer une décomposition en plusieurs classificateurs binaires et de combiner les résultats obtenus par chacun des classificateurs binaires pour aboutir à la décision finale.

Dans notre cadre applicatif, deux schémas de combinaison ont été testés :

1. schéma de décomposition un-contre-un (UCU) (*OVO—One Versus One*)
2. schéma d'ordre de rang (OR) permettant de prendre en compte l'ordre naturel existant entre les classes de qualité.

Soit $t_{i,j}, \forall i \in [1, 5], j \in [2, 5]$ un problème binaire avec $t_{i,j} \in \{+1, -1\}$, où le chiffre 5 représente le nombre de classes finales de qualité tel que défini par la recommandation ITU-R BT.500.11. Soit $h_i(\cdot)$ et $f_i(\cdot)$ respectivement la fonction de décision obtenue après entraînement du problème binaire i et la sortie obtenue. Le tableau 3.1 présente les dix problèmes binaires utilisés dans notre étude, et le tableau 3.2 présente les fonctions de décision utilisées dans le second schéma de combinaison utilisé.

classe	$t_{5,4}$	$t_{5,3}$	$t_{5,2}$	$t_{5,1}$	$t_{4,3}$	$t_{4,2}$	$t_{4,1}$	$t_{3,2}$	$t_{3,1}$	$t_{2,1}$
5	+1	+1	+1	+1	-	-	-	-	-	-
4	-1	-	-	-	+1	+1	+1	-	-	-
3	-	-1	-	-	-1	-	-	+1	+1	-
2	-	-	-1	-	-	-1	-	-1	-	+1
1	-	-	-	-1	-	-	-1	-	-1	-1

TABLE 3.1: Transformation d'un problème de classification à cinq classes en 10 problèmes binaires obtenus lors de l'utilisation d'une approche UCU.

Contrairement au schéma de décomposition un-contre-un, l'information sur le rang des classes est conservée lors de l'application du schéma de décomposition OR, *i.e.* $\forall c_1, c_2 : t_{i,c_1} > t_{i,c_2} \rightarrow c_1 > c_2$.

La résolution d'un problème multi-classes est effectuée en le transformant en une combinaison de problèmes binaires. Dès lors, la décision finale est prise à partir de celle obtenue pour chacune des fonctions de décision. Plusieurs stratégies de combinaison peuvent être adoptées afin d'obtenir la décision finale. Le critère du vote majoritaire est celui que nous avons utilisé dans cette application, car il s'applique facilement à la décomposition de type RO.

classe	t_1	t_2	t_3	t_4
5	+1	+1	+1	+1
4	-1	+1	+1	+1
3	-1	-1	+1	+1
2	-1	-1	-1	+1
1	-1	-1	-1	-1

TABLE 3.2: Transformation d'un problème de classification à cinq classes en quatre problèmes binaires obtenus lors de l'utilisation du schéma de décomposition OR.

3.2.3 Résultats

Afin de mesurer l'efficacité du modèle développé, trois mesures de cohérence M_C ont été introduites et sont définies comme suit :

$$M_C = 1 - \frac{1}{m} \sum_{i=1}^m LO_C(D(i), MOS_i) \quad (3.6)$$

où $D(i)$ représente la classe prédite par le multi-classificateur, MOS_i correspond à la vérité terrain pour l'image i et $LO_C : C \in \{1, 2, 3\}$ est l'une des trois fonctions de perte utilisées :

$$LO_1(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ 1 & \text{sinon} \end{cases} \quad (3.7)$$

$$LO_2(y_1, y_2) = \begin{cases} 0 & \text{if } |y_1 - y_2| \leq 1 \\ 1 & \text{sinon} \end{cases} \quad (3.8)$$

$$LO_3(y_1, y_2) = \begin{cases} 0 & \text{if } y_1 = y_2 \\ \frac{m}{m_{y_2}} & \text{else} \end{cases} \quad (3.9)$$

où m_{y_2} fait référence au nombre d'images de la classe y_2 dans la base de référence. M_1 est une mesure classique du taux de reconnaissance. M_2 est une mesure de cohérence de rang pour la prédiction de la qualité. Elle permet uniquement la sélection des classes adjacentes à la classe de référence par la méthode développée. M_3 est une mesure prenant en compte la proportion relative à la représentation de chacune des cinq classes dans la base de données (ce qui correspond au taux de reconnaissance balancé). Le classificateur prendra en compte les classes peu représentées au même titre que les autres classes.

Les deux schémas de combinaison précédemment cités ont été testés sur la base d'images LIVE. Une sélection d'un ensemble d'apprentissage et d'un ensemble de test est ensuite opérée afin de pouvoir entraîner notre modèle. Chacun de ces deux ensembles est construit en respectant les distributions de qualité présentes initialement dans la base globale.

De plus, comme la base est relativement petite, la sélection du modèle est opérée selon une mesure de validation croisée de type *leave-one-out*. En outre, deux tests ont été menés :

1. l'ensemble des 27 attributs est conservé,
2. une opération de sélection des attributs est réalisée selon l'algorithme *best-first-search* [95].

	Les 27 attributs		Sélection d'attributs	
	UCU	OR	UCU	OR
M_1	0.862	0.823	0.701	0.653
M_2	0.986	1.000	0.938	1.000
M_3	0.810	0.781	0.709	0.598

TABLE 3.3: Mesures de cohérence calculées sur la base de test.

Ainsi, puisque nous avons deux modèles de décomposition (UCU et OR) et utilisé deux procédures liées aux attributs, nous obtenons au final quatre schémas de classification. Le tableau 3.3 présente les résultats obtenus pour chacun des quatre schémas testés. Le schéma de décomposition OR (qui intègre la cohérence des relations sémantiques inter-classes) permet d'obtenir un taux de classification de 100%, lorsque la mesure M_2 est utilisée. Ceci n'est pas surprenant puisque cette mesure ne considère pas la sélection d'une classe adjacente à la classe réelle comme une erreur. Ainsi, un exemple qui appartient à la classe de qualité « bonne qualité » sera considéré comme étant bien classé s'il est considéré comme un exemple de la classe directement supérieure (*i.e.*, « Très bonne qualité ») ou un exemple de la classe directement inférieure (*i.e.*, « mauvaise qualité »). Cela signifie en outre que le modèle OR tolère un certain degré d'incertitude quant à la classe d'appartenance de l'exemple, ce qui n'est pas le cas lorsque le schéma de décomposition UCU est utilisé.

Finalement, lorsqu'un processus de sélection d'attributs est appliqué, on remarque que le taux de reconnaissance est plus faible. Cela peut s'expliquer en partie par le fait que les attributs utilisés ont un rôle complémentaire dans la mesure de la qualité et que le simple fait de ne pas tenir compte de certaines de ces caractéristiques prive le système d'informations utiles pour prendre la décision finale.

3.3 De l'incertitude sur la classification

Dans les exemples précédents, nous avons appliqué le paradigme du vote majoritaire pour prendre la décision finale quant à la classe d'appartenance de l'exemple. Cependant, l'application d'un tel choix peut se révéler drastique. Si l'on compare les résultats obtenus avec le modèle de décomposition OR à ceux obtenus avec le modèle UCU, on constate que non seulement les taux de reconnaissance sont différents mais que cette différence peut être imputée au fait que la classification d'un exemple peut être incertain lorsqu'il se situe à la frontière de deux classes et que dans ce cas, il conviendrait de prendre la décision finale en intégrant cette incertitude.

Parmi tous les outils qui s'offrent à nous dans ce domaine, nous décidons d'utiliser la théorie de l'évidence (aussi appelée théorie de Dempster-Shafer ou théorie des fonctions de croyance) [96], qui permet d'une part de traiter des informations incertaines, et d'autre part de combiner des informations provenant de plusieurs sources. Ce cadre théorique permettra de traiter l'incertitude relative à chacun des classificateurs binaires qui peuvent être considérés comme des sources d'information imprécises et incertaines. Conceptuellement, la décision finale est prise en fonction de la confiance que l'on accorde au classificateur binaire. Cette confiance peut revêtir plusieurs aspects : un taux de reconnaissance, une probabilité de vraisemblance, une probabilité *a posteriori*, etc. Cependant, un SVM ne permet pas d'aboutir directement à de telles mesures de confiance. Notre choix s'est porté sur le calcul d'une probabilité *a posteriori* de classification. Afin d'affecter une probabilité *a posteriori* au classificateur (ce qui peut se traduire par une confiance que l'on accorde à la décision prise par un classificateur), une fonction paramétrique de type fonction logistique est utilisée [97] :

$$p(\mathbf{x}_i = 1|\Psi_i) = \frac{1}{1 + \exp(a\Psi_i + b)}, \quad (3.10)$$

Les paramètres a et b sont estimés en utilisant la procédure du maximum de vraisemblance. Ainsi, la confiance que l'on accorde à la classification d'un exemple \mathbf{x}_i par le classificateur Ψ_i est définie par la probabilité conditionnelle $p(\mathbf{x}_i = 1|\Psi_i)$.

Dans notre étude, le schéma de décomposition UCU est utilisé. Reste alors à combiner les résultats ainsi obtenus afin de prendre la décision finale d'appartenance à l'une des cinq classes de qualité, selon le formalisme de la théorie de l'évidence.

3.4 Combinaison des classificateurs sous raisonnement incertain

3.4.1 Éléments de théorie de l'information

Soit $\Omega = \{\omega_1, \dots, \omega_N\}$, l'ensemble des N classes finales possibles pour un pixel, appelé cadre de discernement. Dans notre cadre applicatif, $N = 5$ et Ω correspond aux cinq classes $\{\omega_E, \omega_B, \omega_{MO}, \omega_{MA}, \omega_{TM}\}$, respectivement pour les classes de qualité excellente, bonne, moyenne, mauvaise et très mauvaise. Au lieu de restreindre ces mesures à cet ensemble (comme le ferait la théorie des probabilités contrainte par son axiome d'additivité), la théorie de l'évidence s'étend sur l'ensemble puissance de Ω , noté 2^Ω , l'ensemble des 2^N sous-ensembles de Ω . On définit alors une fonction de masse m qui représente la croyance que l'on accorde aux différents états du système, à un instant donné. Cette fonction est également connue en tant que masse de croyance élémentaire. Elle est définie de 2^Ω dans $[0, 1]$ et vérifie $\sum_{A \subseteq \Omega} m(A) = 1$ et $m(\emptyset) = 0$ où $m(A)$ représente la croyance que l'on place dans la proposition A . Il est à noter que A représente soit un singleton ω_n , soit une disjonction d'hypothèses.

Ici, $m(A)$ quantifie la croyance que l'on accorde au fait que la classe cherchée appartienne au sous-ensemble A de Ω (et à aucun autre sous-ensemble de A). Les propositions A telles que $m(A) > 0$ sont appelées éléments focaux. De plus, dans le cas où l'ensemble des hypothèses est exhaustif et exclusif, la masse de l'ensemble vide \emptyset est nulle.

Deux fonctions de masse initiales m_1 et m_2 représentant les informations respectives de deux sources différentes, peuvent être combinées selon la règle de Dempster [98].

$$m(A) = \frac{\sum_{B \cap C = A} m_1(B)m_2(C)}{1 - K}, \quad \forall A \in \Omega, \quad A \neq \emptyset. \quad (3.11)$$

K est appelé le facteur de conflit et représente le désaccord entre les deux sources. Il correspond à la masse de l'ensemble vide $K = \sum_{B \cap C = \emptyset} m_1(B)m_2(C)$.

Après combinaison, il reste à prendre une décision quant à l'élément le plus « probable » de Ω . Plusieurs règles de décision sont alors possibles, mais une des plus utilisées reste celle du maximum de probabilité pignistique $\text{BetP}(\omega)$. Cette règle de décision présentée par SMETS [99] utilise la transformation pignistique, qui permet de répartir uniformément la masse associée à un sous-ensemble de Ω sur chacun de ses éléments :

$$\text{BetP}(\omega) = \sum_{\omega \in A \subseteq \Omega} \frac{m(A)}{|A|}, \quad \forall \omega \in \Omega \quad (3.12)$$

$|A|$ étant le cardinal de A . La décision va alors vers l'élément de Ω dont cette valeur est la plus grande :

$$\omega^* = \text{Arg} \{ \max_{\omega \in \Omega} [\text{BetP}(\omega)] \}. \quad (3.13)$$

3.4.2 Construction des fonctions de masse

L'une des difficultés réside dans le choix de la modélisation des fonctions de masse. Parmi les modélisations existantes, celle proposée par DENŒUX [100] a été utilisée car elle est basée sur la notion de distance aux voisins et permet d'intégrer divers critères de voisinage. Ainsi, la masse $m(\{\omega_n\})$ est définie comme une fonction décroissante de la distance d entre le vecteur \mathbf{x}_i à classer et le barycentre de la classe ω_n :

$$\begin{cases} m(\omega_n) = \alpha \exp(-\gamma_n d^2) \\ m(\Omega) = 1 - m(\omega_n) \end{cases} \quad (3.14)$$

où $0 < \alpha < 1$ correspond à la probabilité *a posteriori* associée au classificateur binaire Ψ_l ; $\gamma_n > 0$ dépend de la classe ω_n et est calculé par minimisation d'un critère d'erreur basé sur l'algorithme SEM (*Stochastic Expectation Maximization*) [101].

Au lieu de se restreindre à une seule classe, la théorie de l'évidence permet d'agréger dans un même élément focal plusieurs classes candidates selon un critère de similarité. Ainsi, un

élément à classifier est associé à un ensemble de classes, qui est au plus l'ensemble Ω . Dès lors, la contrainte d'affectation a été relâchée de manière à construire l'ensemble A tel que :

$$A = \{\omega_l \in \Omega / \max(p_l) - \delta_l \leq p_l \leq \max(p_l)\} \quad (3.15)$$

où $l \in \{1, \dots, 5\}$ et δ_l est une constante ad-hoc liée au classificateur utilisé.

Dans ce cas, les classes pour lesquelles la probabilité *a posteriori* associée appartient à l'intervalle $\max(p_l) - \delta_l \leq p_l \leq \max(p_l)$ sont considérées comme candidates pour le processus de classification durant l'étape de fusion.

3.4.3 Mesure des performances

La méthode de classification sous raisonnement incertain a été testée sur la base d'images LIVE [102]. Une sélection d'un ensemble d'apprentissage et d'un ensemble de test est ensuite opérée afin de pouvoir entraîner notre modèle. Chacun de ces deux ensembles est construit en respectant les distributions de qualité présentes initialement dans la base globale. Un processus de bootstrap à 999 répliquions sur la base d'apprentissage afin de stabiliser les résultats.

	JPEG2000	JPEG	Bruit blanc	Flou gaussien	Fast fading	Base entière
UCU-VM	0.921	0.951	0.968	0.944	0.905	0.909
UCU-DS	0.945	0.976	0.981	0.962	0.910	0.923

TABLE 3.4: Taux de reconnaissance pour la méthode UCU-VM (avec Vote Majoritaire) et pour la méthode UCU-DS (avec règle de combinaison de Dempster-Shafer).

Le tableau 3.4 présente les taux de reconnaissance pour chacune des métriques comparées. On constate que la méthode intégrant un raisonnement incertain permet d'obtenir des taux de classification supérieurs sur l'ensemble de la base LIVE à ceux obtenus en utilisant la classification par vote majoritaire. En outre, pour la dégradation de type fast fading, le gain obtenu n'est pas significatif. Ceci s'explique en partie par le fait que cette dégradation est le résultat de l'application successive de deux distorsions : JPEG2000 et modification de bits. Ce type de dégradation est particulier dans le sens où il implique deux types d'artefacts. Néanmoins, on observe tout de même une amélioration du taux de reconnaissance en appliquant un raisonnement incertain.

3.5 Schéma de régression

Même si l'estimation de la qualité par un observateur humain est plus en adéquation avec un processus de classification qu'avec un processus de notation scalaire, force est de constater

que la nécessité d'aboutir à une note de qualité est indiscutable. Ceci est inhérent à l'existence d'un ordre total dans l'espace \mathbb{R} permettant un classement des valeurs par ordre croissant (ou décroissant).

Dans ce cas, en considérant la base d'apprentissage A telle que définie dans la section 3.2, l'idée est d'estimer une fonction f qui présente au plus une déviation maximale ε traduisant la dépendance entre un exemple \mathbf{x}_i et la classe d'affectation y_i et qui soit la plus régulière possible. C'est ainsi que l'utilisation des SVM dans un contexte de régression est connue sous le terme « ε -SV régression ». Dans ce cas, les fonctions de coûts L_ε utilisées sont linéaires ou quadratiques et ne prennent en compte que les déviations $|y_i - f(\mathbf{x}_i)| > \varepsilon$. Cette fonction de coût est communément appelée fonction ε -insensible et est définie par

$$L_\varepsilon(x, y) = \begin{cases} 0 & \text{si } |y - f(\mathbf{x})| \leq \varepsilon \\ |y - f(\mathbf{x})| - \varepsilon & \text{sinon} \end{cases} \quad (3.16)$$

Les motivations relatives à l'utilisation d'un schéma ε -SV régression sont :

1. la solution en régression ne fait intervenir qu'un nombre limité d'exemples \mathbf{x}_i , ce qui permet de faire une sélection des exemples pertinents pour chacune des cinq classes de qualité,
2. la forme de la solution dépend du noyau utilisé. Dans ce cas, l'exploration de plusieurs espaces d'hypothèses est possible de manière à ne retenir que celui dans lequel les taux de corrélation avec les valeurs DMOS sont maximum.

3.6 Mesure des performances du schéma général

3.6.1 Apparatus

De manière à évaluer les performances du schéma développé, deux bases d'images sont utilisées : la base LIVE et la base TID2008.

Une base d'apprentissage (TrainC1) issue de la base LIVE est construite. Elle est composée de 12 images de référence et de toutes les versions dégradées associées. De plus, afin de procéder à l'apprentissage des fonctions de régression f_1, f_2, \dots, f_5 , cinq bases d'apprentissage (TrainR1, TrainR2, \dots , TrainR5) sont créées ; une pour chaque classe de qualité.

Les images non utilisées de la base LIVE constituent la base d'apprentissage TestC1. De même, la base TID2008 servira de base de test.

Les paramètres du schéma de classification multi-classe basé sur les SVM ainsi que des cinq schémas de régression sont déterminés par une technique de validation croisée à 10 parties. Qui plus est, une technique de bootstrap à 999 réplifications est appliquée pour valider les

résultats. Afin de comparer l'approche développée, quatre techniques d'évaluation de la qualité des images avec référence sont utilisées : il s'agit de l'indice MS-SSIM [13], de la mesure VSNR [16], de l'indice VIF [15] et de la mesure du PSNR. Toutes ces techniques d'évaluation ont la particularité commune d'être appliquées sur la composante achromatique des images. Les valeurs quantitatives des performances sont évaluées à l'aide de trois mesures de corrélation (Pearson–LCC, Kendall–KROCC, et de Spearman–SROCC) telles que définies dans le chapitre 1.

3.6.2 Résultats

Le tableau 3.5 présente les résultats obtenus lors du calcul des trois facteurs de corrélation entre les métriques d'évaluation de la qualité et les valeurs subjectives DMOS de la base de test TestC1. En utilisant l'approche développée, les résultats montrent que les corrélations sont plus élevées lorsque la base entière est considérée. Pour la plupart des sous-ensembles de la base de test (*i.e.*, par type de dégradation), la méthode développée permet une amélioration de la corrélation avec les scores des observateurs humains. Même si les améliorations induites ne sont pas toutes statistiquement significatives (au sens du test de Fisher), la consistance observée dans l'augmentation des valeurs de corrélation peut-être interprétée comme un indicateur de validité de l'approche développée. Une seconde interprétation des résultats concerne les attributs utilisés. Étant donné que la qualité des attributs revêt une importance somme toute particulière dans les performances des techniques de classification et/ou de régression par apprentissage statistique, les résultats obtenus tendent à démontrer (de manière empirique) que les 27 attributs utilisés dans cette étude sont pertinents. Néanmoins, on peut se demander si l'utilisation d'une base d'apprentissage extraite de la base LIVE pour régler les différents paramètres de nos SVM ne vient pas quelque peu biaiser les résultats.

Les tableaux 3.6 à 3.8 présentent les résultats du calcul des coefficients de corrélation pour les cinq méthodes d'évaluation de la qualité avec les valeurs subjectives DMOS en considérant la base de test TID2008.



Il est à noter qu'aucune nouvelle phase d'apprentissage n'a été effectuée, et que les paramètres obtenus suite à la phase d'entraînement du modèle sur les ensembles d'apprentissage sont conservés.

Les résultats illustrent le fait que la méthode proposée permet d'obtenir des taux de corrélation plus élevés qu'avec les quatre méthodes concurrentes, lorsque la base d'images complète est considérée. À l'exception des dégradations « Bruit Haute Freq. », « Bruit quantif. », « Trans. JPEG », « Distor. de bloc », « Décal d'intensité » et « Chgt. contraste », le schéma MLIQM est le mieux corrélé avec le jugement humain.

Si l'on considère la dégradation « Bruit Haute Freq. », la différence de corrélation n'est pas statistiquement significative. La dégradation « Bruit quantif. » (qui est une distorsion de

LCC						
	JPEG2K	JPEG	Bruit blanc	Flou gaussien	Fast Fading	Base entière
MLIQM	0.9861	0.9703	0.9652	0.9808	0.9505	0.9761
MS-SSIM	0.9091	0.8804	0.9108	0.8795	0.8499	0.9370
VIF	0.9722	0.9622	0.9741	0.9780	0.9636	0.9638
VSNR	0.8934	0.8327	0.8005	0.8056	0.6598	0.2164
PSNR	0.8730	0.8418	0.9792	0.7740	0.8752	0.7999
KROCC						
	JPEG2K	JPEG	Bruit blanc	Flou gaussien	Fast Fading	Base entière
MLIQM	0.8605	0.8491	0.9206	0.8919	0.8615	0.8601
MS-SSIM	0.8708	0.8165	0.8853	0.8570	0.8514	0.8183
VIF	0.8612	0.8283	0.9148	0.8889	0.8741	0.8560
VSNR	0.7142	0.7460	0.757	0.6242	0.5021	0.4702
PSNR	0.7037	0.6359	0.8941	0.5847	0.7067	0.6171
SROCC						
	JPEG2K	JPEG	Bruit blanc	Flou gaussien	Fast Fading	Base entière
MLIQM	0.9809	0.9746	0.9843	0.9891	0.9741	0.9807
MS-SSIM	0.9802	0.9628	0.9851	0.9733	0.9683	0.9513
VIF	0.9720	0.9552	0.9894	0.9821	0.9773	0.9718
VSNR	0.8306	0.8910	0.8181	0.7895	0.6808	0.6351
PSNR	0.8898	0.8409	0.9853	0.7816	0.8903	0.8197

TABLE 3.5: Coefficients de corrélation obtenus entre les valeurs prédites par les quatre méthodes d'évaluation de la qualité standard et par la méthode développée et les valeurs subjectives DMOS en considérant la base de test *TestC1*.

	Bruit gaussien	Bruit couleur	Bruit corrélé	Bruit masqué	Bruit haute freq.	Bruit impuls.
MLIQM	0.9012	0.8815	0.8761	0.8145	0.9367	0.9105
MS-SSIM	0.7437	0.7777	0.759	0.7849	0.8183	0.6227
VIF	0.8656	0.8928	0.8578	0.8900	0.9441	0.8146
VSNR	0.7448	0.7635	0.7501	0.7578	0.8828	0.6237
PSNR	0.9126	0.8968	0.9377	0.7892	0.9168	0.7148
	Bruit quantif.	Flou gaussien	Débruitage	JPEG	JPEG2000	Trans. JPEG
MLIQM	0.8107	0.9415	0.9021	0.9297	0.9561	0.9001
MS-SSIM	0.7546	0.8743	0.9171	0.9293	0.933	0.8147
VIF	0.7442	0.9388	0.8968	0.9327	0.9169	0.872
VSNR	0.811	0.9159	0.9193	0.906	0.9344	0.6755
PSNR	0.8044	0.8703	0.7731	0.8328	0.8062	0.6561
	Trans. JPEG2000	Bruit non écentré	Distor. de bloc	Décal. d'intensité	Chgt. contraste	Base entière
MLIQM	0.8041	0.7716	0.8981	0.6719	0.8653	0.8112
MS-SSIM	0.7964	0.6661	0.729	0.6686	0.7722	0.7927
VIF	0.8307	0.7366	0.834	0.5896	0.8816	0.7769
VSNR	0.7605	0.5654	0.2733	0.2415	0.4284	0.6642
PSNR	0.8737	0.6514	0.5496	0.1463	0.4253	0.2209

TABLE 3.6: Coefficients de corrélation linéaire (LCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.

quantification) peut être interprétée comme une perte de couleur pouvant induire d'importantes déformations structurelles sur l'image lors d'une forte quantification. Dans ce cas, les informations capturées par les attributs dédiés aux déformations structurelles sont importantes, et ceci est parfaitement mesuré dans le cas de l'indice de qualité MS-SSIM. Dans notre cas, ces informations se trouvent en compétition avec les 22 attributs restant ce qui a pour effet d'en atténuer le poids. Néanmoins, la différence de corrélation est faible.

Pour les dégradations orientées compression (sauf la dégradation « Trans. JPEG » qui concerne des erreurs de transmission JPEG), le schéma développé permet une meilleure corrélation des scores prédits avec le jugement humain. Cependant, une faible différence de corrélation est observée pour les dégradations « Trans. JPEG » et « Distor. de bloc ». Pour cette dernière, les erreurs aboutissent à l'apparition de blocs de taille et d'intensité différentes, ce qui peut être considéré comme des erreurs de transmission, puisque le bloc reconstruit n'est pas identique à celui d'origine.

Les dégradations « Décal d'intensité » et « Chgt. contraste » concernent des changements d'intensité et de contraste. Elles ne peuvent pas être considérées uniquement comme des processus de dégradation de la qualité, mais comme des changements de l'aspect naturel des images. Dans ce cas, il se peut que ces changements ne soient pas perçus comme des dégradations mais comme une amélioration des images [103].

	Bruit gaussien	Bruit couleur	Bruit corrélé	Bruit masqué	Bruit haute freq.	Bruit impuls.
MLIQM	0.7012	0.6905	0.7015	0.6642	0.6551	0.6483
MS-SSIM	0.6091	0.6014	0.6181	0.6119	0.6377	0.4862
VIF	0.6783	0.6868	0.6784	0.6892	0.6741	0.6278
VSNR	0.563	0.5852	0.5627	0.5508	0.6619	0.4506
PSNR	0.6997	0.6839	0.7403	0.6026	0.7675	0.5328
	Bruit quantif.	Flou gaussien	Débruitage	JPEG	JPEG2000	Trans. JPEG
MLIQM	0.6219	0.8365	0.8015	0.7442	0.8812	0.6915
MS-SSIM	0.6648	0.8161	0.8317	0.7490	0.856à	0.6702
VIF	0.6494	0.8186	0.7577	0.7183	0.8503	0.658
VSNR	0.6332	0.7713	0.7646	0.7344	0.8034	0.5991
PSNR	0.6219	0.6944	0.6098	0.675	0.5807	0.5933
	Trans. JPEG2000	Bruit non eccentricé	Distor. de bloc	Décal. d'intensité	Chgt. contraste	Base entière
MLIQM	0.6601	0.6229	0.6534	0.6012	0.5651	0.6704
MS-SSIM	0.6766	0.5274	0.5038	0.5355	0.4748	0.6579
VIF	0.658	0.601	0.622	0.3498	0.5788	0.5860
VSNR	0.599	0.3989	0.1376	0.1743	0.2783	0.5274
PSNR	0.7244	0.4765	0.5325	0.117	0.3187	0.2396

TABLE 3.7: Coefficients de corrélation de rang de Kendall (KROCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.

	Bruit gaussien	Bruit couleur	Bruit corrélé	Bruit masqué	Bruit haute freq.	Bruit impuls.
MLIQM	0.9061	0.8954	0.9205	0.8815	0.9061	0.8641
MS-SSIM	0.8088	0.8056	0.8222	0.8103	0.8694	0.6909
VIF	0.8797	0.8757	0.8698	0.8683	0.9075	0.8327
VSNR	0.7734	0.7798	0.7673	0.7348	0.8822	0.6438
PSNR	0.5104	0.6869	0.523	0.6209	0.6401	0.7459
	Bruit quantif.	Flou gaussien	Débruitage	JPEG	JPEG2000	Trans. JPEG
MLIQM	0.8312	0.9612	0.9855	0.9348	0.9715	0.8512
MS-SSIM	0.8563	0.9562	0.9588	0.9329	0.9697	0.8681
VIF	0.7970	0.9540	0.9161	0.9168	0.9709	0.8585
VSNR	0.8255	0.9323	0.929	0.9181	0.9517	0.8053
PSNR	0.715	0.6089	0.8017	0.6608	0.6015	0.8091
	Trans. JPEG2000	Bruit non eccentricé	Distor. de bloc	Décal. d'intensité	Chgt. contraste	Base entière
MLIQM	0.8764	0.8173	0.8039	0.7158	0.8051	0.8941
MS-SSIM	0.861	0.7388	0.7109	0.724	0.635	0.8544
VIF	0.8501	0.7619	0.8324	0.5096	0.8188	0.7491
VSNR	0.7914	0.5723	0.1952	0.2495	0.425	0.6978
PSNR	0.5922	0.5167	0.3311	0.1784	0.4534	0.3356

TABLE 3.8: Coefficients de corrélation de rang de Spearman (SROCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.

3.6.3 Discussion

L'approche basée sur l'utilisation d'un processus de classification suivi d'une étape de régression permettant d'aboutir à une note de qualité des images semble pertinente au regard des résultats obtenus pour deux bases d'images test. Ceci confirme en partie l'hypothèse émise en introduction selon laquelle le SVH procède par classification et combinaison des informations reçues pour aboutir à un jugement final. Cette stratégie de classification peut alors être modélisée par un classificateur de grande dimension puisque de nombreux détails peuvent venir influencer la prise de décision finale. En outre, l'incertitude liée aux processus de classification peut être prise en compte dans le cadre de la combinaison des informations sous raisonnement incertain.

L'aptitude des êtres humains à réaliser des opérations d'ordonnement peut être modélisée par une fonction de décision de type fonction de régression non-linéaire. La confiance que l'on peut porter à l'encontre des jugements humains peut être modélisée par la marge souple (*soft margin*) utilisée dans les SVM, tant pour l'étape de classification que de régression.

L'algorithme de classification multi-classe développé a été utilisé avec succès pour la segmentation d'images de microscopie cellulaire, comme décrit dans la section 3.7.

D'après les résultats précédemment obtenus, on observe que les dégradations sont globalement bien prises en compte dans les attributs développés, ainsi que dans les métriques existantes, tel l'indice MS-SSIM. On peut alors se demander si ces mêmes critères ne pourraient pas être utilisés à d'autres fins, *i.e.* pour

1. identifier le type de dégradation utilisé,
2. quantifier la quantité de distorsion appliquée.

Cette approche est développée dans la section 3.8.

3.7 Application en segmentation d'images cellulaires

L'analyse d'images microscopiques dans les laboratoires d'anatomie et de cytologie pathologiques tient une place très importante pour la détection des cellules anormales afin d'établir un diagnostic. Les cellules sont visuellement évaluées par un technicien lors d'une phase de screening. Cette analyse est d'un intérêt capital pour le pathologiste car son diagnostic dépend de la bonne reconnaissance des cellules anormales ou suspectes. Or le screening visuel est difficile et le faible nombre de cellules anormales comparativement au nombre de cellules normales implique une grande concentration du technicien. À cause de ce facteur subjectif des faux-négatifs peuvent apparaître. Pour pallier ceci, une approche consiste à aider le technicien par un système de screening semi-automatique afin de confirmer l'analyse visuelle. Dans cette optique, une méthode de segmentation des cellules d'images de microscopie a été conçue. Les cellules traitées sont constituées d'un cytoplasme et d'un noyau. Cependant, leur configuration

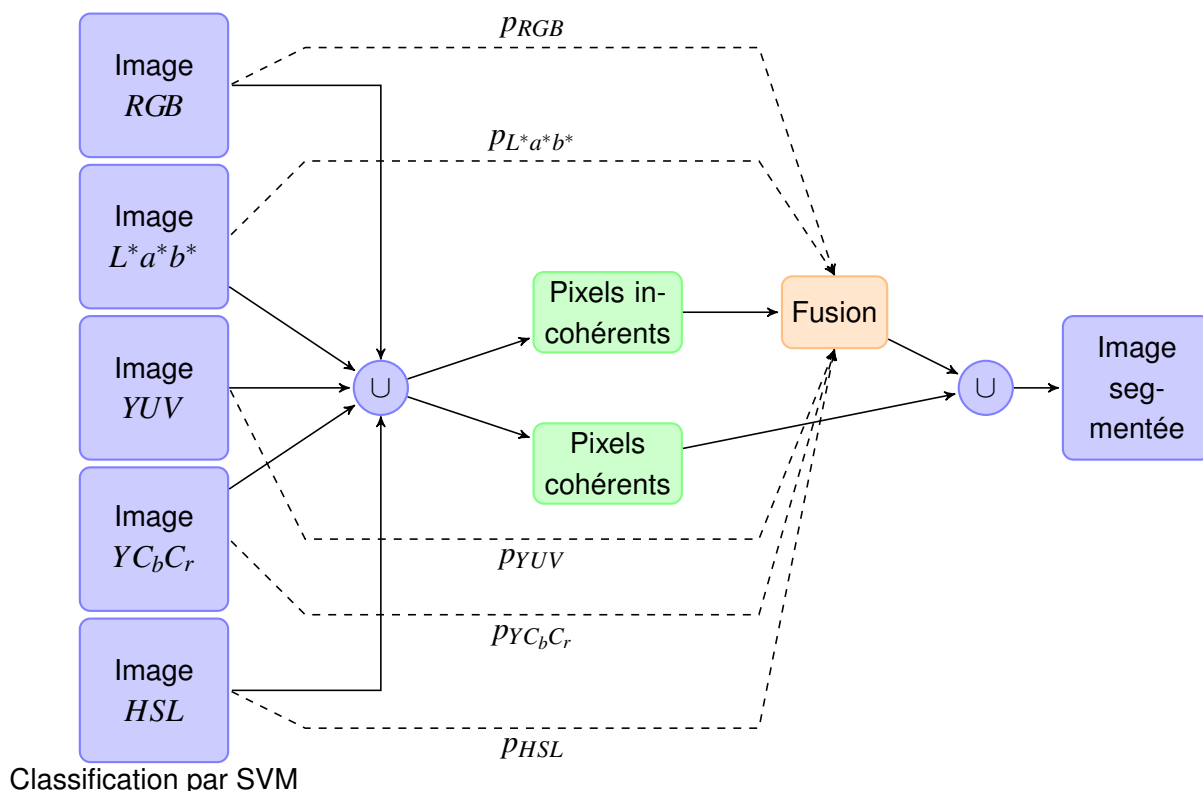


FIGURE 3.3: Synopsis du schéma proposé ou p_i représente la probabilité a posteriori calculée dans l'espace couleur i .

spatiale et colorimétrique est extrêmement variable, ce qui rend difficile leur extraction précise et une méthode de segmentation peut bien fonctionner sur certaines images et moins bien sur d'autres.

En général, un seul schéma de segmentation est appliqué afin d'obtenir des résultats de segmentation les plus fidèles possibles aux vérités terrains disponibles. Or, améliorer les taux de classification, et donc la qualité de la segmentation induit inéluctablement la définition et le développement de protocoles de classification de plus en plus sophistiqués, et par la même de plus en plus sensibles à la moindre variation de la structure de l'image.

Une autre voie peut cependant être envisagée : celle de la combinaison des résultats obtenus par différents schémas. Le classificateur multi-classes sous raisonnement incertain développé précédemment a été appliqué en collaboration avec mes collègues Olivier Lézoray et Gilles Lebrun de l'UCBN.

3.7.1 Création et fusion des cartes de segmentation

La figure 3.3 illustre le synopsis du schéma de segmentation proposé. Une première étape de segmentation par classification des pixels est réalisée dans les cinq espaces couleur retenus. A l'issue de cette phase, et pour chacune des segmentations, une probabilité *a posteriori* p_i est calculée, ce qui permet d'obtenir un indice de confiance de la classification pour chaque pixel de l'image par rapport aux classes finales. Puis, dès lors qu'il y a divergence de classification pour un pixel, une étape de fusion est opérée en prenant en compte la probabilité *a posteriori*. Une carte de segmentation finale est ainsi générée.

3.7.1.1 Segmentation par classification des pixels

La segmentation par classification des pixels a été obtenue à l'aide du schéma développé dans la section précédente. Dans ce cadre applicatif, trois fonctions de décision sont utilisées pour réaliser la classification en trois classes {noyau, fond, cytoplasme} correspondant à leurs discriminations en un contre tous.

Cinq espaces couleur ont été sélectionnés dans cette étude. Il s'agit des espaces *RVB*, $L^*a^*b^*$, *YUV*, YC_bC_r et *HSL* [128].

Un ensemble d'images segmentées constituant une vérité terrain des classes de pixels sert de base d'apprentissage. Chacune des trois fonctions de décision réalise un apprentissage par espace couleur, ce qui permet d'obtenir cinq cartes de segmentation distinctes avec pour chaque pixel de chaque carte une probabilité d'appartenance *a posteriori* p_i (Eq. 3.10) pour chaque classe. Cette probabilité correspond à un indice de confiance pour chaque classe et pour chaque espace couleur. L'objectif est ensuite de fusionner ces résultats afin d'aboutir à une classification finale, et donc une segmentation en trois classes. La théorie de l'évidence présentée précédemment (section 3.4) est utilisée pour sélectionner la classe à laquelle le pixel sera finalement affecté.

Afin de procéder à la fusion, une nouvelle carte de segmentation est générée par intersection des cartes obtenues dans les cinq espaces couleur. Seuls les pixels présentant des incohérences de classification seront traités par la théorie de l'évidence. Un « pixel incohérent » correspond à un pixel classé dans plus d'une classe lorsque l'union de toutes les cartes est réalisée.

3.7.1.2 Traitement des pixels incohérents

Les 5 fonctions de masses initiales (m_i) $i \in [1, \dots, 5]$ sont donc créées après calcul des k régions candidates à la fusion et avant décision du vote majoritaire. Ainsi, pour une même carte de segmentation, un pixel qui a été initialement affecté à une seule des trois classes

$\{\omega_F, \omega_C, \omega_N\}$ peut, en relâchant la contrainte d'affectation, être affecté à une classe parmi au plus les trois classes. La contrainte correspond simplement à la création d'un intervalle calculé à partir de la valeur maximale de la probabilité *a posteriori* p_i tel que $[p_i, p_i - \delta_i]$, avec δ_i fonction de l'espace couleur associé. Ainsi, toutes les classes dont les probabilités *a posteriori* sont incluses dans cet intervalle sont considérées comme des classes candidates à l'affectation du pixel.

Les fonctions de masses permettent de prendre en compte une incertitude associée à chacune des cartes de segmentation. Ainsi, on regroupe dans le même élément focal les classes qui se trouvent très proches les unes des autres, et la prise de décision ne s'effectue qu'après avoir combiné les résultats des autres projections. Dans cette étude, deux formulations de la distance d ont été étudiées. Il s'agit 1) de la distance euclidienne, 2) de la distance de Mahalanobis permettant de comparer la dispersion spatio-colorimétrique de deux nuages couleurs, définie par

$$d = \sqrt{(x - \mu_i)^T \Sigma_i^{-1} (x - \mu_i)} \quad (3.17)$$

où μ_i représente le vecteur moyenne et Σ_i^{-1} la matrice de covariance inverse associée à l'hypothèse ω_n .

Lors de l'utilisation de la distance euclidienne seule, la formulation de la distance euclidienne pondérée préconisée en 1994 par la Commission Internationale de l'Eclairage (CIE) lors de l'utilisation de l'espace couleur $L^*a^*b^*$ a été appliquée. Même si l'expression du calcul de la distance entre couleurs doit s'exprimer dans un espace couleur autre que l'espace RVB , ces formules ne sont valables que pour des petites distances et leurs utilisations en dehors de ce contexte n'est pas validée par la CIE.

3.7.2 Résultats expérimentaux

3.7.2.1 Évaluation de la qualité de la segmentation

Dans cette étude, le terme « vérité terrain » est utilisé en lieu et place d'image de référence dans laquelle tous les pixels ont été étiquetés en fonction de leur classe d'appartenance.

Lorsque l'on dispose d'une vérité terrain, les mesures de qualité sont toutes basées sur la prise en compte de la taille des régions et du taux de recouvrement entre les régions R_i de l'image segmentée I et les régions V_j de l'image de référence V . Parmi toutes les métriques disponibles, la mesure définie par MARTIN a été retenue puisqu'elle permet de mesurer la dissimilarité entre la segmentation résultat et la segmentation de référence [104].

Cette mesure est basée sur l'erreur $E(s)$ calculée en chaque pixel s par :

$$E(s) = \frac{\text{card}(V_j \setminus R_i)}{\text{card}(V_j)}. \quad (3.18)$$

ou sur l'erreur $E'(s)$

$$E'(s) = \frac{\text{card}(R_i \setminus V_j)}{\text{card}(R_i)}. \quad (3.19)$$

La dissimilarité est ensuite fournie par l'erreur locale de cohérence fournissant une mesure de qualité :

$$\text{ELC}(I, V) = \frac{1}{h \times l} \sum_s \min\{E(s), E'(s)\} \quad (3.20)$$

où h et l représentent respectivement la hauteur et la largeur en pixels de l'image V . Il existe une erreur globale de cohérence qui n'a pas été retenue ici car cette dernière a le défaut de favoriser une sur-segmentation ou une sous-segmentation [105].

3.7.2.2 Résultats

La technique proposée est appliquée sur une base d'images médicales composée d'images microscopiques cellulaires. La figure 3.4 présente les pixels incohérents obtenus après intersection des cartes de segmentation dans les cinq espaces couleur (figure 3.4(b)). Les désaccords interviennent principalement au niveau du cytoplasme et des noyaux lorsque l'on compare les résultats avec l'image initiale (figure 3.4(a)).

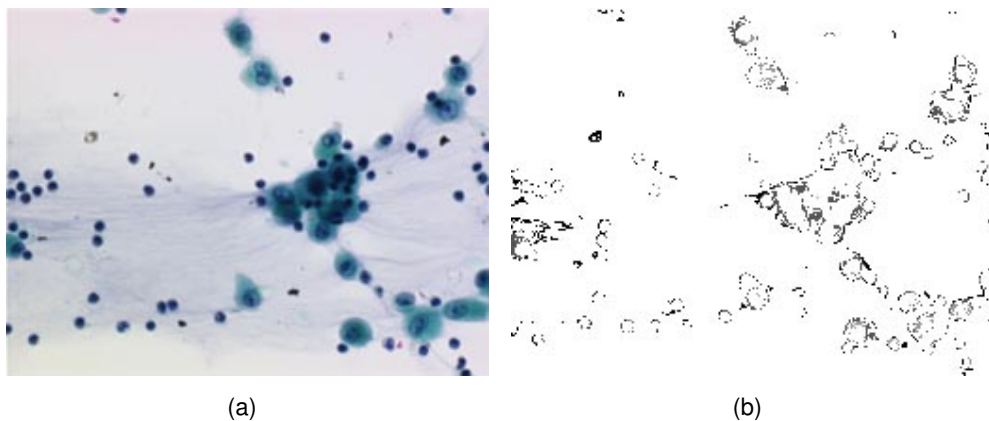


FIGURE 3.4: *Pixels incohérents obtenus par intersections des diverses cartes de segmentation en fonction d'une image originale.*

Le tableau 3.9 contient les taux de reconnaissance moyens par classe. Le taux de reconnaissance associé aux pixels du fond de l'image est supérieur à ceux associés aux deux autres classes que sont le Cytoplasme et le Noyau. Ceci n'est pas surprenant étant donné le nombre important d'exemple appartenant à la classe du Fond par rapport à ceux des classes du Cytoplasme et du Noyau. L'idée est alors de prendre en compte les divergences de classification dans les espaces couleur pour améliorer les taux de reconnaissance.

Fond	Cytoplasme	Noyau
95.63 ± 4.21	84.92 ± 3.81	89.62 ± 3.21

TABLE 3.9: Taux de reconnaissance moyen des pixels par classe.

Le tableau 3.10 présente les moyennes en pourcentage des pixels incorrectement et correctement classifiés ainsi que la moyenne en pourcentage des pixels présentant une incohérence de classification. On note qu'il subsiste un pourcentage de pixels (9.4%) pour lesquels il n'est plus possible de revenir sur la classification puisque ces derniers ont été associés à une classe erronée de manière unanime (selon la comparaison avec les images de vérité terrain). Ainsi, le taux de classification maximal que l'on puisse atteindre est fonction de ces affectations. Seuls les pixels pour lesquels subsiste un désaccord entre les classificateurs pourront subir une révision de classification, soit de 3.47% au maximum. Ainsi, pour améliorer le schéma de classification, il est également nécessaire d'améliorer les pré-traitements, à savoir, les résultats obtenus par les SVMs.

Moyenne des classifications par SVMs dans les différents espaces couleur		
correcte	incorrecte	incohérente
87.13 ± 6.12	9.4 ± 5.32	3.47 ± 1.13

TABLE 3.10: Pourcentage des pixels classifiés correctement, incorrectement selon les vérités terrain.

Le tableau 3.11 présente la moyenne des taux de classification corrects obtenus sur la base pour les SVMs et la méthode proposée utilisant la distance euclidienne (d_1) et la mesure de distance de Mahalanobis (d_2). Est également présentée la mesure de qualité de la segmentation finale.

Méthode	moyenne TR	Moyenne qualité
SVM	87.13 ± 6.12	0.45 ± 0.04
SVM combinés (d_1)	87.67 ± 5.62	0.39 ± 0.03
SVM combinés (d_2)	88.12 ± 5.43	0.38 ± 0.04

TABLE 3.11: Moyennes du taux de reconnaissance et de qualité selon le critère de Martin pour la segmentation par les SVMs, la combinaison des SVMs intégrant la distance euclidienne (d_1) et la distance de Mahalanobis (d_2).

Les résultats obtenus démontrent que la combinaison des probabilités des pixels incohérents permet d'améliorer sensiblement la qualité de la segmentation basée sur un processus de classification des pixels. On note effectivement une augmentation du pourcentage de reconnaissance pour chacune des trois classes. L'incohérence se situe principalement sur les pixels

appartenant au cytoplasme et aux noyaux. Le gain de qualité se portant essentiellement sur la classification de ces pixels, la qualité de la segmentation d'un point de vue cytologique est grandement améliorée. On note également que l'utilisation de mesure de similarité autre que la distance euclidienne permet d'obtenir de meilleurs taux de reconnaissance, ce qui se traduit par une meilleure qualité de segmentation finale.

3.8 Estimation de la combinaison de dégradation

Lors de mon séjour au département d'Informatique de l'Université de Sherbrooke (Canada) durant mon CRCT (*Congé pour Recherche et Conversion Thématique*) obtenu au titre de l'année 2008/2009, j'ai travaillé avec le Pr. Pierre-Marc Jodoin et l'étudiant en maîtrise André-Louis Caron sur la problématique d'estimation du type et du degré de dégradation dans une image en partant des attributs de qualité existants [106].

Les métriques développées dans la dernière décennie permettent de mesurer la fidélité ou la qualité des images ayant subies une modification. Malgré des valeurs de corrélation avec le jugement humain élevées, les critères mis en oeuvre dans ces méthodes algorithmes ne permettent pas directement d'estimer ni le type de dégradation utilisé ni la quantité de distorsion appliquée sur une image originale I . Qui plus est, si l'on considère un score de qualité, les méthodes ne permettent pas non plus de déterminer si une ou plusieurs dégradations ont été appliquées sur l'image. Ainsi, si l'on ne considère que le cas du bruit et du flou comme dégradation possible, aucune des métriques ne permet d'identifier la distorsion et la quantité utilisées. Bien qu'il existe des méthodes permettant d'estimer le niveau de dégradation appliquée dans le cas d'un seul type de distorsion [107, 108, 109, 110], les techniques sont beaucoup plus rares (voire inexistantes) en ce qui concerne l'estimation d'une combinaison de dégradations.

L'expertise développée dans le cadre de cette collaboration repose sur le développement d'une méthode d'estimation de la présence et de la quantité de plusieurs types de dégradation présents dans une image J , étant donné une image de référence I . Plus spécifiquement, nous cherchons à estimer la quantité de bruit α et de flou β présents dans une image.

3.8.1 L'approche retenue

Une extension de l'indice SSIM est utilisée pour estimer les paramètres de distorsion α et β . Cette extension est basée sur les attributs de distorsion de luminance l , de distorsion structurelle s et de contraste c entre une image I et sa version dégradée J (Cf section 4.2.1). Ces trois attributs permettent de définir les axes d'un espace 3D que nous appelons espace SSIM. Dans cet espace, chaque paire d'image (I, J) est associée à un point 3D (l, c, s) . Du fait que l'image J dérive de l'image I par application de deux paramètres (α, β) , la méthode développée se focalise sur la relation existante entre ces paramètres et le point 3D de l'espace

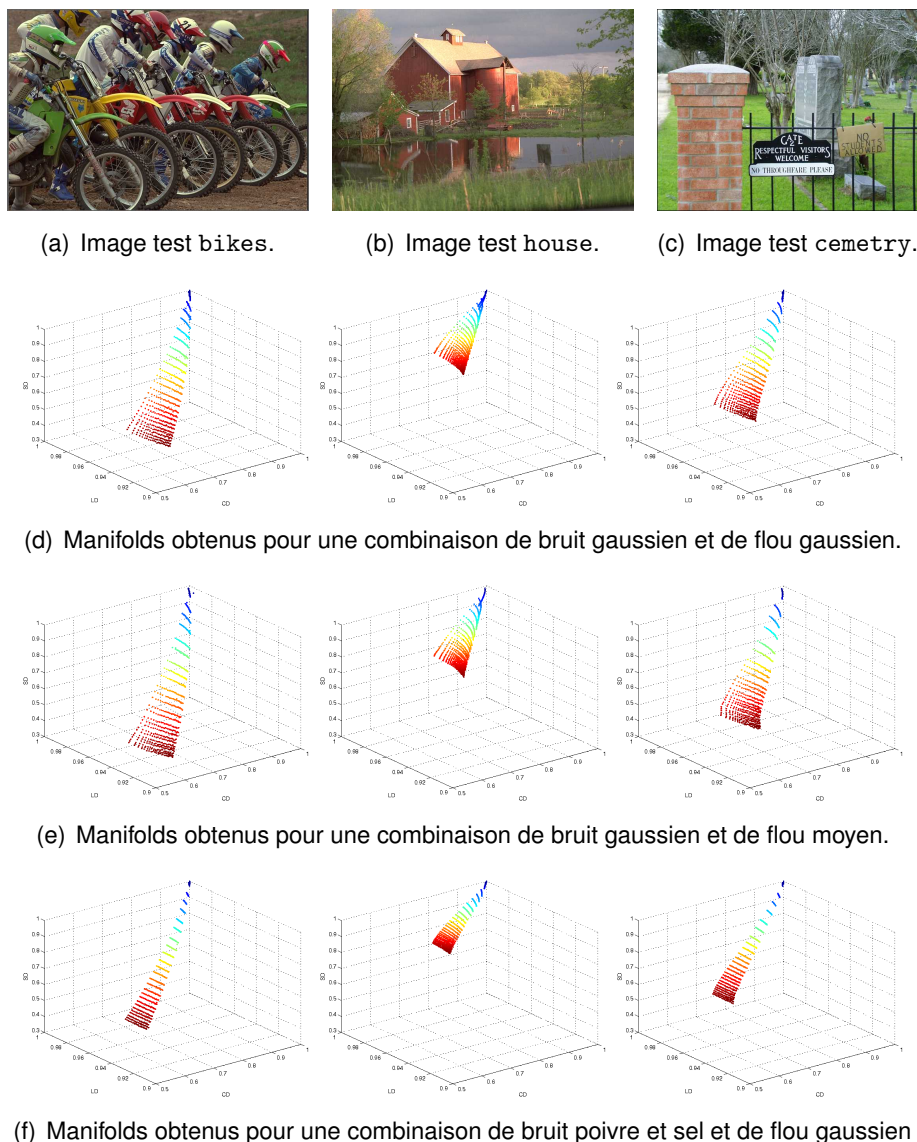


FIGURE 3.5: *Manifolds SSIM obtenus pour les trois images tests. Chaque point dans l'espace (l, c, s) correspond à une distorsion spécifique (α, β) . Ces manifolds correspondent à une combinaison de bruit gaussien et de flou gaussien (d), de bruit gaussien et de flou moyen (e) et de bruit poivre et sel et de flou gaussien (f).*

SSIM. En remarquant que la transformation entre le point 2D (α, β) et le point 3D (l, c, s) est bijective, le problème d'estimation multi-distorsions peut-être formalisé comme un problème classique d'optimisation.

La réduction de bruit et la rehaussement de la netteté des images sont deux opérations qui sont souvent critiques dans la production d'images claires. Néanmoins, l'utilisation des techniques de débruitage et de déconvolution nécessitent généralement de connaître la nature des

opérations appliquées et des valeurs des paramètres. En supposant que le processus de dégradation est linéaire et invariant en position, et que la scène 3D est constituée d'objets grossièrement situés à la même distance du capteur, l'image dégradée J est obtenue comme suit :

$$J = h_\beta * I + \mathcal{N}_\alpha \quad (3.21)$$

où h_β est un filtre passe-bas, \mathcal{N}_α le bruit et $*$ l'opération de convolution. Dans cette étude, l'estimation de divers types de bruit et de flou est réalisée par un paramètre. Concernant le flou, un filtre gaussien centré réduit (défini par la valeur de l'écart-type) et un filtre passe-bas uniforme (défini par la taille du filtre) sont considérés. De même, un bruit gaussien centré réduit (défini par la valeur de l'écart-type) et un bruit de type poivre et sel (défini par le pourcentage de pixels altérés) sont considérés. En procédant ainsi, quelle que soit la combinaison de bruit et de flou utilisée, deux paramètres sont finalement à estimer : α et β . À partir de (3.21), il est possible d'associer un point 2D (α, β) à un point 3D de l'espace SSIM tel que

$$(\alpha, \beta) \rightarrow (l(I, h_\beta * I + \mathcal{N}_\alpha), c(I, h_\beta * I + \mathcal{N}_\alpha), s(I, h_\beta * I + \mathcal{N}_\alpha)). \quad (3.22)$$

Afin d'illustrer cette relation, 676 points ont été arrangés en treillis régulier de taille 26×26 dans l'espace (α, β) . Ce treillis est ensuite projeté dans l'espace SSIM selon (3.22). Comme illustré dans la figure 3.5, quel que soit le contenu de l'image d'origine I , le treillis projeté forme un manifold étonnamment lisse. Ceci suggère fortement qu'un changement dans l'espace 2D (α, β) peut être prédit dans l'espace (l, c, d) . Plus spécifiquement, nous pouvons empiriquement observer que

- il existe une relation bijective entre un point 2D (α, β) et son correspondant dans l'espace 3D (l, c, s) ,
- les manifolds obtenus présentent deux directions privilégiées qui correspondent aux variations des valeurs de α et β .

3.8.2 Estimation de la distorsion vue comme un problème d'optimisation

L'objectif est d'estimer les quantités de bruit α^* et de flou β^* dans une image J . Considérant la transformation bijective entre l'espace (α, β) et l'espace SSIM, la procédure d'estimation des distorsions peut être reformulée comme un processus d'optimisation. La figure 3.6 illustre cette approche. En supposant que le point (α, β) correspond à un unique point 3D (l^*, c^*, s^*) dans l'espace SSIM (matérialisé par un point rouge), l'objectif est de trouver la paire $(\hat{\alpha}, \hat{\beta})$ tel que $(I, I * h_{\hat{\beta}} + \mathcal{N}_{\hat{\alpha}})$ corresponde au point 3D $(\hat{l}, \hat{c}, \hat{s})$ (matérialisé par un point bleu) situé au plus près du point (l^*, c^*, s^*) . Dès lors, le problème de l'estimé des valeurs (α^*, β^*) revient à un problème d'optimisation qui peut-être formalisé comme suit :

$$(\hat{\alpha}, \hat{\beta}) = \arg \min_{\alpha, \beta} E(I, J, \alpha, \beta) \quad (3.23)$$

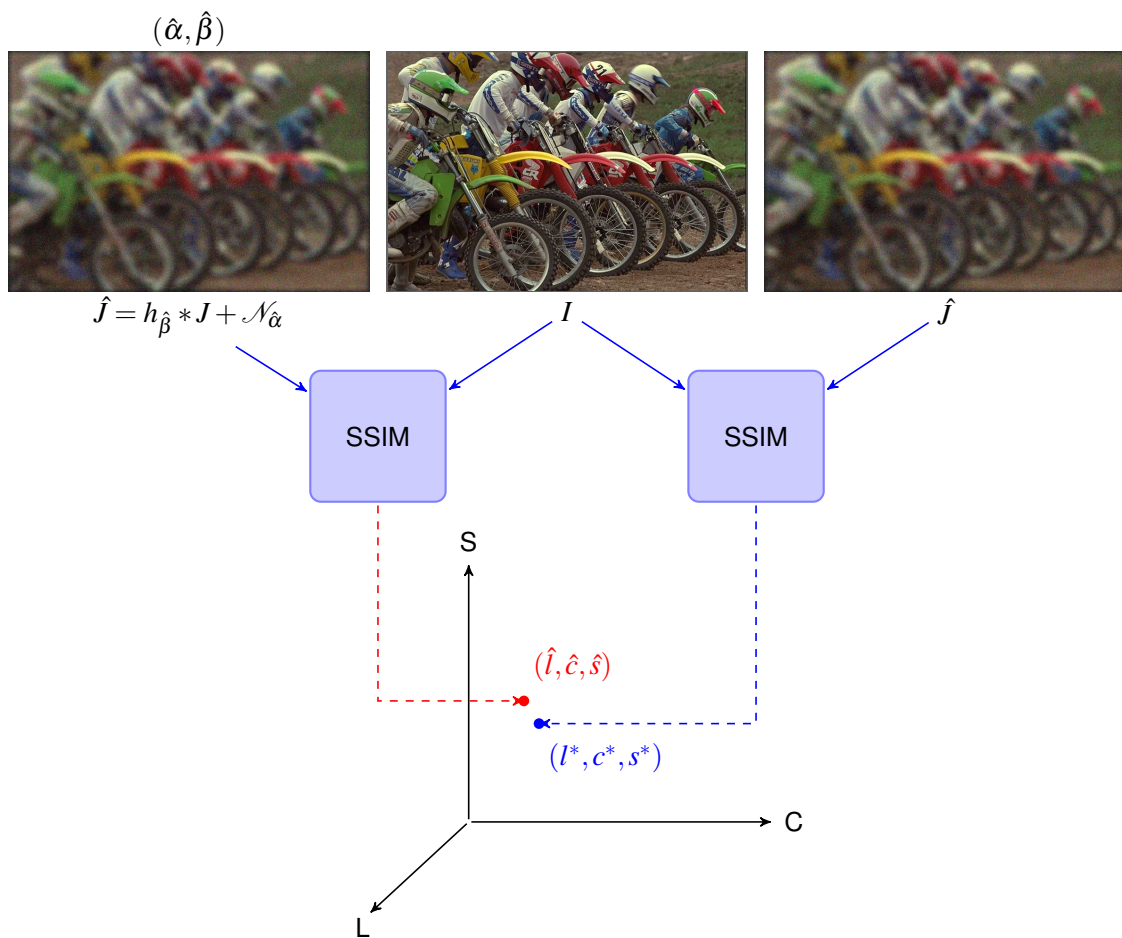


FIGURE 3.6: Synopsis de la procédure d'estimation des distorsions.

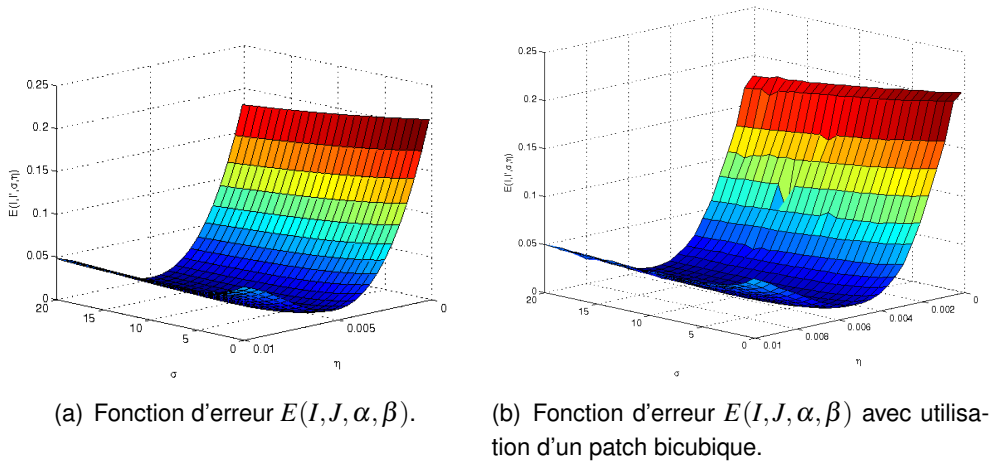


FIGURE 3.7: Fonction d'erreur $E(\cdot)$ pour l'image *bikes* avec et sans utilisation d'un patch bicubique. Le minimum est à $(\frac{\alpha_{MAX}}{2}, \frac{\beta_{MAX}}{2}) = (10, 0.005)$.

où $\alpha \in [0, \sigma_{MAX}]$, $\beta \in [0, \eta_{MAX}]$, et $E(\cdot)$ est la distance euclidienne classique $\sqrt{(l - l^*)^2 + (c - c^*)^2 + (s - s^*)^2}$. De manière à concrétiser la forme de cette fonction d'erreur, nous avons calculé sa valeur sur plus de 10.000 exemples (α, β) avec la solution $(\alpha^*, \beta^*) = (\frac{\alpha_{MAX}}{2}, \frac{\beta_{MAX}}{2})$. Les résultats sont présentés Figure 3.7(a). La fonction d'erreur est globalement lisse et possède un seul minimum global. Cependant, étant donné que la forme tridimensionnelle de la fonction d'erreur $E(I, J, \alpha, \beta)$ n'est pas connue, les optimisateurs de descente de gradient ne sont pas applicables. Plus précisément, un tel problème est un problème d'optimisation non-contraint sans dérivée [111].

Trois stratégies de recherche adaptées à notre problème ont été utilisées :

1. **la recherche par force brute–BF**. Dans ce cas, l'idée utilisée pour retrouver les valeurs $\hat{\alpha}, \hat{\beta}$ étant données les images I et J est de considérer toutes les valeurs possibles de (α, β) et de choisir celle dont le point 3D (l, c, s) correspond est le plus proche (au sens de la distance euclidienne) du point (l^*, c^*, s^*) .
2. **la recherche par simplex–SI** [112]. L'objectif est de réduire le temps de recherche qui peut se révéler prohibitif, en visitant simplement un nombre réduit de valeurs (α, β) avant d'atteindre le minimum global. Cette procédure commence la recherche avec trois solutions (appelées *sommets*) qui forment un triangle dans l'espace de recherche (α, β) . Les nouvelles positions sont ensuite définies de manière itérative à l'intérieur et à l'extérieur du triangle. Les erreurs sont ensuite calculées avec ces nouvelles positions et comparées aux erreurs obtenues avec les anciennes positions, *i.e.* avec les sommets du triangle. Une nouvelle position deviendra un sommet du triangle à la condition que la nouvelle valeur de l'erreur soit inférieure à l'ancienne. Une fois les sommets du nouveau triangle définis, le processus est itéré jusqu'à ce que le diamètre du cercle inscrit au triangle soit inférieur à un seuil fixé ou lorsque le nombre de points visités est supérieur à un seuil maximal

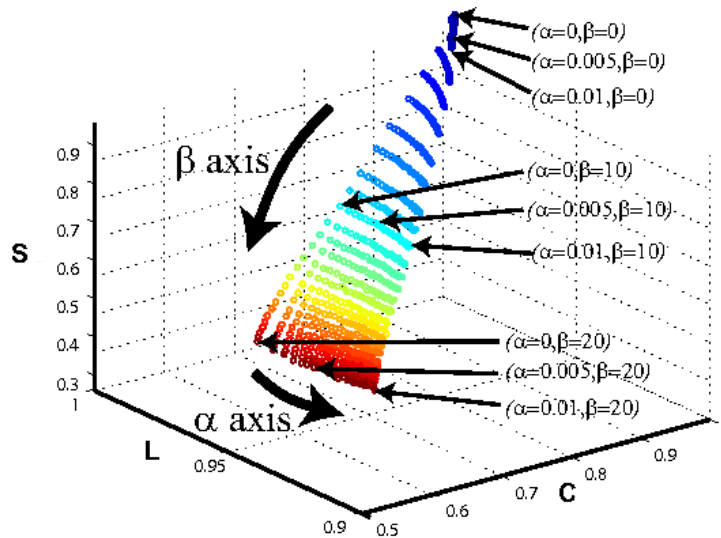


FIGURE 3.8: *Manifold SSIM associé à l'image cemetery. Ce manifold peut être vu comme un patch paramétrique sur lequel chaque point est défini par un couple (α, β)*

prédéfini.

3. **la recherche selon l'algorithme newUOA-NU [113]**. Bien que la méthode du simplexe soit une solution classique pour les problèmes d'optimisation non-contrainte sans dérivées, une technique récente permet d'obtenir une convergence en un temps plus faible avec une meilleure stabilité numérique et une robustesse améliorée lors du traitement de fonctions de plus grande complexité. Les algorithmes d'optimisation à régions de confiance (*Trust-region optimization algorithms*) sont des techniques d'optimisation différentiables destinées à minimiser une fonction réelle définie sur un espace euclidien (voire hilbertien) tout en effectuant un nombre réduit d'évaluation. Une approche classique [114] consiste à construire une approximation de la fonction objective dans une région de confiance et à itérer une série de minimisation sur ce modèle. À chaque itération, un nouveau point candidat est obtenu suite à la minimisation de l'approximation. Si la minimisation est correcte, le nouveau point est conservé et vient grossir les rangs de la région candidate. Dans le cas contraire, l'approximation est jugée inexacte et est mise à jour, entraînant une réduction de la taille de la région de confiance.

3.8.3 De l'espace 2D à l'espace 3D

À partir de la figure 3.5, il a été précédemment fait mention que l'espace (α, β) forme un manifold relativement lisse dans l'espace SSIM, avec la particularité d'avoir deux principales directions associées aux paramètres α et β . Dans la figure 3.8, on constate que la position d'un point 3D est étroitement liée aux valeurs (α, β) . En d'autres termes, un point 3D est déterminé

en fonction de l'amplitude des paramètres α et β , qui peuvent alors être perçus comme des paramètres de navigation dans l'espace 3D. Ceci signifie que si la paramétrisation de la relation bijective $(\alpha, \beta) \leftrightarrow (l, c, s)$ est correctement modélisée, une très bonne prédiction de la projection du point (α, β) dans l'espace SSIM pourra être faite. Une telle paramétrisation permettra alors de s'affranchir de l'utilisation de (3.21) et des équations de calcul des coefficients l , s et d issus de l'indice SSIM.

La construction de la relation bijective est inspirée par la forme des manifolds 3D. Comme il est possible de le constater sur les figures 3.5 et 3.8, les manifolds 3D ont une forme proche d'un patch paramétrique [115]. En outre, la relation bijective $(\alpha, \beta) \leftrightarrow (l, c, s)$ est également proche de la définition d'un patch paramétrique, qui est défini par

$$(x(s, t) \ y(s, t) \ z(s, t))^T = S.M.G.M^T.T^T \quad (3.24)$$

où $s \in [0, 1]$, $t \in [0, 1]$ et (x, y, z) les coordonnées du point 3D. Dans le cas des patches bicubiques, on a $S = (1 \ s \ s^2 \ s^3)$, $T = (1 \ t \ t^2 \ t^3)$, G une matrice carrée 4×4 contenant 16 points de contrôle, et M fait référence à la matrice 4×4 définissant la nature du patch (Bezier, Hermite, Catmull-Rom, etc.). Puisque les manifolds obtenus dans l'espace 3D sont lisses, il est possible de plaquer un patch dessus en utilisant la matrice de Vandermonde et 16 points de contrôle obtenus après un échantillonnage uniforme de l'espace (α, β) avec un treillis de taille 4×4 . Dès lors que la matrice G est calculée, chaque paire (α, β) peut être mise en relation avec un point de l'espace SSIM selon la procédure suivante :

$$\begin{aligned} (s, t) &= (\alpha/\alpha_{\text{MAX}}, \beta/\beta_{\text{MAX}}) \\ (ld, cd, sd) &= S.M.G.M^T.T^T \end{aligned} \quad (3.25)$$

où M est la matrice de Vandermonde définie par

$$M = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1/3 & 1/9 & 1/27 \\ 1 & 2/3 & 4/9 & 8/27 \\ 1 & 1 & 1 & 1 \end{pmatrix}^{-1}$$

Comme nous pouvons l'observer sur la figure 3.7 (b), la fonction d'erreur $E(I, J, \alpha, \beta)$ obtenue par un patch bicubique a une forme proche de la fonction d'erreur originale.

3.8.3.1 Optimisation basée sur l'utilisation de patch

L'utilisation d'un patch permet de mettre en correspondance un point 2D (α, β) à un point 3D SSIM (l, c, s) avec un faible temps de calcul (se résumant à quatre multiplications de matrices). Étant donné qu'une procédure basée sur l'utilisation d'un patch ne modifie pas l'équation (3.23), les procédures d'optimisation préalablement présentées peuvent prendre en compte cette correspondance sans changer leur fonctionnalité. Seule la procédure d'association bijective $(\alpha, \beta) \leftrightarrow (l, c, s)$ est modifiée. Nous obtenons deux nouvelles méthodes :

1. recherche par force brute avec patch–PBF,
2. recherche par simplex avec patch–PSI.

3.8.3.2 Algorithme de Newton-Raphson–NR

L'utilisation d'un patch bicubique permet une reformulation du problème : « Soit $A = (l, c, s)$ un point 3D associé à un couple d'images (I, J) , trouver sa projection (s, t) sur le patch tel que $d(A, B(s, t))$ soit minimum, avec $B(s, t) = (x(s, t), y(s, t), z(s, t))$ ». L'objectif est alors de trouver les valeurs de paramètres s et t minimisant la distance $d(A, B(s, t))$. Cependant, lors de l'utilisation de patch bicubique, il n'existe pas de solution déterministe, ce qui requiert de trouver les racines d'un polynôme d'ordre 5. Selon PLASS et STONE [116], une solution consiste à initialiser les paramètres s et t à une valeur donnée et à rechercher la solution minimisant la distance $d(A, B(s, t))$ en utilisant une technique classique de point fixe. Dans notre cas, nous supposons que les meilleures valeurs des paramètres s et t correspondent à celles pour lesquelles les dérivées partielles sont nulles :

$$\frac{\partial}{\partial s} d(A, B(s, t)) = 0 \quad \text{et} \quad \frac{\partial}{\partial t} d(A, B(s, t)) = 0$$

Selon la formulation de Newton-Raphson, on obtient :

$$s^{[k+1]} = s^{[k]} - \frac{F_s}{F'_s} \quad \text{et} \quad t^{[k+1]} = t^{[k]} - \frac{F_t}{F'_t}$$

où $F_s = \frac{\partial}{\partial s} d(A, B(s, t))$ et $F_t = \frac{\partial}{\partial t} d(A, B(s, t))$.

Les voisins les plus proches du point A et des points de contrôle du patch sont sélectionnés. Le barycentre est ensuite calculé selon les nouvelles valeurs des paramètres s et t puisque les points 3D sont des points paramétriques. Le poids de chaque point de contrôle est ensuite calculé comme la distance euclidienne inverse entre le point A et les points de contrôle du patch. Lorsque la convergence est atteinte, les valeurs $\hat{\alpha}$ et $\hat{\beta}$ sont calculées en multipliant s et t par α_{MAX} et β_{MAX} .

3.8.3.3 Utilisation de patches raffinés

Bien que les patches bicubiques paramétriques approximent très bien les manifolds SSIM, on constate néanmoins que certaines approximations finales sont éloignées des valeurs réelles, ce qui induit une forte distorsion sur le résultat final. Une solution mise en œuvre pour réduire ces distorsions repose sur le principe des patches raffinés. Le concept est identique pour tous les optimisateurs utilisés. Lors de la convergence de l'une des techniques d'optimisation basées sur l'utilisation de patch vers une solution $(\hat{\alpha}, \hat{\beta})$, l'objectif est de plaquer un nouveau patch dans le voisinage de la solution et d'itérer le processus sur ce patch local jusqu'à convergence. Le

pseudo-code utilisé est présenté dans l'algorithme 1. À partir de ce pseudo-code, on remarque que l'espace de recherche défini par les paramètres (α_1, β_2) et (α_2, β_1) est réduit à chacune des itérations.

Algorithm 1: Patches raffinés

input : $I, J, \alpha_{\text{MAX}}$ et β_{MAX}
output: $\hat{\alpha}$ et $\hat{\beta}$

- 1 $\alpha_1 \leftarrow 0, \quad \beta_1 \leftarrow 0$
- 2 $\alpha_2 \leftarrow \alpha_{\text{MAX}}, \quad \beta_2 \leftarrow \beta_{\text{MAX}}$
- 3
- 4 **for** $i \leftarrow 0$ **to** iterMax **do**
- 5 $G \leftarrow$ Calculer les 16 points SSIM couvrant la région (α_1, β_1) à (α_2, β_2)
- 6 $(\hat{\alpha}, \hat{\beta}) \leftarrow \text{Optimizer}(I, J, G)$
- 7 $\hat{\alpha} \leftarrow \hat{\alpha}(\alpha_2 - \alpha_1) + \alpha_1$
- 8 $\hat{\beta} \leftarrow \hat{\beta}(\beta_2 - \beta_1) + \beta_1$
- 9 $\alpha_1 \leftarrow \hat{\alpha} - \frac{\alpha_{\text{MAX}}}{4^i}, \quad \alpha_2 \leftarrow \hat{\alpha} + \frac{\alpha_{\text{MAX}}}{4^i}$
- 10 $\beta_1 \leftarrow \hat{\beta} - \frac{\beta_{\text{MAX}}}{4^i}, \quad \beta_2 \leftarrow \hat{\beta} + \frac{\beta_{\text{MAX}}}{4^i}$

De par la nature des patches utilisés, les valeurs $(\hat{\alpha}, \hat{\beta})$ obtenues pour les processus d'optimisation est normalisée, *i.e.*, comprise entre 0 et 1. Ces valeurs nécessitent alors d'être correctement replacées dans l'espace (α, β) (procédure réalisée sur les lignes 7 et 8 de l'algorithme). Cette méthode a été combinée avec la méthode de recherche par simplex et est dénommée RPSI (*Raffinement du Patch par Simplex*)

3.8.4 Comparaison des méthodes

Afin de tester les méthodes développées, 58 images incluant les images de référence de la base LIVE et de la base COREL sont utilisées. Chacune de ces images a ensuite subi 225 dégradations (α, β) linéairement distribuées entre $(0, 0)$ et $(\alpha_{\text{MAX}}=25, \beta_{\text{MAX}}=20)$. Au final, 13.050 images ont été générées. Pour chacune de ces images, la quantité de dégradation estimée $(\hat{\alpha}, \hat{\beta})$ est comparée avec les données réelles (α, β) .

En outre, de manière à pouvoir évaluer la pertinence des résultats obtenus avec les sept techniques d'estimation développées, quatre méthodes issues de l'état de l'art sont utilisées : deux sont liées à l'estimation de la quantité de bruit avec ou sans référence et les deux autres sont dédiées à l'estimation de la quantité de flou avec ou sans référence.

L'estimation de bruit sans utilisation de l'image de référence est basée sur la méthode développée par Buckley [107] et est dénommée *EVAR*. La méthode soustrait à l'image dégradée J un estimé \hat{I} de l'image de référence I . L'image \hat{I} est estimée en suivant une stratégie

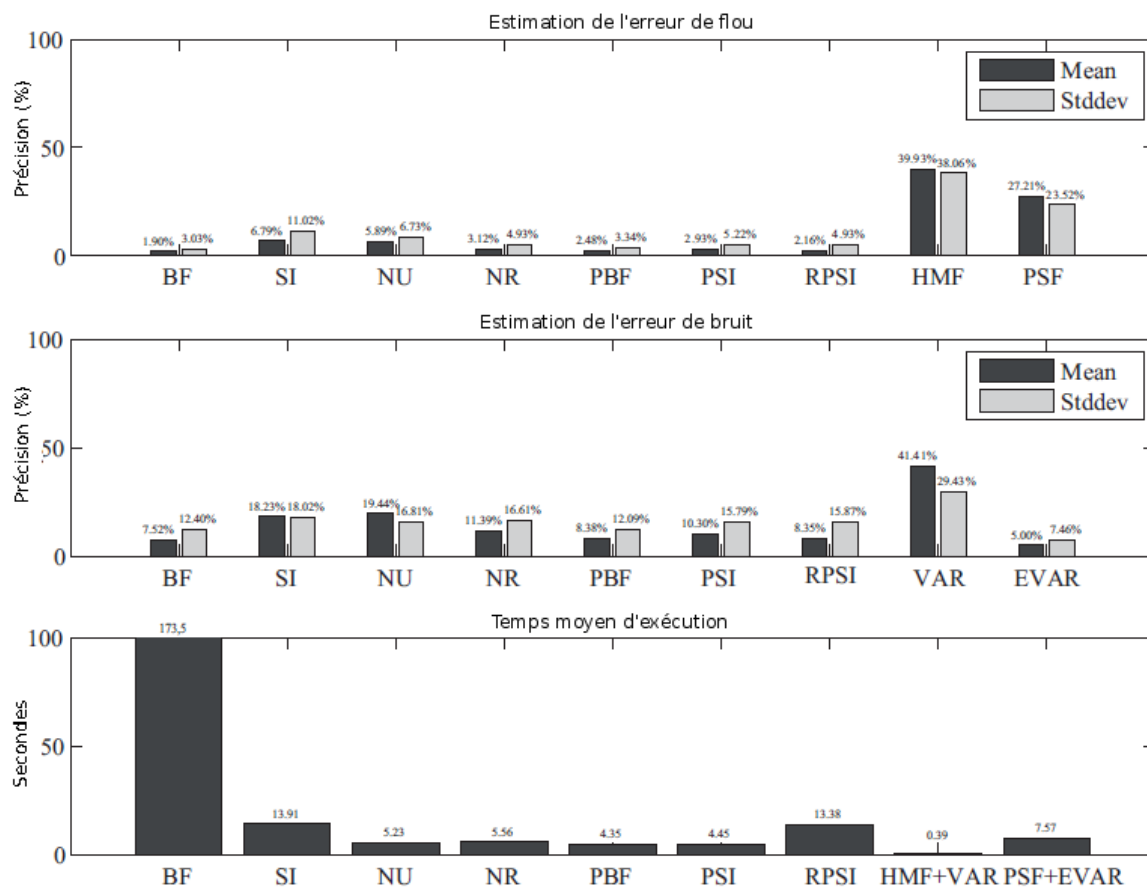


FIGURE 3.9: Comparaison de la performance et de la précision des algorithmes de recherche utilisés lors de l'estimation de bruit gaussien et de flou gaussien. Les graphiques supérieurs et intermédiaires présentent la moyenne et l'écart-type des erreurs d'estimation de la quantité de bruit et de flou. Les graphiques inférieurs présentent le temps moyen d'exécution des méthodes Matlab. Les méthodes testées sont les suivantes : **BF** – Estimation par force brute ; **SI** – Estimation par simplex ; **NR** – Estimation par l'algorithme de Newton-Raphson ; **NU** – Estimation par l'algorithme newUOA ; **PBF** – Estimation par force brute avec patch ; **PSI** – Estimation par simplex avec patch ; **RPSI** – Estimation selon le raffinement du patch avec simplex ; **HMF** – Estimation par utilisation d'un filtre homomorphe ; **PSF** – Estimation par utilisation de la fonction d'étalement du point ; **VAR** – Estimation par calcul de la variance du bruit ; **EVAR** – estimation par l'algorithme de Buckley.

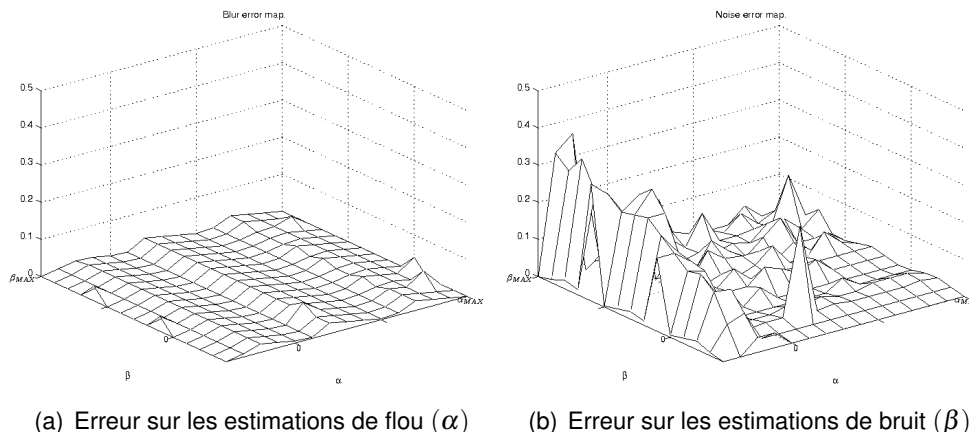


FIGURE 3.10: Distribution de l'erreur des estimations sur l'ensemble de l'espace de recherche pour l'image *bikes* en utilisant l'algorithme de recherche par force brute. L'estimation a été calculée sur un treillis régulier de taille 15×15 sur lequel la dynamique des valeurs est $(0, 0)$ à $(\alpha_{MAX}, \beta_{MAX})$.

basée sur l'idée qu'une image exempte de toute dégradation peut-être approchée par une spline. Ainsi l'image d'origine peut être reconstruite à partir de sa version dégradée J en approximant cette dernière par plusieurs patches. La variance du bruit est alors estimée comme $\text{var}(\hat{I}(x, y) - J(x, y))$. La méthode d'estimation de bruit en ayant accès à l'image de référence procède naïvement en calculant la variance du bruit $\text{var}(I(x, y) - J(x, y))$. Cette méthode est dénommée *VAR*.

Afin d'estimer la quantité de flou présente dans une image de manière aveugle, l'algorithme Lucy-Richardson [117, 118] est utilisé afin d'estimer le maximum de vraisemblance de la fonction d'étalement du point (*PSF-Point Spread Function*). Cet algorithme permet simultanément d'estimer la PSF et de restaurer l'image dégradée. Il sera référencé comme la méthode *PSF*. L'estimation avec référence de la quantité de flou est basée sur l'utilisation d'un filtre homomorphe, tel que défini en annexe A. La référence à l'utilisation de cette méthode se fera sous le terme *HMF*.

La figure 3.9 présente la moyenne des erreurs et l'écart-type obtenus pour chacune des onze méthodes testées, ainsi que le temps moyen d'exécution associé. Si l'on se focalise sur la précision des estimations, l'algorithme de recherche brute (BF) produit les meilleurs résultats avec une erreur d'estimation de flou de $\hat{\beta} = 0.95\%$ et de $\hat{\alpha} = 5.40\%$ pour le bruit. En outre, on constate que plusieurs stratégies de recherche permettent d'obtenir des valeurs $(\hat{\alpha}, \hat{\beta})$ proches de celles obtenues en utilisant la technique de recherche brute avec un temps d'exécution moindre. Comme montré dans le graphique inférieur, la recherche par simplexe (SI), la recherche par simplexe utilisant l'approche par patch (PSI) et la technique de recherche basée sur l'algorithme de Newton-Raphson (NR) sont plus rapides d'un facteur 10 que la recherche par force brute. En particulier, le schéma NR est environ 43 fois plus rapide que la méthode BF,

(a) image originale I 

(b) ($\alpha^* = 16, \beta^* = 7$) pour image de gauche ; ($\hat{\alpha} = 16.9, \hat{\beta} = 7.4$) pour image de droite ;
 (c) ($\alpha^* = 13, \beta^* = 10\%$) pour image de gauche ; ($\hat{\alpha} = 13.5, \hat{\beta} = 10.5\%$) pour image de droite

FIGURE 3.11: Exemple de résultats obtenus avec la méthode de recherche brute utilisant des patches. Soit une image originale I (a) et ses versions dégradées présentées à gauche pour (b) et (c) avec les paramètres ($\alpha^* = 16, \beta^* = 7$) pour une dégradation par bruit gaussien combinée à un flou gaussien et ($\alpha^* = 13, \beta^* = 10\%$) pour une dégradation par bruit poivre & sel combinée à un flou par filtre moyenneur, respectivement. Les images de droites (dans (b) et (c)) présentent les versions dégradées de l'image I en se basant sur les paramètres estimés ($\hat{\alpha}, \hat{\beta}$).

tout en ayant des pourcentages d'erreur d'estimation de $\hat{\beta} = 6.35\%$ et $\hat{\alpha} = 11.55\%$.

Bien que la méthode BF soit la plus précise, elle est néanmoins environ 18 fois plus lente que la seconde plus lente. Toutes les autres méthodes sont significativement plus rapides que la plus lente des techniques. Si l'on se focalise sur les résultats des méthodes utilisant les patches, l'algorithme NR produit les résultats les moins précis. Un résultat surprenant concerne la technique PSI qui se révèle plus précise que la méthode SI. Une explication possible est de mettre cela sur le compte des patches qui rendent la fonction d'erreur plus lisse et qui peut être, de fait, dépourvue de minima locaux (Cf. Figure 3.7). En outre, les erreurs d'estimation obtenues en utilisant la technique des patches raffinés (RSI) sont parmi les plus faibles malgré un coût calculatoire plus élevé. Ceci permet ainsi de justifier le fait que l'utilisation des patches permet de réduire non seulement les coûts calculatoires, mais aussi d'avoir de faibles erreurs de prédiction. Si l'on considère l'ensemble des méthodes testées, la technique PBF offre le meilleur compromis du triptyque vitesse, simplicité, précision.

Il est néanmoins difficile d'estimer le niveau de bruit avec une grande précision. Ceci peut être expliqué par le fait que le manifold devient étroit sur l'axe du bruit lorsque l'on s'approche de l'origine $(1, 1, 1)$ tel que présenté dans la figure 3.5. La forte proximité des valeurs dans cette région entravent même les performances de l'algorithme de force brute. Un exemple de la distribution des erreurs d'estimation sur l'ensemble de l'espace de recherche est présentée

dans la figure 3.10.

L'approche proposée a également été testée sur deux autres distorsions : un bruit poivre & sel combiné à un flou par filtrage moyen. Les résultats obtenus ont mis en évidence que les techniques permettent une estimation de la quantité de bruit avec la même précision que dans le cas du bruit gaussien, tandis qu'elles sont un peu moins précises pour l'estimation du flou. Le temps d'exécution reste quasiment identique. À l'instar des conclusions formulées dans le cas du bruit gaussien combiné à un flou gaussien, les méthodes les plus précises se révèlent être la technique BF et celles basées sur l'utilisation de patches.

La figure 3.11 présente un exemple de résultats obtenus dans le cas de l'application d'une combinaison de flou gaussien et de bruit gaussien (image de gauche de la figure 3.11(b)) et d'une combinaison de bruit poivre & sel et d'un flou par filtre moyenneur (image de gauche de la figure 3.11(c)) appliquées sur une image de référence (figure 3.11(a)). Après estimation des quantités de dégradations, l'image de référence est à nouveau dégradée avec les paramètres estimés. Les images résultantes sont affichées dans la partie droite des figures 3.11(b) et (c).

3.9 Conclusion

Dans ce chapitre, deux nouveaux concepts ont été abordés. Le premier concerne l'utilisation d'un schéma de classification de données afin de développer une méthode d'évaluation de la qualité des images, et le second détourne l'utilisation classique d'attributs de qualité afin de pouvoir estimer les niveaux et les types de dégradations utilisés.

L'utilisation d'un processus de classification et de régression basé sur les Séparateurs à Vaste Marge permet de plagier le fonctionnement du SVH qui utilise plusieurs échelles interne d'évaluation de qualité selon le principe du modèle perception-conscience défini par Freud. L'application des SVM dans un cadre multi-classe implique la définition d'une stratégie de combinaison des résultats réalisée sous raisonnement incertain. Ceci permet de prendre en compte les incertitudes des classificateurs impliqués. Cette stratégie a également été appliquée avec succès pour la segmentation des images de microscopie cellulaire, ce qui démontre une méthodologie suffisamment générique pour être appliquée dans différents domaines.

Une fois le processus de classification exécuté, une régression par SVM est effectuée afin de passer d'une représentation vectorielle de l'image à un score de qualité. Une fonction de régression est construite pour chacune des classes de qualité de façon à garantir une complexité faible de ladite fonction.

Dans une seconde approche, les attributs de qualités sont détournés de leur utilisation classique de manière à définir un processus d'estimation du niveau et du type de distorsion dans le cas de multi-dégradation. Les attributs retenus sont les trois facteurs définis dans l'indice SSIM de manière à construire un espace 3D dans lequel chacun des axes correspond à un des trois

facteurs. La méthode développée se focalise sur la relation bijective entre les paramètres de dégradation et le point 3D de l'espace SSIM.

Lorsque l'on cherche à quantifier la performance d'une méthode de mesure de qualité, il devient impératif de pouvoir comparer la méthode développée avec les méthodes existantes. Si l'on dispose d'une vérité terrain, on cherchera à reproduire le plus fidèlement possible cette vérité. La méthode qui aura ainsi le plus fort taux de corrélation avec cette vérité sera réputée comme étant la meilleure. Or la mesure de corrélation n'est peut être pas si fiable que cela et peut réserver quelques surprises comme nous le verrons dans le chapitre suivant.

Optimisation de la corrélation par échelles de différence

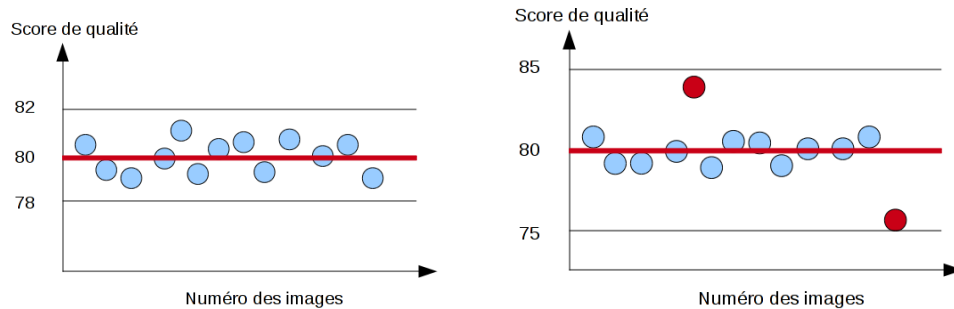
La qualité d'une solution algorithmique de mesure de qualité passe par l'optimisation de ses paramètres afin que les scores prédits soient le plus fidèle possible à une vérité terrain disponible, ce qui se traduit par des valeurs de coefficients de corrélation élevés. Une méthode d'optimisation basée sur l'utilisation conjointe d'une échelle perceptive des différences et d'un algorithme génétique a été développée afin de sélectionner les valeurs optimales des paramètres de la fonction de qualité développée.

Sommaire

4.1	Échelle de différence par maximum de vraisemblance	89
4.2	Procédure d'optimisation d'algorithmes d'évaluation de la qualité.	91
4.3	Conclusion	100

DÈS LORS que l'on cherche à estimer la performance d'un ou plusieurs systèmes d'évaluation de la qualité des images, les critères de corrélation sont couramment utilisés. Ainsi, plus le système aura une valeur de corrélation élevée avec les scores donnés par les observateurs humains, et meilleure sera jugée la métrique algorithmique.

Supposons que nous disposions de deux méthodes algorithmiques d'évaluation de la qualité q_1 et q_2 . Les figures 4.1(a) et 4.1(b) présentent un échantillon des scores obtenus pour chacune des deux métriques. Pour chacune des sous-figures, le score égal à 80 correspond à la valeur MOS obtenue pour chacune des images testées et les disques de couleurs correspondent aux valeurs prédites par chacune des métriques et pour chaque image testée. Supposons en outre que la valeur de corrélation SROCC ait été calculée pour chacun de ces échantillons et qu'elle soit identique pour les deux métriques q_1 et q_2 . Dans ce cas de figure, il conviendrait de dire que les deux métriques testées ont globalement la même efficacité. Or, en observant un peu plus en détail ces deux figures, il est possible d'affirmer que ces deux métriques n'ont pas le même comportement. En effet, concernant la figure 4.1(b), on observe que les scores prédits sont très proches de la valeur de référence, excepté pour deux valeurs. Ceci pourrait être interprété comme un problème dans la phase de conception de la métrique associée q_2 qui échoue à prédire correctement les scores pour deux images. En ce qui concerne la métrique d'évaluation q_1 (Fig.4.1(a)), de telles exceptions ne sont pas observables, même si les scores sont moins proches du score de référence. Ainsi, en tenant compte de la variance réduite dans les scores de corrélation, il est possible de conclure que la métrique q_1 est globalement meilleure que la métrique q_2 .



(a) Distribution des scores prédits par la métrique q_1 . (b) Distribution des scores prédits par la métrique q_2 .

FIGURE 4.1: Échantillon des scores prédits pour les métriques algorithmique d'évaluation de la qualité q_1 et q_2 .

Quoiqu'il en soit, l'interprétation des valeurs MOS obtenues est difficile. Supposons, par exemple, qu'un observateur humain attribue des notes de 3 et 4 (sur une échelle discrète à dix valeurs) pour deux images compressées (ou ayant subies un processus d'altération quelconque) et que pour deux autres images, il attribue les scores 7 et 8. Bien que la différence numérique de chaque couple d'images soit identique (et égale à 1), il est néanmoins quasiment impossible de conclure que la différence perceptuelle entre les images de la première paire est plus grande, identique, ou plus petite que la différence perceptuelle des images de la seconde paire. Les scores subjectifs permettent seulement un classement de la qualité des images par ordre croissant (ou décroissant).



Cette problématique d'évaluation de la sensation de différence existe également dans le domaine de la vision des couleurs, où l'un des enjeux consiste à associer une grandeur numérique de différence de couleur à une sensation de perception de cette différence. La CIE (Commission Internationale de l'Éclairage) a tenté d'apporter une solution à cette problématique en introduisant des espaces colorimétriques perceptuels (CIE Lab, CIE Luv, etc.) dans lesquels les calculs de différence colorimétrique sont régis par des équations à complexité variable.

Dans ces travaux menés en commun avec le Pr. Laurence T. Maloney de l'Université de New-York, le Dr. Kenneth Knoblauch de l'Unité Cerveau et Vision de l'INSERM et le Pr. Alan C. Bovik de l'Université du Texas à Austin, une méthode originale d'échelle de différence par maximum de vraisemblance (MLDS – *Maximum Likelihood Difference Scaling*) a été utilisée de manière à circonvenir les limitations imposées par les méthodes de notation de la qualité des images.

4.1 Échelle de différence par maximum de vraisemblance

La psychophysique est un domaine de recherche de la psychologie expérimentale ; domaine dans lequel on cherche à déterminer les relations quantitatives existantes entre un stimulus physique et sa perception par un être humain. Gustav Fechner [119] est considéré comme le fondateur de la psychophysique. Il fut le premier à avoir entrepris une mathématisation systématique de la relation entre stimulation et sensation à partir des travaux de Ernst Weber.

La psychophysique repose essentiellement sur le concept de seuil de détection. Il s'agit de la limite à partir de laquelle un individu est dans l'incapacité de dire si un stimulus est présent. Par extension, c'est aussi le seuil au-dessous duquel il ne parvient pas à distinguer deux stimuli. Dans ce cas, on utilise l'expression « seuil différentiel ».

Les algorithmes classiques de calcul du MOS sont basés sur une méthode psychophysique introduite par Stevens en 1946 [120] connu sous le nom d'estimation d'amplitude. En réponse aux critiques émises sur la fiabilité des données collectées par cette méthode, plusieurs autres méthodes d'échelle ont été développées, et notamment la technique MLDS. Cette méthode a été par le passé utilisée avec succès lors d'une étude conjointe avec le Pr. Maloney et le Dr. Knoblauch portant sur l'estimation de l'impact des niveaux de dégradation sur la perception de la qualité des images [121].

Une série d'images est constituée d'une image de base ϕ_1 et de l'ensemble de ses versions dégradées ϕ_2, \dots, ϕ_n , indexées par un niveau de dégradation croissant. Par convention, si l'image i a subi une dégradation d'un degré supérieur à l'image j , on notera $\phi_i > \phi_j$. Dans la suite du texte, le couple (i, j) fera référence au couple (ϕ_i, ϕ_j) .

Sur chaque présentation, l'observateur humain regarde simultanément deux paires de stimuli (i, j) et (k, l) affichées à l'écran. Chacun de ces quatre stimulus correspond à quatre niveaux de distorsion différents, incluant la possibilité d'avoir l'image originale. Nous ferons référence à ces deux paires en tant que quadruplet $i, j; k, l$. L'observateur juge alors la différence perceptuelle entre la première paire de stimuli (i, j) et la seconde paire (k, l) . Ainsi, tout au long du test psychophysique, l'observateur est amené à juger les différences de toutes les combinaisons de quadruplet possibles pour les N stimuli de la série ϕ_1, \dots, ϕ_p à p niveaux de dégradation.

L'objectif de la technique MLDS est d'assigner des valeurs numériques (ψ_1, \dots, ψ_p) qui peuvent être utilisées pour prédire la façon dont les observateurs ordonnent les paires de chacun des quadruplets. Ces valeurs numériques constituent l'*échelle des différences*. En principe, l'objectif est d'obtenir des valeurs numériques traduisant la sensation de l'observateur humain lors de l'évaluation des différences paire à paire de manière à avoir

$$\|\psi_i - \psi_j\| > \|\psi_k - \psi_l\|. \quad (4.1)$$

si et seulement si la différence perçue sur la paire (i, j) est estimée plus grande que sur la paire (k, l) .

Cependant, si les différences $\|\psi_i - \psi_j\|$ et $\|\psi_k - \psi_l\|$ sont proches, la fiabilité du jugement de l'observateur humain n'est plus assurée. Dans ce cas, il convient de prendre en compte cette variation stochastique du jugement dans un modèle modifié [122]. Soit $L_{ij} = \|\psi_i - \psi_j\|$ la longueur de l'intervalle (ψ_i, ψ_j) . Le modèle utilisé est un modèle gaussien d'égale variance dans lequel la variable est la différence des longueurs des intervalles :

$$\delta(i, j; k, l) = L_{ij} - L_{kl} = \|\psi_i - \psi_j\| - \|\psi_k - \psi_l\| \quad (4.2)$$

Le signal δ est dans ce cas supposé être corrompu par une erreur ε de loi gaussienne de moyenne 0 et d'écart-type σ de façon à former la variable du jugement

$$\Delta(i, j; k, l) = \delta(i, j; k, l) + \varepsilon. \quad (4.3)$$

Étant donné le quadruplet $(i, j; k, l)$, l'observateur est alors réputé avoir choisi la paire (i, j) lorsque $\Delta(i, j; k, l) > 0$. Ainsi, en utilisant ce modèle, lorsque l'amplitude de $\delta(i, j; k, l)$ est petite relativement à σ , l'observateur peut donner des réponses différentes pour un même stimuli. Le degré d'inconsistance dépend de l'amplitude de $\delta(i, j; k, l)$ par rapport à σ . Cette dépendance peut être également utilisée pour tester le modèle [122, 123].

Maloney et Yang [122] ont proposé une méthode d'estimation des valeurs d'échelle par maximisation de la vraisemblance. Or, étant donné que la règle de décision implique une simple combinaison des réponses internes, elles peuvent également être estimées selon un modèle linéaire généralisé (*GLM-Generalized Linear Model*) [121, 123].

Soit R_t la réponse donnée par l'observateur lors de la présentation du t -ième quadruplet $i, j; k, l$. La réponse R_t est codée de la façon suivante :

$$R_t = \begin{cases} 0 & \text{si différence perceptuelle de } (i, j) \text{ supérieure à } (k, l) \\ 1 & \text{sinon} \end{cases} \quad (4.4)$$

Dans ce cas, le modèle linéaire généralisé peut être exprimé par

$$g(\mathbb{E}[P(R = 1)]) = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p, \quad (4.5)$$

où l'opérateur de prédiction linéaire est lié à la valeur souhaitée de la réponse de l'observateur par la fonction $g(\cdot)$. Dans le cas du modèle de choix binaire (ce qui est le cas ici), la fonction $g(\cdot)$ est une fonction sigmoïde inverse de type fonction gaussienne. Les coefficients β_i qui correspondent aux valeurs $\psi - i$ de l'échelle, sont estimés par un processus itératif permettant d'avoir une solution en maximum de vraisemblance. En outre, le premier point ψ_1 ainsi que le dernier point ψ_p de l'échelle sont respectivement fixés à 0 et 1 de manière à obtenir une échelle normalisée et pouvoir ainsi comparer les diverses échelles.

4.2 Procédure d'optimisation d'algorithmes d'évaluation de la qualité.

4.2.1 L'algorithme d'évaluation de la qualité

L'indice multi-échelle de qualité MS-SSIM [14] implique un facteur de distorsion dédié à la luminance (l) auquel sont adjoints un facteur de distorsion de contraste (c) et un critère de mesure de distorsion de structure (s) calculés entre deux images f et g , tel que :

$$\text{MS-SSIM}(f, g) = \underbrace{\left[\frac{2\mu_f\mu_g + C_1}{\mu_f^2 + \mu_g^2 + C_1} \right]^{\alpha_M}}_{l(f,g)} \prod_{i=1}^M \underbrace{\left[\frac{2\sigma_f\sigma_g + C_2}{\sigma_f^2 + \sigma_g^2 + C_2} \right]^{\beta_i}}_{c(f,g)} \underbrace{\left[\frac{2\sigma_{f,g} + C_3}{\sigma_f^2\sigma_g^2 + C_3} \right]^{\gamma_i}}_{s(f,g)} \quad (4.6)$$

où μ_f et μ_g représentent l'intensité moyenne des images f et g ; σ_f et σ_g sont les écarts-types calculés sur les images f et g ; $\sigma_{f,g} = \frac{1}{N-1} \sum_{i=1}^N (f_i - \mu_f)(g_i - \mu_g)$. C_1 , C_2 et C_3 sont des constantes de stabilisation. $M = 5$ et à l'échelle $i = 1$, l'image est de résolution originale. Les valeurs originales des exposants sont $\beta_1 = \gamma_1 = 0.0448$, $\beta_2 = \gamma_2 = 0.2856$, $\beta_3 = \gamma_3 = 0.3001$, $\beta_4 = \gamma_4 = 0.2363$, et $\alpha_5 = \beta_5 = \gamma_5 = 0.1333$ [14].

4.2.2 Évolution de la corrélation en fonction du niveau de compression

La méthodologie précédemment décrite a ainsi été appliquée sur un large panel d'images afin d'évaluer l'impact du schéma de compression JPEG2000 sur la précision de la prédiction de l'indice MS-SSIM vis-à-vis des jugements humains.

4.2.2.1 Apparatus

Trente observateurs ont participé au test psychophysique. Tous ces observateurs ont une vision normale des couleurs et une acuité visuelle de 10/10 avec ou sans correction. Les images utilisées dans ce test sont les images de référence de la base LIVE. Aucune image dégradée issue de la base LIVE n'a été utilisée. Le type de dégradation utilisée est la compression JPEG2000. L'implantation de ce standard de compression utilisé ici est issue du projet JasPer [124]. L'algorithme a été appliqué sur chacune des 15 images de référence selon huit niveaux : 0.1000, 0.3057, 0.5627, 0.7684, 0.9741, 1.1798, 1.3854 et 1.5912 bpp. À ces 15×8 images dégradées, sont ajoutées les 15 images de référence afin de constituer notre base d'images.

Dans une première étape, l'aptitude des observateurs à classer une image de référence et ses huit versions dégradées par ordre de qualité décroissante a été testée. Cette étape permet

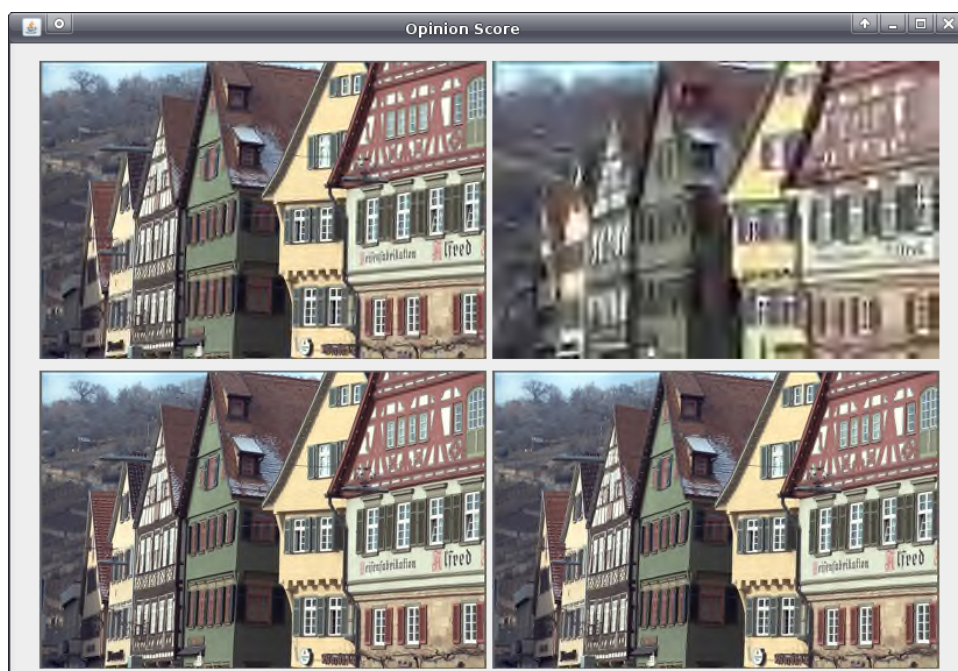


FIGURE 4.2: Exemple d'une épreuve présentée lors du test psychophysique associée à la technique MLDS. Les stimuli sont arrangés en deux paires d'images (i, j) et (k, l) . Pour chacune des paires affichées, le stimulus affiché à droite correspond à l'image la moins dégradée. L'observateur humain doit alors décider si la variation de qualité entre les stimuli i et j est plus grande ou plus petite que la variation de qualité entre les stimuli k et l . Dans cet exemple, la majorité des observateurs devraient juger que la paire affichée en haut exhibe la plus grande variation de qualité.

de vérifier qu'il n'y a pas de difficulté à discriminer les images adjacentes dans l'échelle des différences, faute de quoi la mise en place d'une telle échelle pourrait s'avérer impossible.

Durant une seconde étape, les observateurs humains visualisent un quadruplet d'images arrangé en deux paires (i, j) et (k, l) telle que présentées dans la figure 4.2. La première paire est affichée dans la partie haute de la fenêtre tandis que la seconde paire est affichée dans la partie basse. Afin de faciliter la prise de décision par l'observateur humain, l'image la moins dégradée est systématiquement affichée à gauche. L'observateur doit alors sélectionner la paire qui affiche la plus grande différence de qualité. Ces jugements sont ensuite utilisés pour construire l'échelle de différence de manière à capturer les effets de compression additionnelle sur la qualité des images.

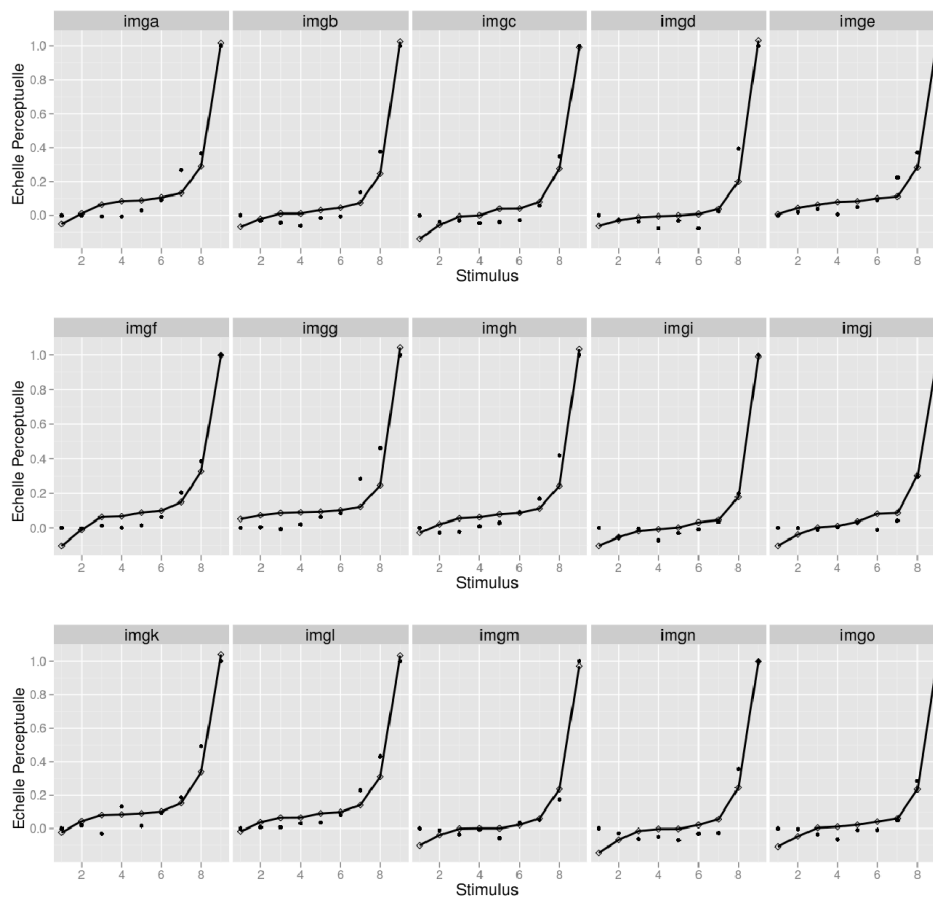


FIGURE 4.3: Résultats obtenus pour chacune des images de la base. Les points noirs représentent les valeurs MLDS et la courbe noire correspond aux scores MS-SSIM.

4.2.2.2 Résultats

De manière à pouvoir comparer les valeurs MLDS et les scores prédits par l'algorithme d'évaluation de qualité MS-SSIM, la différence des scores obtenus pour deux images consécutives (en terme de niveaux de compression) est calculée et cumulée le long de l'échelle de compression. Ces scores cumulés sont ensuite ajustés aux valeurs MLDS en utilisant un fonction de régression logistique.

Les résultats obtenus et présentés dans la figure 4.3 montrent que l'algorithme d'évaluation de la qualité des images MS-SSIM capture de manière satisfaisante les changements perceptuels lorsque le taux de compression croît. Cependant, il est possible d'observer que l'indice MS-SSIM échoue quelque fois à prédire efficacement de tels changements dans la perception de la qualité, comme cela se produit par exemple avec l'image `imgk`.

Une étude des raisons de tels échecs dans la prédiction de l'évolution de la qualité a été réalisée et les résultats sont présentés dans la figure 4.5 où l'efficacité de chacun des trois

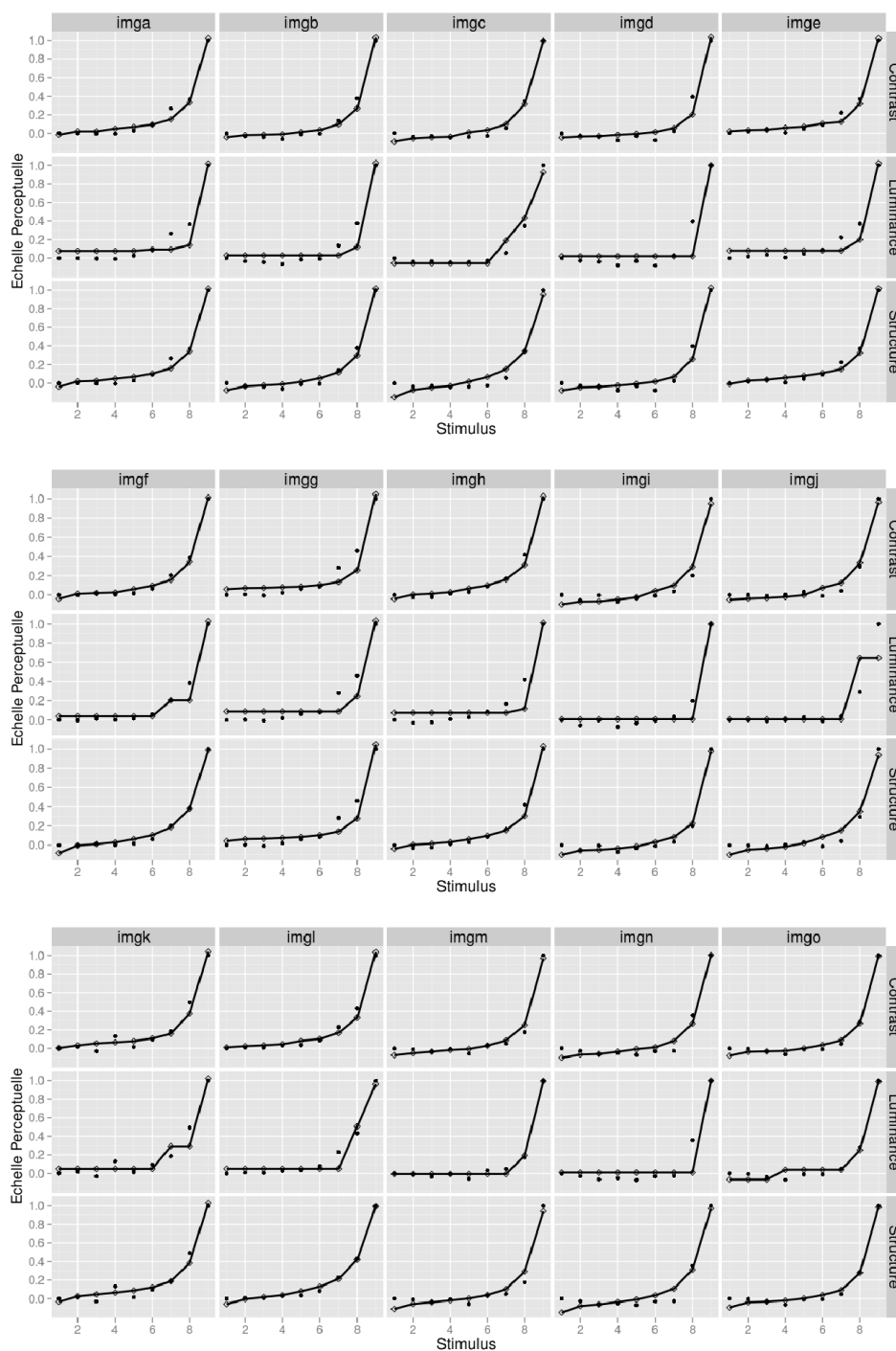


FIGURE 4.4: Résultats obtenus pour chacune des images de la base. Les points noirs représentent les valeurs MLDS et la courbe noire correspond à chacun des trois facteurs utilisés dans le calcul de l'indice MS-SSIM.

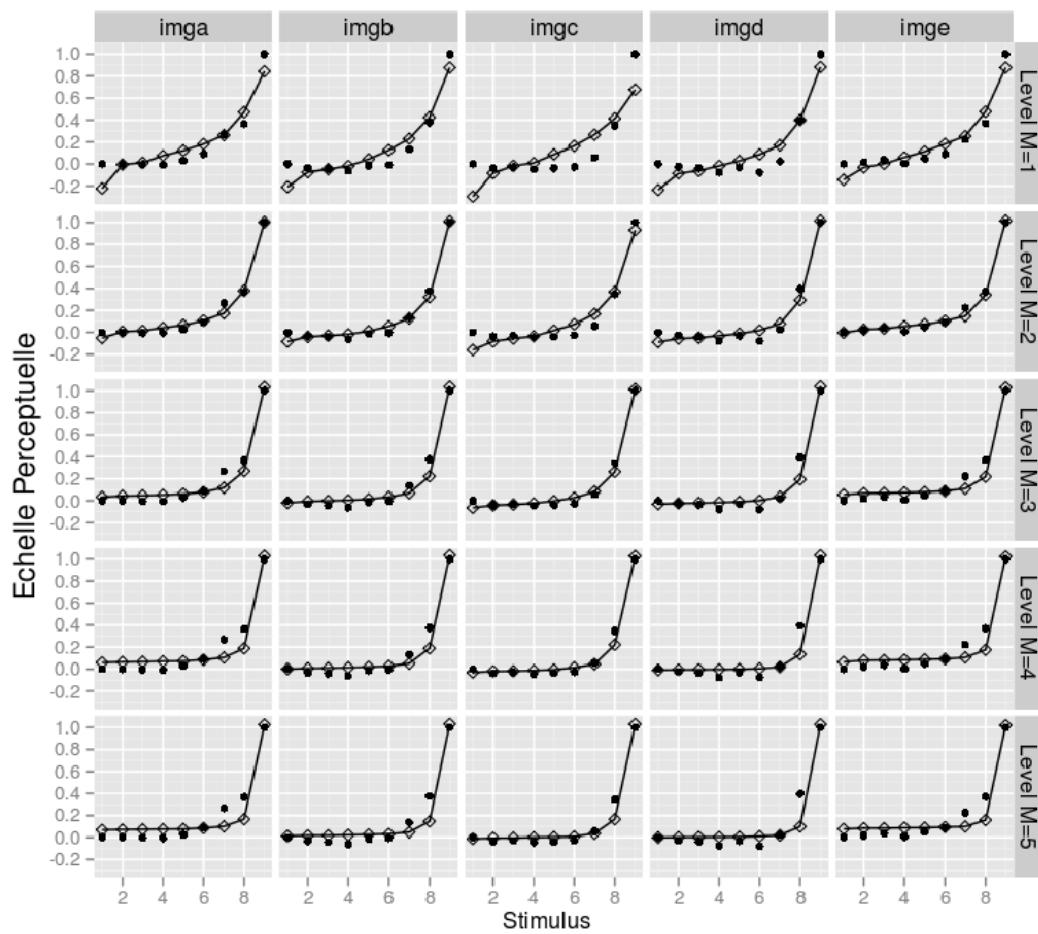


FIGURE 4.5: Évolution de la valeur de l'attribut de structure à divers niveaux de décomposition pour cinq images (*imga* à *imge*). Les points noirs représentent les valeurs MLDS et la courbe noire correspond au facteur de distorsion des structures utilisés dans le calcul de l'indice MS-SSIM.

critères utilisés pour le calcul de l'indice MS-SSIM est analysée. La première ligne de chacune des sous-figures correspond au critère de distorsion au contraste $\prod_{i=1}^M c_i(I, J)^{\beta_i}$, la deuxième ligne correspond aux valeurs du critère de luminance $l_M(I, J)^{\alpha_M}$ tandis que la troisième ligne correspond au critère de distorsion de structure $\prod_{i=1}^M s_i(I, J)^{\gamma_i}$. À première vue, le critère de distorsion de structure semble être le moins corrélé des trois facteurs avec les valeurs MLDS, et plus spécifiquement au début de l'échelle des différences. En outre, une faible adéquation est également observable de manière globale sur la figure 4.3.

Étant donné que le troisième facteur est initialement calculé sur cinq niveaux, l'intérêt est de connaître l'influence de chacun des niveaux. La figure 4.4 présente l'évolution des valeurs de distorsion de structure par rapport aux valeurs MLDS sur chacun des cinq niveaux de décomposition. Force est de constater que l'adéquation entre les deux mesures est évolutive : une faible correspondance est observée sur les deux premiers niveaux alors que la meilleure correspondance est notée à partir du troisième niveau de décomposition.

Une solution permettant une meilleure corrélation de ce troisième facteur serait d'introduire un facteur de pondération afin de modifier l'influence du troisième facteur sur le score global MS-SSIM. Dès lors, on peut légitimement s'interroger sur l'utilisation de facteurs de pondération communs entre le facteur de distorsion au contraste et de distorsion de structure. On peut même envisager une approche multirésolution pour le facteur de distorsion de luminance, alors que dans la formulation initiale, seul le facteur de distorsion de luminance au cinquième niveau de décomposition est pris en compte.

4.2.3 Optimisation par algorithme génétique

4.2.3.1 Stratégie de recherche

Dès lors, le principal objectif est d'estimer la valeur des quinze coefficients de l'indice MS-SSIM. Dans ce cas, l'expression de l'indice MS-SSIM peut être revue selon une formule paramétrique à 15 inconnues, telle que :

$$\text{MS-SSIM}(I, J, \alpha_i, \beta_i, \gamma_i; i = 1, \dots, M) = \prod_{i=1}^M \left[l_i(I, J)^{\alpha_i} c_i(I, J)^{\beta_i} s_i(I, J)^{\gamma_i} \right] \quad (4.7)$$

sous les contraintes $\sum_{i=1}^M \alpha_i + \beta_i + \gamma_i = 1$ et $\forall i \in [1, \dots, M], 0 \leq \alpha_i \leq 1, 0 \leq \beta_i \leq 1, 0 \leq \gamma_i \leq 1$.

La recherche des valeurs optimales des exposants peut alors être formulée comme la minimisation de la fonction d'erreur :

$$E(\alpha_i, \beta_i, \gamma_i; i = 1, \dots, M) = \min \left(\sum_{j=1}^K (\text{MLDS}_j(I, J) - \text{fSSIM}_j(I, J, \alpha_i, \beta_i, \gamma_i))^2 \right) \quad (4.8)$$

avec K le nombre des images testées pour lesquelles les valeurs MLDS sont disponibles, et $\text{fSSIM}_j(\cdot)$ la différence cumulée des valeurs MS-SSIM calculées pour deux images de taux de

Exposants	α_1	α_2	α_3	α_4	α_5
Valeur	0.1920	0.2169	0.2026	0.2136	0.1749
IC	[0.0989,0.2415]	[0.1877,0.2791]	[0.1692,0.2384]	[0.1765,0.2868]	[0.0814,0.2304]
Exposant	β_1	β_2	β_3	β_4	β_5
Valeur	0.9612	0.0097	0.0097	0.0097	0.0097
IC	[0.8288,0.9681]	[-0.0145,0.0933]	[0.0084,0.0112]	[0.0084,0.0112]	[-0.0133,0.1012]
Exposant	γ_1	γ_2	γ_3	γ_4	γ_5
Valeur	0.0082	0.1586	0.8167	0.0083	0.0082
IC	[0.0073,0.0086]	[0.1241,0.2530]	[0.7250,0.8501]	[0.0073,0.0086]	[0.0073,0.0086]

TABLE 4.1: Valeurs des 15 exposants et les intervalles de confiance (IC) à 95% utilisant une optimisation par algorithme génétique.

compression consécutifs et pour lesquelles une régression logistique a été réalisée. Étant donné que la fonction d'erreur est non-convexe et peut contenir plusieurs minima locaux, le choix d'une stratégie de recherche adaptée est crucial.

Parmi toutes les techniques d'optimisation, les algorithmes génétiques offrent de notre point de vue toutes les garanties nécessaires à la recherche d'un minimum global [125]. Dans notre approche, un chromosome est représenté par un vecteur à 15 dimensions $(\alpha_1, \dots, \alpha_M, \beta_1, \dots, \beta_M, \gamma_1, \dots, \gamma_M)$, et l'équation (4.8) représente la fonction de fitness. Qui plus est, puisque la fonction d'erreur est non-convexe, une procédure de bootstrap à 999 répliques a été utilisée afin de garantir une stabilité des résultats.

4.2.3.2 Résultats de l'optimisation

Le tableau 4.1 présente les résultats obtenus pour chacun des 15 exposants ainsi que l'intervalle de confiance à 95% associé, après minimisation de la fonction (4.8). Afin de mesurer la pertinence de ces nouvelles valeurs, une comparaison avec les valeurs originales est réalisée. La figure 4.6 présente le résultat de cette comparaison où les points noirs représentent les valeurs MLDS, la courbe continue noire les valeurs MS-SSIM calculées avec les exposants originaux, et la courbe noire pointillée fait référence aux valeurs MS-SSIM calculées avec les nouveaux exposants du tableau 4.1. Pour chacune des images testées, on observe un meilleur niveau de correspondance entre les valeurs MLDS et les nouvelles valeurs MS-SSIM. Ceci valide l'hypothèse initialement formulée sur l'utilisation de valeurs différentes pour chacun des quinze exposants pour optimiser l'indice MS-SSIM. En outre, si l'on considère les valeurs des exposants pour le facteur de distorsion de structure, on observe que le troisième niveau de décomposition est celui de prime importance puisque la valeur de l'exposant associé est la plus grande au regard des quatre autres valeurs. Ce qui implique que ce niveau de décomposition est celui qui est le mieux corrélé avec les valeurs MLDS. Cette constatation rejoint l'observation formulée précédemment lors de l'analyse de la figure 4.5.

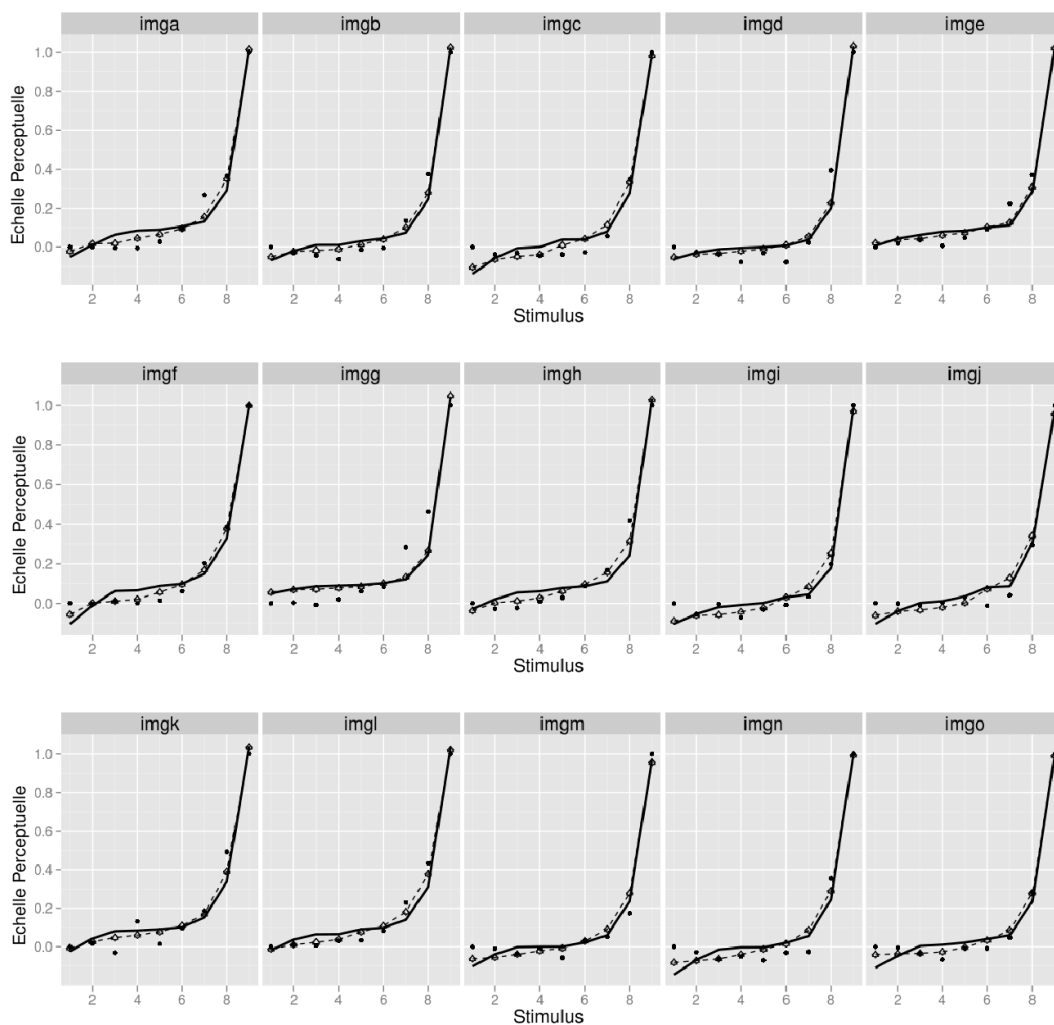


FIGURE 4.6: Résultats obtenus pour les images testées. Les points noirs représentent les valeurs MLDS, la courbe continue noire les valeurs MS-SSIM calculées avec les exposants originaux, et la courbe noire pointillée fait référence aux valeurs MS-SSIM calculées avec les nouveaux exposants du tableau 4.1.

	JP2K		JPEG		Bruit blanc	
	Original	New	Original	New	Original	New
CC	0.783	0.810	0.730	0.742	0.9153	0.9142
KROCC	0.884	0.884	0.849	0.852	0.8887	0.8878
SROCC	0.980	0.991	0.962	0.981	0.9825	0.9813
	Flou gaussien		FastFading		All	
	Original	New	Original	New	Original	New
CC	0.8864	0.8623	0.725	0.788	0.7980	0.8142
KROCC	0.8591	0.8413	0.859	0.876	0.8021	0.8543
SROCC	0.9725	0.9627	0.965	0.974	0.9464	0.9762

TABLE 4.2: Coefficients de corrélation calculés sur la base d'images LIVE pour les valeurs originales des exposants utilisées par l'indice de qualité MS-SSIM et les nouvelles valeurs.

Si l'on considère la situation de la mesure de distorsion de luminance pour laquelle seule la valeur au dernier niveau de décomposition est initialement prise en compte, le résultat obtenu après minimisation de la fonction (4.8) montre que les cinq niveaux de décomposition contribuent de manière quasiment identique dans l'estimation de la dégradation de luminance. Finalement, les nouvelles valeurs des exposants contribuent à une meilleure corrélation locale avec les valeurs MLDS.

4.2.4 Validation du protocole

Afin de juger la pertinence des nouvelles valeurs des exposants préalablement obtenues, une comparaison de la corrélation obtenue avec les valeurs DMOS est réalisée sur deux bases d'images : la base LIVE et la base TID2008. Trois mesures de corrélations sont utilisées : le coefficient de corrélation linéaire (LCC), le coefficient de corrélation de rang de Kendall (KROCC) et le coefficient de corrélation de rang de Spearman (SROCC).

De plus, pour pouvoir affirmer qu'une différence de corrélation est statistiquement significative entre les deux indices MS-SSIM, un test d'hypothèse basé sur la variance des erreurs résiduelles entre les valeurs DMOS et les scores prédits par chacun des deux indices MS-SSIM est réalisé. Ce test est basé sur le test de Fisher permettant de déterminer si les variances de deux populations sont égales, ce qui est réalisé en comparant le ratio des deux variances calculées. Dans notre cas, l'hypothèse nulle correspond au cas où les erreurs résiduelles de l'indice MS-SSIM original sont statistiquement non distinguables (à un niveau de confiance de 95%) des erreurs résiduelles de l'indice MS-SSIM amélioré. Le seuil ratio pour lequel les deux ensembles des erreurs résiduelles sont distinguables peut être obtenu à partir de la loi de Fisher.

Le tableau 4.2 présente les résultats obtenus sur la base LIVE. À la lecture des résultats, il convient d'affirmer que la performance de l'algorithme d'évaluation de la qualité MS-SSIM

	Dégradations											
	Bruit gaussien		Bruit couleur		Bruit corrélé		Bruit masqué		Bruit haute freq.		Bruit impuls.	
	Original	New	Original	New	Original	New	Original	New	Original	New	Original	New
CC	0.7994	0.7700	0.8151	0.7913	0.8278	0.8340	0.8341	0.8224	0.8861	0.8333	0.6672	0.6399
KROCC	0.6139	0.5767	0.6013	0.5677	0.6148	0.6241	0.6117	0.5977	0.6419	0.5887	0.4846	0.4575
SROCC	0.8099	0.7767	0.8055	0.7748	0.8215	0.8265	0.8099	0.7923	0.8706	0.8211	0.6899	0.6547
	Bruit quantif.		Flou gaussien		Débruitage		JPEG		JPEG2000		Trans. JPEG	
CC	0.8524	0.8355	0.9384	0.9292	0.9638	0.9485	0.9629	0.9796	0.9727	0.9823	0.8784	0.8983
KROCC	0.6569	0.6514	0.8169	0.7793	0.8316	0.8013	0.7489	0.7664	0.8559	0.8876	0.6637	0.6891
SROCC	0.8488	0.8361	0.9563	0.9355	0.9587	0.9458	0.9328	0.9571	0.9697	0.9812	0.8663	0.8852
	Trans. JPEG2000		Bruit non centré		Distor. de bloc		Décal. d'intensité		Chgt. contraste		All	
CC	0.8414	0.8437	0.7417	0.7388	0.7290	0.8666	0.7322	0.7259	0.7721	0.5468	0.8332	0.8532
KROCC	0.6766	0.6957	0.5254	0.5335	0.5038	0.6309	0.5345	0.5427	0.4748	0.4068	0.6577	0.6699
SROCC	0.8609	0.8849	0.7375	0.7434	0.7109	0.8353	0.7239	0.7402	0.6349	0.5430	0.8543	0.8601

TABLE 4.3: Coefficients de corrélation calculés sur la base d'images TID2008 pour les valeurs originales des exposants utilisées par l'indice de qualité MS-SSIM et les nouvelles valeurs.

utilisant les nouvelles valeurs des exposants est meilleure que l'indice MS-SSIM original. Cependant, ceci n'est pas exact pour les images dégradées par un bruit blanc ou un flou gaussien. Néanmoins, en considérant la totalité de la base LIVE, la valeur de corrélation SROCC est statistiquement significativement plus élevée lorsque les nouvelles valeurs des exposants est utilisée. Ceci est probablement dû à l'utilisation de la dégradation JPEG2000 pour construire l'échelle des différences.

Le tableau 4.3 présente les résultats obtenus lors de l'utilisation de la base TID2008. Si l'on considère les artefacts liés à la compression (dégradations « JPEG », « JPEG2000 », « Trans. JPEG » et « Trans. JPEG2000 »), on constate que les différences sont statistiquement significatives au profit de l'indice MS-SSIM calculé avec les nouvelles valeurs des exposants. En outre, la dégradation « Distor. de bloc », qui correspond à des distorsions locales de blocs d'intensité différentes, peut également être considérée comme le résultat d'un processus de compression. Pour cette dégradation, le gain de corrélation est statistiquement significatif. En revanche, le gain n'est pas statistiquement significatif lorsque la base TID est évaluée dans sa totalité.

4.3 Conclusion

Dans ces travaux, nous avons défini une méthodologie permettant de mesurer l'évolution des corrélations entre les scores prédis par une solution algorithmique et les notes de qualité attribuées par un être humain. Nous avons utilisé une nouvelle méthode psychophysique qui est une échelle des différences calculée par maximum de vraisemblance MLDS (*Maximum Likelihood Difference Scaling*) de manière à pouvoir outrepasser les limitations induites par les méthodes classiques d'évaluation subjective de la qualité (tests comparatifs n-aires, tests

de mesure absolue, etc.). Nous avons appliqué cette méthodologie sur un large panel d'images afin d'évaluer l'impact du schéma de compression JPEG2000 sur la précision de la prédiction de l'indice MS-SSIM vis-à-vis des jugements humains. Nous avons ainsi découvert que l'indice MS-SSIM présente des variations de précision, principalement dues au poids du troisième facteur de mesure de distorsion. La précision peut-être améliorée en modifiant les valeurs de pondération de chacun des trois facteurs de distorsion impliqués dans le calcul de l'indice MS-SSIM.

Dans le chapitre suivant, une synthèse des travaux présentés tout au long de ce manuscrit sera réalisée, suivie de perspectives pouvant être formulées.

Conclusion générale et perspectives

Sommaire

5.1 Synthèse	103
5.2 Perspectives	105

CE DOCUMENT résume mes activités de recherche de ces dix dernières années ; activités qui ont été pour la plupart d'entre elles réalisées au sein de l'Université de Caen Basse-Normandie. Est présentée maintenant une conclusion générale ouvrant la voie à plusieurs perspectives.

5.1 Synthèse

Dans le chapitre 2, deux méthodes d'évaluation agnostique de la qualité des images sans référence ont été développées. L'idée inhérente à l'approche retenue réfère au principe des statistiques de scènes naturelles (SSN). Ce principe, qui s'inscrit dans la thématique de la perception, repose sur l'hypothèse que les fonctions perceptuelles du SVH se sont adaptées aux statistiques des scènes naturelles dans lequel l'être humain évolue.

Le point commun des deux méthodes développées porte sur l'exploitation des statistiques calculées dans le domaine fréquentiel de la TCD, qui constitue notre cadre de travail. Dans un premier temps, le schéma développé intègre des caractéristiques « classiques » de distorsion induites par des propriétés du SVH telles que la sensibilité aux orientations des structures, aux variations de la netteté des structures des images lors de la vision fovéale et aux variations de contraste. Ainsi, quatre facteurs sont mis en place. Au lieu d'appliquer une combinaison linéaire de ces facteurs, une modélisation de leur distribution est réalisée à l'aide d'une fonction gaussienne généralisée multivariable. Cette métrique a été appliquée avec succès en biométrie afin de mesurer la qualité des données biométriques.

Dans une seconde approche, l'idée appliquée a été de dériver un modèle généralisé des SSN utilisant les coefficients DCT locaux, et d'utiliser les paramètres du modèle comme autant de facteurs que nécessaire. Ces derniers sont utilisés pour prédire la qualité des images.

À l'instar de ce qui a été réalisé précédemment, un modèle probabiliste généralisé multivariable est ensuite calculé à partir des caractéristiques retenues, ce qui permet d'aboutir à un schéma d'évaluation agnostique de la qualité des images. L'hypothèse majeure retenue dans cette approche s'appuie sur la capacité de l'algorithme d'évaluation de la qualité des images à apprendre la façon dont le modèle SSN varie en fonction des différents niveaux perceptuels des distorsions des images. L'algorithme est directement entraîné sur le modèle statistique paramétrique généralisé des coefficients DCT d'une image à divers niveaux de résolution.

Contrairement à ce qui a été réalisé dans la première approche, quasiment aucune utilisation directe des modélisations perceptuelles spécifiques existantes (telles que les décompositions corticales de l'aire V1 du cortex, les effets de masquage, etc.) n'est réalisée. Néanmoins, notre approche est perceptuellement cohérente puisque les modèles issus des SSN se veulent le reflet des propriétés statistiques du monde environnant influençant l'évolution des fonctions perceptuelles du SVH.

Le chapitre 3 a été le théâtre d'investigations dans le développement d'une métrique d'évaluation de la qualité lorsque l'on dispose d'une image de référence. En considérant la notation de la qualité d'une image comme un processus non naturel pour un observateur humain, le fil conducteur suivi dans cette approche est l'utilisation d'un schéma de classification multi-classes. Le choix du classificateur s'est porté sur le séparateur à vaste marge (SVM). Le nombre de classe utilisé dans ce schéma est fonction du nombre de classe de qualité utilisé dans les tests subjectifs de mesure de la qualité, en l'occurrence : cinq. Ainsi, en utilisant un processus de décomposition de type un-contre-un, dix classificateurs binaires SVM sont déployés. Au lieu de combiner les résultats obtenus par chacun des classificateurs par un processus de vote majoritaire, un processus de combinaison sous raisonnement incertain est appliqué. Ceci permet de prendre en compte les incertitudes de chacun des classificateurs binaires. Outre l'application de ce schéma pour la classification de la qualité des images, une utilisation dans le domaine de la segmentation des images de microscopie cellulaire a été réalisée.

Une fois exécutée la classification d'une image, il reste à calculer la note de qualité associée. Pour ce faire, une fonction de régression par SVM est générée pour chacune des classes de qualité. Ceci permet d'optimiser la fonction associée à chacune des classes et de maximiser la corrélation des scores prédits avec les notes de qualité associées aux observateurs humains.

Quels que soient les facteurs de distorsions développés, ces derniers permettent d'estimer un niveau relatif de dégradation de certains éléments jugés importants (contours, netteté, etc.) de l'image par rapport à une image originale. Plus surprenant, ces facteurs n'ont pas été utilisés pour estimer le niveau absolu de dégradation, ni la nature même de la dégradation appliquée. Dans une seconde application, ces attributs de qualité ont été détournés de leur utilisation classique afin de définir un processus d'estimation du niveau et du type de distorsion dans le cas de multi-dégradation. Les attributs retenus sont les trois facteurs définis dans l'indice SSIM de manière à construire un espace 3D dans lequel chacun des axes correspond à un des trois facteurs. La méthode développée se focalise sur la relation bijective entre les paramètres de

dégradation et le point 3D de l'espace SSIM.

Le chapitre 4 aborde l'utilisation des échelles de différence afin d'améliorer les niveaux de corrélation des méthodes algorithmiques d'évaluation de la qualité avec les scores donnés par les observateurs humains. L'échelle des différences par maximum de vraisemblance utilisée permet de dépasser les limites des méthodes traditionnelles de mesure subjective de la qualité. En effet, avec ces dernières, l'observateur juge de la qualité des images seules, ce qui revient à estimer l'amplitude de la qualité et ce qui pose la problématique de la fiabilité des données recueillies. En utilisant les échelles de différence, l'observateur doit juger les différences perceptuelles entre deux couples d'images, ce qui permet de répondre aux critiques émises précédemment. Cette technique a été appliquée dans le cas des images dégradées selon le schéma de compression JPEG2000. Elle a permis ainsi d'évaluer l'impact du schéma de compression test sur la précision de la prédiction de l'indice MS-SSIM vis-à-vis des jugements humains et d'ajuster les valeurs de pondération des trois facteurs de distorsion intégrés dans l'indice MS-SSIM.

5.2 Perspectives

Plusieurs extensions des travaux de recherche précédents peuvent être envisagées :

1. Dès lors que l'on dispose d'un indice fiable de qualité des images alors que l'image source n'est pas disponible, il est possible de l'intégrer dans les appareils d'acquisition. Ainsi, dans le domaine biomédical (images de microscopie cellulaire, d'IRMd, etc.), la calibration de l'appareil pourrait être asservie par un tel indice de qualité. L'opérateur serait alors assuré d'avoir une image de qualité optimale. Parallèlement, les images acquises par de tels systèmes sont très volumineuses. Il convient alors de les compresser pour en réduire le volume.

Actuellement, le schéma de compression avec perte le plus populaire est sans conteste le standard JPEG qui date des années 1990. De par son implantation hardware, il est déployé sur nombre d'appareils photos numériques et systèmes d'acquisition. Néanmoins, force est de constater que plus le taux de compression augmente, plus la qualité de l'image se dégrade rapidement. Il devient alors de plus en plus difficile d'identifier le contenu de l'image. Afin de pallier cet inconvénient, le standard JPEG2000 a été créé au début du XXI^e siècle. Ce standard permet d'atteindre des taux de compression supérieurs à ceux obtenus par JPEG, à qualité constante. Une particularité intéressante de ce standard consiste à appliquer un codage par région d'intérêt (*ROI-Region Of Interest*). Il s'agit d'une région de l'image qui est codée avec une plus grande précision, et correspond généralement à des régions présentant un intérêt particulier (ex : visage, plaque d'immatriculation, etc.). Cette plus grande précision se fait au détriment des autres zones de l'image qui sont alors compressées à un taux inférieur et donc beaucoup plus nettement dégradées. Dans le cadre de JPEG0000, la sélection du ROI est faite par l'utilisateur (*i.e.*, manuellement) ce qui peut s'avérer fastidieux et source d'erreur. Un second inconvénient

réside dans le fait que la région d'intérêt sera nécessairement dégradée (même si cela se fait à un degré moindre que le reste de l'image), ce qui pourra se révéler rédhibitoire pour des données sensibles et pour lesquelles aucune perte de précision n'est tolérée.

Un objectif sera alors de développer un outil de compression hybride intelligent permettant d'aboutir à un système de compression automatique par région d'intérêt. L'étiquetage des zones en fonction de leur importance sera réalisée par un schéma d'apprentissage statistique générique, adaptable à tout type de données numériques. La fonction de coût du schéma d'apprentissage intégrera la notion de qualité des données. Ainsi, le processus de mesure de la qualité des zones sensibles permettra de contrôler la puissance des dégradations et d'assurer ainsi une distorsion minimale (voire nulle) en utilisant un schéma de compression sans perte.

2. Avec l'avènement des télévisions 3D, se pose également la question de la qualité des images perçues par le téléspectateur. Dans ce cas, la notion de qualité devra intégrer le concept de profondeur avec les éventuels effets d'occlusion des objets de la scène 3D. De plus, si le téléspectateur se déplace pendant qu'il regarde un tel programme, une modification de la scène vue peut être effective (à l'instar de ce qui se passe lorsqu'un être humain se déplace dans une scène réelle). Dans ce cas, il sera nécessaire de prendre en compte le changement dynamique de la scène dans le temps, un peu comme cela est le cas dans les vidéos. La problématique liée au développement d'une métrique de qualité pour les scènes 3D devra également intégrer une notion de fluidité temporelle et de modification de la position de visualisation.
3. L'extension des méthodes de l'évaluation de la qualité des images sans référence aux vidéos requiert une compréhension de la dimension temporelle inhérente aux vidéos. En effet, une séquence vidéo n'est pas pas qu'une simple succession d'images (frames). De fait, la corrélation existante entre deux frames successives est très élevée et il est possible de déduire le contenu d'une frame à partir d'une autre en fournissant des informations de changement temporel du signal. Dès lors, la compréhension des mécanismes de base liés au mouvement et à la dimension temporelle est nécessaire de manière à développer une nouvelle méthode fiable et efficace d'évaluation de la qualité des vidéos. Cette problématique est actuellement abordée dans le projet de recherche ANR QuiAVU ANR-08-SECU-007-04.

Une des raisons pour lesquelles la mise au point d'une métrique de qualité sans référence de vidéo est un problème complexe, réside dans la très grande diversité des distorsions qui peuvent être rencontrées. Bien que de nombreuses études aient directement portées sur la modélisation des distorsions spatiales, très peu d'études ont eu pour objet la modélisation de la distorsion temporelle. De même, si la communauté s'accorde sur le fait que le mouvement est une donnée importante à prendre en compte lorsque l'on cherche à quantifier la qualité d'une vidéo, ce consensus s'étiolo lorsque l'on cherche à considérer l'influence d'attributs liés au mouvement, tels que la vitesse de déplacement, la taille de l'objet, le mouvement de la caméra, la direction du déplacement, et la saillance du mou-

vement (qui est un attracteur de l'attention). La modélisation des effets de masquage est très bien maîtrisée dans le domaine spatial, alors qu'elle pose encore de nombreux défis dans le domaine temporel. Qui plus est, lorsque l'on considère le cas de vidéos transmises par des canaux de communication (chaînes câblées, réception satellite, téléphones intelligents, etc.), les distorsions engendrées sont de natures spatiales et temporelles. Les distorsions spatiales (comme le flou) altèrent seulement l'apparence spatiale des frames de la vidéo, tandis que les distorsions temporelles (telles que les effets de gigue ou *jitter*) peuvent ne pas être directement visibles car ils affectent le mouvement des pixels et, dès lors, ne sont visibles qu'à travers une succession de frames. Il existe une troisième catégorie de distorsion qui sont de nature spatio-temporelle, telle que l'effet fantôme (*ghosting*), l'effet Gibb (*mosquito noise*), l'effet de smearing, etc. Le degré de gêne visuelle des distorsions temporelles et spatio-temporelles est à rapprocher du traitement du mouvement réalisé par les aires corticales du cerveau. Dès lors, l'amélioration de la modélisation des distorsions temporelles (incluant les distorsions spatio-temporelles) passe par une amélioration des modèles du traitement du mouvement par le cerveau humain.

L'approche retenue dans cette quête est la même que celle utilisée lors du développement d'une métrique agnostique de qualité, à savoir les statistiques des scènes naturelles. Ainsi, le traitement du mouvement par le cerveau humain est également adapté aux statistiques du mouvement dans son environnement naturel. Dès lors, l'objectif sera de développer une analyse statistique des mécanismes de transport de l'information de mouvement dans une scène visuelle en mouvement. Le flux optique est l'un de ces mécanismes largement utilisé dans les méthodes de détection de mouvement dans une vidéo. Néanmoins, peu d'études ont été consacrées à la régularité statistique de l'information de flux tant dans des vidéos exemptes de dégradation que dans les vidéos altérées.

Roth et Black [126] ont analysé les dérivées du flux optique dans certaines vidéos spécifiques ne contenant qu'un nombre restreint de mouvement. Dans cette étude, les objets sont statiques et le mouvement perçu est obtenu par des mouvements de caméra. Dans ce cadre strict, la régularité statistique des dérivées du flux optique est observée. Cependant, lorsque les contraintes sont relâchées (mouvement des objets, etc.), la régularité précédemment observée est perdue.

Les prochains travaux porteront sur l'examen des statistiques de mouvement et leur impact sur la perception visuelle de la qualité des vidéos. Deux points feront l'objet d'une attention particulière :

- (a) la régularité statistique des modèles de mouvement à travers les vidéos naturelles (séquence de l'environnement naturel captée par une caméra numérique).
- (b) évolution de la régularité statistique en fonction des distorsions.

Une étude préliminaire a permis de pointer une certaine régularité des statistiques dans des séquences vidéo de contenu similaire (*i.e.*, entre une séquence non altérée et ses différentes versions dégradées). Cependant, cette régularité n'existe plus lorsque les contenus varient fortement (distinction entre mouvement de caméra et objets en mouvement).

Une étude sur les raisons de cette perte de régularité devra être menée afin de développer un palliatif.

De plus, une analyse des caractéristiques contenant des informations pertinentes de mouvement influant la perception de la qualité des vidéos devra être conduite. Pour ce faire, il sera nécessaire d'étudier conjointement les statistiques spatiales et temporelles du flux optique afin de déterminer si la régularité statistique est conservée et si elle évolue de manière consistante et monotone en fonction de l'évolution de la qualité perçue des images.

Bibliographie

- [1] S. Freud. *Abrégé de la psychanalyse* (PUF, Paris, 1955).
- [2] C. Delgorge, C. Rosenberger, G. Poisson and P. Vieyres. *Towards a new tool for the evaluation of the quality of ultrasound compressed images*. IEEE Transactions on Medical Imaging 25, 1502 (2006).
- [3] P. L. Callet, C. Viard-Gaudin, S. Péchard and E. Caillault. *No reference and reduced reference video quality metrics for end to end qos monitoring*. Special Issue on multimedia Qos evaluation and management technologies E89, 289 (2006).
- [4] ITU-P Recommendation P.910. *Subjective video quality assessment methods for multimedia application*. Technical report, International Telecommunication Union, Geneva, Switzerland (2008).
- [5] ITU-R Recommendation BT.500-11. *Methodology for the subjective assessment of the quality of television pictures*. Technical report, International Telecommunication Union, Geneva, Switzerland (2002).
- [6] M. Pedersen and J. Y. Hardeberg. *Survey of full-reference image quality metrics*. Technical Report 5, Høgskolen i Gjøviks rapportserie (2009).
- [7] S. Winkler. *Video quality measures standards – current status and trends*. In *Intern. Conference on Information, Communications and Signal Processing 09* (2009), pp. 978–982.
- [8] Z. Wang and A. C. Bovik. *Mean squared error : Love it or leave it ? - a new look at signal fidelity measures*. IEEE Signal Processing Magazine 26, 98 (2009).
- [9] Z. Wang and E. P. Simoncelli. *Translation insensitive image similarity in complex wavelet domain*. In *IEEE Inter. Conf. Acoustics, Speech, Signal* (Philadelphia, PA, 2005), volume 2, pp. 573–576.
- [10] B. A. Wandell. *Foundations of Vision* (Sinauer Associates, Sunderland, Massachusetts, 1995).
- [11] W. S. Geisler and M. S. Banks. *Handbook of Optics* (McGraw-Hill, 1995), chapter Visual performance.
- [12] W. N. Charman. *Handbook of optics : Devices, measurements and properties* (MacGraw-Hall, 1995), volume II, chapter Optics of the eye. 2nd edition.
- [13] Z. Wang, E. P. Simoncelli and A. C. Bovik. *Multi-scale structural similarity for image quality assessment*. In *IEEE Asilomar Conference on Signals, Systems, and Computers* (2003), pp. 1398–1402.
- [14] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli. *Image quality assessment : From error measurement to structural similarity*. IEEE Transactions on Image Processing 13 (2004).

- [15] H. R. Sheikh and A. C. Bovik. *Image information and visual quality*. IEEE Transactions on Image Processing 15, 430 (2006).
- [16] D. M. Chandler and S. S. Hemami. *VSNR : A wavelet-based visual signal-to-noise ratio for natural images*. IEEE Transactions on Image Processing 16, 2284 (2007).
- [17] Z. Wang and E. P. Simoncelli. *Reduced-reference image quality assessment using a wavelet-domain natural image statistic model*. In *SPIE Human Vision and Electronic Imaging* (2005), volume 5666, pp. 149–159.
- [18] U. Engelke, M. Kusuma, H.-J. Zepernick and M. Caldera. *Reduced-reference metric design for objective perceptual quality assessment in wireless imaging*. Image Communication 24, 525 (2009).
- [19] A. M. Tourapis, A. Leontaris, K. Suhring and G. Sullivan. *H.264/14496-10 avc reference software manual*. Technical report, Joint Video Team (JVT) of ISO/IEC MPEG and ITU-T VCEG (2009).
- [20] H. R. Wu and M. Yuen. *A generalized block-edge impairment metric for video coding*. IEEE Signal Processing Letters 4, 317 (1997).
- [21] Z. Wang, A. C. Bovik and B. L. Evans. *Blind measurement of blocking artifacts in images*. In *International Conference on Image Processing (Vancouver, BC, 2000)*, pp. 981–984.
- [22] W. Gao, C. Mermer and Y. Kim. *A de-blocking algorithm and blockiness metric for highly compressed images*. IEEE Transactions on Circuits and Systems for Video Technology 12, 1150 (2002).
- [23] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu and X. Yang. *A locally adaptive algorithm for measuring blocking artifacts in images and videos*. Signal Processing : Image Communication 19, 499– (2005).
- [24] R. Muijs and I. Kirenko. *A no-reference blocking artifact measure for adaptive video processing*. In *European Signal Processing Conference (Eusipco)* (2005).
- [25] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi. *A no-reference perceptual blur metric*. In *IEEE International Conference on Image Processing (ICIP)* (Rochester, NY, 2002), volume 3, pp. 57—60.
- [26] E. Ong, W. Lin, Z. Lu, X. Yang, S. Yao, F. Pan, L. Jiang and F. Moschetti. *A no-reference quality metric for measuring image blur*. In *International Symposium on Signal Processing and Its Applications* (2003), pp. 469–472.
- [27] Y.-C. Chung, J.-M. Wang, R. Bailey, S.-W. Chen and S.-L. Chang. *A non-parametric blur measure based on edge analysis for image processing applications*. In *IEEE Conf. on Cybernetics and Intelligent Systems* (2004), pp. 356–360.
- [28] Z. P. Sazzad, Y. Kawayoke and Y. Horita. *No-reference image quality assessment for jpeg2000 based on spatial features*. Signal Processing : Image Communication 23, 257– (2008).

- [29] N. Ramin. *Vers une métrique sans référence de la qualité spatiale d'un signal video dans un contexte multimedia*. Ph.D. thesis, Université de Nantes (2009).
- [30] S. Winkler. *Vision models and quality metrics for image processing applications*. Ph.D. thesis, Dept. Elect. Eng., Ecole Polytechnique Fédérale de Lausanne (2002).
- [31] P. Marziliano, F. Dufaux, S. Winkler and T. Ebrahimi. *Perceptual blur and ringing metrics : Application to jpeg2000*. Signal Processing : Image Communication 19, 163– (2004).
- [32] X. Feng and J. Allebach. *Measurement of ringing artifacts in JPEG images*. In *SPIE* (2006), volume 6076, p. 74–83.
- [33] L. Hantao, N. Klomp and I. Heynderickx. *A no-reference metric for perceived ringing artifacts in images*. IEEE Transactions on Circuits and Systems for Video Technology 20, 529 (2010).
- [34] M. Jung, D. Lger and M. Gazelet. *Univariant assessment of the quality of images*. Journal of Electronic Imaging 11, 354 (2002).
- [35] H. Luo. *A training-based no-reference image quality assessment algorithm*. In *International Conference on Image Processing (ICIP)* (2004), pp. 2973—2976.
- [36] Y. R. Tsoy, V. G. Spitsyn and A. V. Chernyavsky. *No-reference image quality assessment through interactive neuroevolution*. In *International Conference on Computer Graphics and Vision* (2007), pp. 23–27.
- [37] R. V. Babu, S. Suresh and A. Perkiş. *No-reference JPEG image quality assessment using GAP-RBF*. Signal Processing 87, 1493– (2007).
- [38] A. K. Moorthy and A. C. Bovik. *Visual importance pooling for image quality assessment*. IEEE J. Selected Topics in Signal Process., Special Issue on Visual Media Quality Assessment 3, 193– (2009).
- [39] A. K. Moorthy and A. C. Bovik. *A two-step framework for constructing blind image quality indice*. IEEE Signal Processing Letters 17, 513– (2010).
- [40] S. Gabarda and G. Cristobal. *Blind image quality assessment through anisotropy*. Journal of the Optical Society of America 24, B42 (2007).
- [41] T. Brandao and M. P. Queluz. *No-reference image quality assessment based on dct-domain statistics*. Signal Processing 88, 822 (2008).
- [42] O. Desforges, L. Bedat and J. Ronsin. *La méthode LAR pour le codage basé région des images couleur*. In *Dix-huitième colloque sur le traitement du signal et des images—GRETSI* (2001), pp. 1037–1040.
- [43] VQEG. *Final report from the video quality experts group on the validation of objective models of video quality assessment*. Technical report (2000).
- [44] J. Mollon. *Tho' she kneel'd in that place where they grew*. J. Exp. Biol 146, 21 (1989).
- [45] C. A. Párraga and T. Troscianko. *Spatiochromatic properties of natural images and human vision*. Current Biology 12, 483 (2002).

- [46] D. J. Field. *Relations between the statistics of natural images and the response of cortical cells*. Journal of the Optical Society of America 4, 2379 (1998).
- [47] A. Srivastava, A. B. Lee, E. P. Simoncelli and S. C. Zhu. *On advances in statistical modeling of natural images*. Journal of Mathematical Imaging and Vision 18, 17 (2003).
- [48] D. L. Ruderman. *The statistics of natural images*. Network 5, 517 (1994).
- [49] E. P. Simoncelli and E. H. Adelson. *Noise removal via bayesian wavelet coring*. In *Thrid Int. Conf. on Image Proc.* (Lausanne, Switzerland, 1996), pp. 379–382.
- [50] M. J. Wainwright and E. P. Simoncelli. *Scale mixtures of gaussians and the statistics of natural imaes*. In *Adv. in Neural Information Processing System, NIPS* (MIT Press, 2000), pp. 855–861.
- [51] H. R. Sheikh, A. C. Bovik and L. Cormack. *No-reference quality assessment using natural scene statistics : JPEG2000*. IEEE Transactions on Image Processing 14, 1918 (2005).
- [52] J. Huang. *Statistics of natural images and models*. Ph.D. thesis, Division of applied mathematics, Brown University, Providence, RI, USA (2000).
- [53] J. A. van Hateren. *Independant component filters of natural images compared with simple cells in primary visual cortex*. Proc. Royal Statistical Society of London B 265, 359 (1998).
- [54] A. Hyvarinen, J. Karhunen and E. Oja. *Independant Component Analysis* (John Wiley and Sons, New-York, 2001).
- [55] K. D. Miller. *A model for the development of simple-cell receptive fields and the arrangement of orientation columns through activity dependent competetion between on- and off-center inputs*. Journal of NeuroScience 14, 409 (1994).
- [56] J. M. Shapiro. *Embedded image coding using zerotrees of wavelets coefficients*. IEEE Transactions on Image Processing 41, 3445– (1993).
- [57] R. W. Buccigrossi and E. P. Simoncelli. *Image compression via joint statistical characterization in the wavelet domain*. IEEE Transactions on Image Processing 8, 1688 (1999).
- [58] E. Y. Lam and J. W. Goodman. *A mathematical analysis of the dct coefficient distributions for images*. IEEE Transactions on Image Processing 9, 1661– (2000).
- [59] J. K. Romberg, H. Choi and R. Baraniuk. *Bayesian tree-structured image modeling using wavelet-domain hidden markov models*. IEEE Transactions on Image Processing 10, 1056– (2001).
- [60] C. S. Furmanski and S. A. Engel. *An oblique effect in human primary visual cortex*. Nature NeuroScience 3, 535 (2000).
- [61] M. Saad, A. C. Bovik and C. Charrier. *A DCT statistics-based blind image quality index*. IEEE Signal Proc. Letters 17, 583 (2010).
- [62] C. Shannon. *A mathematical theory of communication*. Bell System Technical Journal 27, 379 (1948).

- [63] N. Wiener. *Cybernetics or Control and Communication in the Animal and the Machine* (Hermann & Cie Editeurs, Paris, The Technology Press, Cambridge, Mass., John Wiley & Sons Inc., New York, 1948).
- [64] A. Rényi. *Some fundamental questions of information theory*. In *Selected Paper of Alfréd Rényi* (Pál Turán, 1960), volume 3, pp. 526–552.
- [65] P. Flandrin, R. G. Baraniuk and O. Michel. *Time-frequency complexity and information*. In *Int. Conf. on Acoustics Speech and Signal Proc.* (1994), volume 3, pp. 329–332.
- [66] E. Eltoft, T. Kim and T.-W. Lee. *On the multivariate laplace distribution*. *IEEE Signal Processing Letters* 13, 300 (2006).
- [67] Laboratory for Image & Video Engineering, University of Texas (Austin). *LIVE Image Quality Assessment Database*. <http://live.ece.utexas.edu/research/Quality> (2002).
- [68] N. Ponomarenko, M. Carli, V. Lukin, K. E. ans J. Astola and F. Battisti. *Color image database for evaluation of image quality metrics*. In *International Workshop on Multimedia Signal Processing* (Australia, 2008), pp. 403–408.
- [69] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors*. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27, 1615 (2005).
- [70] D. G. Lowe. *Distinctive image features from scale-invariant keypoints*. *International journal of computer vision* 60, 91 (2004).
- [71] U. of Essex. *Faces94 database, face recognition data*. In <http://cswww.essex.ac.uk/mv/allfaces/faces94.html>.
- [72] B. Hemery, C. Rosenberger and H. Laurent. *The ENSIB database : a benchmark for face recognition*. In *Int. Symp. on Signal Proc. and its Applications (ISSPA)* (2007).
- [73] P. Phillips, H. Wechsler, J. Huang and P. Rauss. *The FERET database and evaluation procedure for face recognition algorithms*. *Journal of Image and Vision Computing on Information* (1998).
- [74] A. Martinez and R. Benavente. *The AR face database*. Technical report, Computer Vision Center (1998).
- [75] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman and A. K. Jain. *FVC2002 : Second fingerprint verification competition*. In *Int. Conf. on Pattern Recognition (ICPR'02)* (2002), volume 3, pp. 811–814.
- [76] C. Rosenberger and L. Brun. *Similarity-based matching for face authentication*. In *Int. Conf. on Pattern Recognition (ICPR'2008)* (2008), pp. 1–4.
- [77] E. Tabassi and C. Wilson. *A novel approach to fingerprint image quality*. In *Int. Conf. on Image Processing (ICIP)* (2005), volume 2, pp. 37–40.
- [78] G. Saporta. *Probabilités, Analyse des données et Statistiques* (Technip, 1990).
- [79] D. J. Heeger. *Normalization of cell responses in cat striate cortex*. *Vision Neural Science* 9, 181 (1992).

- [80] E. P. Simoncelli and D. J. Heeger. *A model of neuronal responses in visual area MT*. Vision Research 38, 743 (1998).
- [81] R. Blake and R. Sekuler. *Perception* (McGraw Hill, 2006), 5th edition.
- [82] Z. Wang, A. C. Bovik and E. P. Simoncelli. *Structural approaches to image quality assessment*. In *Handbook of Image and Video Processing* (Academic Press, 2005), pp. 961–974. 2nd edition.
- [83] H. Hemminger and P. Mahler. *Psychophysical saturation scales and the spectral sensitivity in human vision*. Psychological Research 42, 207 (1980).
- [84] R. N. Shepard. *Psychological relations and psychophysical scales : On the status of direct psychophysical measurement*. Journal of Mathematical Psychology 24, 21 (1981).
- [85] A. Bouzerdoum, A. Havstad and A. Beghdadi. *Image quality assessment using a neural network approach*. In *Fourth IEEE Inter. Symp. on Signal Proc. and Information Tech., 2004*. (2004), pp. 330–333.
- [86] P. Gastaldo, R. Zunino, I. Heynderickx and E. Vicario. *Objective quality assessment of displayed images by using neural networks*. Signal Processing : Image Communication 20, 643 (2005).
- [87] M. Narwaria and W. Lin. *Objective image quality assessment based on support vector regression*. IEEE Transactions on Neural Networks 21, 515 (2010).
- [88] C. Charrier, G. Lebrun and O. Lezoray. *A color image quality assessment using a reduced-reference image machine learning expert*. In *SPIE, Image Quality and System Performance V* (San-Jose, California, 2008), volume 6808.
- [89] Z. Wang and A. C. Bovik. *A universal quality index*. IEEE Signal Processing Letters 9, 81 (2002).
- [90] A. B. Watson. *The cortex transform : Rapid computation of simulated neural images*. Computer Vis. Graphics and image proces. 39, 311 (1987).
- [91] J. Lubin. *Digital Images and Human Vision* (MIT Press, 1993), chapter The use of psychophysical data and models in the analysis of display system performance, pp. 163–178.
- [92] S. Daly. *A visual model for optimizing the design of image processing algorithm*. In *ICIP* (1994), volume 2, pp. 16–20.
- [93] E. P. Simoncelli and W. T. Freeman. *The steerable pyramid : a flexible architecture for multi-scale derivative computation*. In *ICIP* (Washington, DC, 1995), pp. 444–447.
- [94] V. N. Vapnik. *Statistical Learning Theory* (Wiley, New York, 1998).
- [95] J. Pearl. *Heuristics : Intelligent Search Strategies for Computer Problem Solving* (Addison-Wesley, 1984).
- [96] G. Shafer. *A mathematical theory of evidence* (Princeton University Press, 1976).

- [97] J. Platt. *Probabilities for SV machines*. In *Advances in Large-Margin Classifiers*, edited by D. S. P. J. Bartlett, B. Schölkopf and A. J. Smola (MIT Press, 2000), Neural Information Processing Systems.
- [98] A. Dempster. *Upper and Lower Probabilities Induced by Multivalued Mapping*. *Ann. Math. Statist.* 38, 325 (1967).
- [99] P. Smets. *Constructing the pignistic probability function in a context of uncertainty*. *Uncertainty in Artificial Intelligence* 5, 29 (1990). Elsevier Science Publishers.
- [100] T. Denoeux. *A k-nearest neighbor classification rule based on dempster-shafer theory*. *IEEE Transactions on Systems, Man and Cybernetics* 25, 804 (1995).
- [101] G. Celeux and J. Diebolt. *The SEM algorithm : A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem*. *Computational Statistics Quartely* pp. 73–82 (1985).
- [102] H. R. Sheikh, Z. Wang, L. Cormack and A. C. Bovik. *Live image quality assessment database release 2*. [Http ://live.ece.utexas.edu/research/quality](http://live.ece.utexas.edu/research/quality).
- [103] A. Trémeau and C. Charrier. *Influence of chromatic changes on the perception of color image quality*. *Color Research & Application* (1999).
- [104] D. Martin, C. Fowlkes, D. Tal and J. Malik. *A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics*. In *ICCV* (Vancouver, BC, 2001).
- [105] S. Philipp-Foliguet. *Evaluation de méthodes de segmentation d'images couleur*. In *Actes de l'école de printemps du GdR-ISIS, Opération Imagerie Couleur* (Pau, France, 2001).
- [106] A. L. Caron, P. M. Jodoin and C. Charrier. *Search strategies for image multi-distortion estimation*. In *ICPR 2010* (Istanbul, Turkey, 2010).
- [107] M. Buckley. *Fast computation of a discretized thin-plate smoothing spline for image data*. *Biometrika* 8, 247 (1994).
- [108] S. Chitale and W. T. Padgett. *Blur identification and correction for a given imaging system* (1999), pp. 268—273.
- [109] S.-C. Tai and S.-M. Yang. *A fast method for image noise estimation using laplacian operator and adaptive edge detection*. In *Int. Symp. on Comm. Cont. and Signal Proc.* (2008), pp. 1077—1081.
- [110] D. Garcia. *Robust smoothing of gridded data in one and higher dimensions with missing values*. *Comput. Statist. Data Anal.* 54, 1167– (2010).
- [111] R. P. Brent. *Algorithms for Minimization Without Derivatives* (Prentice-Hall : Englewood Cliffs, New Jersey, 1973).
- [112] J. Lagarias, J. Reeds, M. Wright and P. Wright. *Convergence properties of the nelder-mead simplex method in low dimensions*. *SIAM J. on Optimization* 9, 112 (1998).

- [113] M. Powell. *The newuoa software for unconstrained optimization without derivatives*. In *Large-Scale Nonlinear Optimization* (2006), volume 83.
- [114] K. Scheinberg. *Derivative free optimization method*. Technical report, IBM Watson Research Center (2000).
- [115] L. Piegl and W. Tiller. *The NURBS book (2nd ed.)* (Springer-Verlag, Inc., New York, USA, 1997).
- [116] M. Plass and M. Stone. *Curve-fitting with piecewise parametric cubics*. SIGGRAPH 17, 229 (1983).
- [117] W. H. Richardson. *Bayesian-based iterative method of image restoration*. Journal of the Optical Society of America 62, 55–59 (1972).
- [118] L. B. Lucy. *An iterative technique for the rectification of observed distributions*. Astronom. J. 79, 745 (1974).
- [119] G. T. Fechner. *Element der psychophysik* (Liepzig, Breitkopf and Härtel, 1860).
- [120] S. Stevens. *On the theory of scales measurement*. Science 103, 677 (1946).
- [121] C. Charrier, L. T. Maloney, H. Cherifi and K. Knoblauch. *Maximum likelihood difference scaling of image quality in compression-degraded images*. Journal of the Optical Society of America 24, 3418 (2007).
- [122] L. T. Maloney and J. N. Yang. *Maximum likelihood difference scaling*. Journal of Vision pp. 573–585 (2003).
- [123] K. Knoblauch and L. T. Maloney. *MLDS : Maximum likelihood difference scaling in R*. Journal of Statistical Software 25, 1 (2008).
- [124] U. University of Victoria. *The JasPer projet home page*. <http://www.ece.uvic.ca/mdadams/jasper/>.
- [125] J. H. Holland. *Adaptation in natural and artificial systems* (MIT Press, Cambridge, MA, USA, 1992).
- [126] S. Roth and M. Black. *On the spatial statistics of optical flow*. In *IEEE Int. Conf. on Computer Vision* (2005), volume 1, pp. 42—49.
- [127] A. Oppenheim, R. Shafer and T. Stockham. *Nonlinear filtering of multiplied and convolved signals*. IEEE Trans. on Audio and Electroacoustics 16, 437 (1968).
- [128] R. Gonzalez and R. Woods. *Digital image processing, 2nd Ed.* (Prince Hall, 2001).

Estimation des dégradations

La fonction d'étalement du point est une fonction mathématique décrivant la réponse d'un système d'imagerie à une source ponctuelle. Lors de l'application de flou, le processus de dégradation est représenté par le produit de convolution $g(x,y) = h_\beta * f(x,y)$, où g est la version floue de l'image de référence f , et h_β est la fonction d'étalement du point (PSF) à estimer. Une manière d'estimer cette fonction connaissant f et g est d'utiliser les filtres homomorphiques [127, 128].

Un filtre homomorphique repose sur l'hypothèse que la valeur achromatique d'un pixel $u(x,y)$ peut s'exprimer sous la forme du produit de l'illumination de ce pixel $i(x,y)$ (associée aux très basses fréquences) et de la réflectance dudit pixel $r(x,y)$ (correspondant aux plus hautes fréquences) :

$$u(x,y) = i(x,y) \cdot r(x,y) \quad (\text{A.1})$$

Le filtre homomorphique est un filtre fréquentiel. Cependant, la non-linearité de l'équation précédente ne permet pas d'accéder directement à un filtre fréquentiel car

$$\mathcal{F}[u(x,y)] \neq \mathcal{F}[i(x,y)] \cdot \mathcal{F}[r(x,y)] \quad (\text{A.2})$$

L'idée est alors de considérer le logarithme népérien de valeur achromatique du pixel afin d'être en mesure de dissocier la composante d'illumination et de réflectance :

$$\mathcal{F}[\log(u(x,y))] = \mathcal{F}[\log(i(x,y))] + \mathcal{F}[\log(r(x,y))] \quad (\text{A.3})$$

où \mathcal{F} correspond à la transformation de Fourier discrète. Ensuite, chacune des composantes est filtrée dans le domaine de Fourier par un filtre $H(u,v)$ atténuant les basses fréquences et réhaussant les hautes fréquences :

$$H(u,v) = (\gamma_H - \gamma_L)[1 - \exp(-c(D^2(u,v)/D_0^2))] + \gamma_L \quad (\text{A.4})$$

où γ_H et γ_L sont les fréquences de coupure respectivement dans les hautes et les basses fréquences, c représente la pente régulant la transition entre γ_L et γ_H , D_0 est une constante

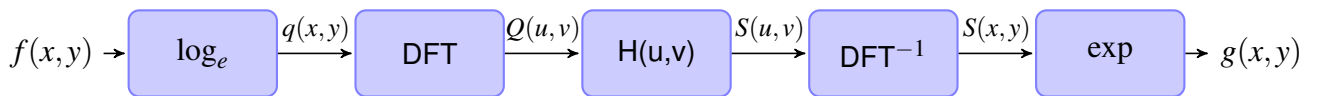


FIGURE A.1: Etapes d'un filtrage homomorphique.

et $D(u, v)$ est la distance du point (u, v) au centre du filtre. La figure A.1 illustre le principe du filtrage homomorphique.

Dans notre cas, h_β peut être défini par :

$$h_\beta(x, y) = \mathcal{F}^{-1} \{ \log[F] - \log[G] \}$$

et

$$\mathcal{F} \{ g(x, y) \} = \mathcal{F} \{ h_\beta * f(x, y) \} = G.H, \tag{A.5}$$

où G et F sont les versions spectrales de g et f .

Curriculum vitæ

ETAT CIVIL

CHARRIER Christophe
42 ans (09/09/69)
Marié, 3 enfants
Nationalité Française

Adresse professionnelle

Université de Caen Basse-Normandie
Laboratoire GREYC, UMR CNRS 6072, Équipe Image
Dept. Services et réseaux de communication
120, rue de l'exode, 50 000, Saint-Lô, France
Courriel : christophe.charrier@unicaen.fr

Université de Sherbrooke
Faculté des Sciences, Département d'Informatique
Département d'Informatique
2500, Bd de l'Université, Sherbrooke, QC, Canada, J1K 2R1
Courriel : christophe.charrier@usherbrooke.ca

POSTES OCCUPÉS

Depuis 2011 PROFESSEUR ASSOCIÉ au département Informatique de l'Université de Sherbrooke (Qc, Canada)

2010-2011 PROFESSEUR INVITÉ au département Informatique de l'Université de Sherbrooke, Qc, Canada

2010-2011 DÉLÉGATION CNRS au laboratoire XLIM-SIC de l'Université de Poitiers

2008-2009 PROFESSEUR INVITÉ au département Informatique de l'Université de Sherbrooke (Qc, Canada)

08-09/2008 PROFESSEUR INVITÉ au département de Génie Informatique et de Génie Électrique de l'Université du Texas à Austin (TX, USA).

Depuis 2001 MAÎTRE DE CONFÉRENCES section CNU 27, Département Services et Réseaux de Communication, IUT Cherbourg Manche, site de Saint-Lô, France

1999-2001 CHERCHEUR ASSOCIÉ au département de génie électrique et de génie informatique, Université Laval, Québec (Qc., Canada)

FORMATION

1998-1999 POST-DOCTORAT à l'Université Laval, Québec (Qc., Canada)

1995-1998 DOCTORAT Informatique option Images à l'Université Jean Monnet, Saint-Étienne, faculté des Sciences et Techniques.

Titre : Vers l'optimisation statistique et perceptuelle de la qualité pour la compression des images couleur par quantification vectorielle.

Soutenu le 24 novembre 1998, à l'Université Jean Monnet,

Rapporteurs : J-M. Jolion, D. Barba.

Examineurs : M. Barlaud, H. Cherifi, G. Eude, K. Knoblauch, M. Kunt, B. Laget (Président), M. Paindavoine, B. Picinbono.

1996-1998 MONITEUR de l'enseignement supérieur à l'Université Jean Monnet, Saint-Étienne

1994-1995 DEA IMAGE à l'Université Jean Monnet, Saint-Étienne, faculté des Sciences et Techniques.

1991-1993 LICENCE ET MAÎTRISE INFORMATIQUE à l'Université de Nantes, faculté des Sciences et Techniques.

RESPONSABILITÉS ADMINISTRATIVES

- Directeur des études du département SRC de l'IUT Cherbourg Manche (2006-2008)
- Chef du département SRC de l'IUT Cherbourg Manche (2003-2006)
- Responsable des relations industrielles du département SRC de l'IUT Cherbourg Manche (2001-2003 ; 2011-...)
- Membre élu de la commission de spécialistes en Informatique de l'Université de Caen (2007-2008)
- Membre de la commission de recrutement des PRAG à l'IUT Cherbourg Manche (2006)
- Membre de la commission de recrutement des ATER à l'IUT Cherbourg Manche (2002 - 2008)
- Membre élu du conseil scientifique du laboratoire LUSAC de l'Université de Caen (2006 - 2007)
- Responsable du site internet du Laboratoire de Radiocommunications et de Traitement du Signal (LRTS), Université Laval, Qc., Canada (1999-2001)
- Responsable du rapport annuel d'activité du Laboratoire de Radiocommunications et de Traitement du Signal (LRTS), Université Laval, Qc., Canada (1999-2001)

ACTIVITÉS DE RECHERCHE

→ **Thème 1** : Mesure de la qualité des images et des vidéos avec ou sans référence.

→ **Thème 2** : Développement de meta-classifieurs multiclassés.

→ **Domaines d'application** : Imagerie médicale, Protection des biens et des personnes

→ **Mots-clés** : Classification, Fusion de données, Apprentissage, Compression, Qualité, Vision humaine, Biométrie.

→ **Publications**

Types de publications	Nombre
Manuscrit de thèse	1
Chapitres de livre	4
Articles de revues internationales avec comité de lecture	11
Articles de revues nationales avec comité de lecture	3
Article de conférences internationales avec actes et comité de lecture	40
Article de conférences nationales avec actes et comité de lecture	16
Séminaire	3
Rapport	1
Total	79

Rayonnement

- Co-organisation de 5 manifestations (réunions GdR ISIS, conférence CORESA, session spéciale à CSIMTA)
- Membre du comité de programme des conférences CORESA, CGIV, IPTA
- Editeur associé de la revue IJRA
- Co-éditeur de 3 numéros spéciaux (IJRA, Image Communication, EURASIP)
- Collaborations internationales (Canada, USA)
- Collaborations nationales (SIC, Thalès, INSERM)
- Comité de lecture de quatre conférences internationales (International Conference on Visualization Imaging and Image Processing, CGIV, International Conference on Computing, Communication and Control Technologies) et de cinq revues (IEEE transactions on Image Processing, IEEE transactions on Information Technology in Biomedicine, IEEE transactions on Multimedia, Journal of Real-Time Imaging, Computer Vision and Image Understanding)
- Expert auprès de l'Agence Nationale de la Recherche (ANR), France (1 expertise).

Encadrement

- Prime d'Encadrement Doctoral et de Recherche – 2008-2012
- 4 thèses (1 soutenue, 3 en cours)
- 6 étudiants en Master de recherche et Licence

ACTIVITÉS D'ENSEIGNEMENT

J'ai eu l'occasion d'enseigner dans différents établissements d'enseignement supérieur ainsi qu'à différents publics comme moniteur de l'enseignement supérieur lors de ma thèse (UFR sciences économiques, Université Jean Monnet), comme chargé de cours lors de mon post-doc (Département de génie électrique et de génie informatique, à l'Université Laval, Canada), comme enseignant-chercheur (IUT Cherbourg Manche, ENSICAEN, IUT de Caen, Université de Sherbrooke–Canada)

Cours enseignés actuellement

Mon service d'enseignement pour l'année universitaire 2011-2012 est résumé dans le tableau ci-après. J'enseigne dans la formation initiale du département SRC de l'IUT Cherbourg Manche. Tous les cours sont en étroite relation avec mes activités de recherche.

Lieu	Module	Public	CM	TD	TP	Vol. eq. TD
DUT SRC	Algorithmique	bac+1	7.5h	18 h	45 h	74.25h
DUT SRC	Base de données	bac+1	4.5h	N/A	18 h	24.75h
DUT SRC	Outils mathématiques	bac+1	4.5h	18h	N/A	24.75h
DUT SRC	Compression vidéo	bac+2	4.5h	4.5h	27h	38.25h
DUT SRC	Système d'information	bac+2	N/A	N/A	24h	24h
DUT SRC	Programmation événementielle	bac+2	N/A	N/A	27h	27h

DÉTAIL DES MODULES ENSEIGNÉS

– Algorithmique

Public : DUT SRC 1ère année

Objectif : Aborder les techniques algorithmiques avancées, apprentissage des notions de complexité et de structure de données tout en mettant en œuvre les algorithmes sous-jacents. Le langage de programmation retenu pour l'ensemble du module est le langage orienté objet Java, puisqu'il constitue un des piliers de la programmation Web.

– Base de données

Public : DUT SRC 1ère année

Objectif : Apprendre aux étudiants à effectuer l'analyse des besoins ainsi que la modélisation d'une base de données. Sont abordées les notions de modèle relationnel et de modèle entité/association. Le langage de manipulation retenu est le langage SQL sous le SGBD MySQL. Les notions de sécurité et d'intégrité des données sont également abordées.

– Outils mathématiques

Public : DUT SRC 1ère année

Objectif : Maîtriser les bases mathématiques de l'informatique. Connaître le vocabulaire de la théorie des ensembles, d'éléments de logique et d'arithmétique.

– **Compression vidéo**

Public : DUT SRC 2ème année

Objectif : Maîtriser les différents codecs vidéos, ainsi que les différents supports et standards de stockage et de diffusion. Les étudiants apprennent ainsi les techniques de compression basées DCT et ondelettes, avant d'aborder les standards MPEG.

– **Système d'information**

Public : DUT SRC 2ème année

Objectif : Savoir mettre en place et concevoir des applications pour la diffusion et la gestion des contenu. En l'occurrence les étudiants manipuleront les CMS (systèmes de gestion de contenu) ainsi que les aspects liés à la sécurité des données, et notamment les problématique d'authentification et de cryptographie.

– **Programmation évènementielle**

Public : DUT SRC 2ème année

Objectif : Acquérir les notions de la base de la programmation évènementielle tout en utilisant une hiérarchie d'objet, telle le DOM (Document Object Model) afin de dynamiser et animer une interface utilisateur. Le langage de programmation utilisé est JavaScript.

Cours enseignés antérieurement

→ DIPLÔME UNIVERSITAIRE DE TECHNOLOGIE (1ER CYCLE) – 2001-2010

– **Outils mathématiques pour l'informatique**

Contenu : Théorie des ensembles, logique, analyse combinatoire, etc.

– **Outils mathématiques pour le signal**

Contenu : Série de Fourier, notion de spectre, applications, etc.

– **Les images informatiques et leurs traitements**

Contenu : Notion de contraste, d'accentuation, détection de contour, format des images, notion de transparence et de palette graphique, Standards de compression des images, etc.

– **Administration et Sécurité des réseaux**

Contenu : Notion de serveurs et clients, systèmes distribués, interconnexion, pare-feu, passerelle, routeur, chiffrement

– **Réseaux internet et services**

Contenu : Configuration de serveurs et de clients, installation, gestion et utilisation des

services Internet

– **Programmation multimédia/hypermédia**

Contenu : XHTML, Javascript, JSP, Web2.0, etc.

– **Programmation orientée objets**

Contenu : Notion d'objets, de classe, de hiérarchie, notions d'évènements, gestion des évènements

→ LICENCE PROFESSIONNELLE (2ÈME CYCLE) – 2005-2010

– **Systèmes d'information**

Contenu : Gestion des sessions, modélisation des traitements, gestion des contenus dynamiques

– **Compression des images**

Contenu : Compression avec perte, compression sans perte, quantification, standards de compression JPEG, JPEG2000, mesure de qualité

→ ECOLES D'INGÉNIEUR (3ÈME CYCLE) – 2006-2010

– **Systèmes intelligents**

Contenu : Apprentissage, agrégation de modèles, réseaux de neurones, algorithmes évolutionnaires, arbres de décision, Validation des modèles

– **Base de données avancées(2)**

Contenu : Indexation de grande bases d'images, systèmes distribués (réseaux pair-à-pair), catégorisation des images, mesure de similarité, descripteurs d'images, rétroaction

– **Informatique décisionnelle**

Contenu : Entrepôt de données, modélisation multidimensionnelle, modèles en étoiles, en flocons, SQL3, OLAP

→ UNIVERSITÉ DE SHERBROOKE (QC., CANADA) – 2009

– **IMN-659 – Analyse de la vidéo**

Contenu : Estimation de mouvement, modélisation du mouvement, segmentation spatio-temporelle, résumé dynamique/statique, Modélisation des objets, standard de compression de type MPEG, mesure de la qualité.

→ UNIVERSITÉ LAVAL (QC., CANADA) – 1999-2001

– **IFT-19965 – Programmation en C++ sous Unix**

Contenu : Introduction à la problématique de la programmation structurée, par objet et à la modélisation UML. Types, structures de contrôle, fonctions, mode de passage des paramètres, pointeurs et allocation dynamique de mémoire. Classes et objets : types,

constructeur et constructeur par recopie, destructeur, surdéfinition des fonctions et des opérateurs, structures de données, fonctions amies. Flots et formatage. Technique de l'héritage, niveaux de protection, héritage multiple et virtuel. Polymorphisme. Fonctions virtuelles. Patrons de fonctions et de classes.

– **IFT-19966 – Algorithmes de l'ingénieur I**

Contenu : Analyse des algorithmes : notions d'ordre, formule de récurrence, induction mathématique. Tableaux dynamiques, listes, piles et files. Arbres binaires, AVL et Splay, parcours d'arbre et récursivité. Adressage dispersé. Monceaux binaires, gauches et auto-ajustables. Algorithmes de tri.

– **GEL-19962 – Analyse des signaux**

Contenu : Définitions et propriétés de signaux usuels : rampe, porte, échelon ; impulsion et concept de distribution ; signaux périodiques et propriétés ; puissance et énergie d'un signal. Systèmes linéaires invariants en temps, convolution, calcul de la sortie d'un système. Transformée de Fourier, série de Fourier ; densités spectrales de puissance et d'énergie ; relation de Parseval.

ACTIVITÉS DE RECHERCHE

Encadrement d'étudiants 2ème et 3ème cycle

Nom	Type de formation	Période	Participation	Sujet
Viviana Weiss	Doctorat	2011-2014	50%	Classification des données des images HARDI dans le cas de maladies neuro-dégénératives
Michele Saad	Doctorat	2009-2012	50%	Mesure de la qualité des images sans référence. Application au cas des arrêts sur image d'une vidéosurveillance.
Mohamad El Abed	Doctorat	2008-2011	50%	Evaluation des systèmes biométriques
Gilles Lebrun	Doctorat	2002-2006	40%	Les machines à vecteurs supports pour la classification. Application en cancérologie
André-Louis Caron	Maîtrise	2009-2010	50%	Estimation de dégradations multiples dans une image
Jeremy Cosson	Master de recherche	2009	50%	Développement d'une librairie de gestion de vidéo de très longue durée
Antoine Aliotti	Master de recherche	2001	50%	Quantification des images couleur par la théorie de l'évidence
Kuldeep Kumar	Licence	2010	100%	Classification de données des images HARDI.
Sébastien Duchossoy	Licence	2009	100%	Intelligence virtuelle d'un avatar
Benjamin Leglinel	Licence	2008	100%	Interface Web JPEG2000

Collaborations nationales et internationales

- ❶ Monsieur Alan BOVIK, Professeur titulaire
Laboratoire LIVE – Université de Texas
Department of Electrical and Computer Engineering
Austin, USA
- ❷ Monsieur Maxime DESCOTEAUX, Professeur adjoint
Université de Sherbrooke
Faculté des Sciences
Sherbrooke (Québec)
- ❸ Monsieur Pierre-Marc JODOIN, Professeur adjoint

Université de Sherbrooke
Faculté des Sciences
Sherbrooke (Québec)

- ④ Madame Anne-Laure JOUSSELME, Chercheur
Defence R & D Canada - Valcartier (DRDC Valcartier)
Val-Bélair (Quebec)
- ⑤ Monsieur Kenneth KNOBLAUCH, Chargé de recherche
INSERM 486 : Institut Cellule Souche et Cerveau
Département Neurosciences Intégratives
Lyon, France
- ⑥ Monsieur Chaker LARABI, Maître de Conférences
Laboratoire XLIM, dept. SIC – Université de Poitiers
Futuroscope Chasseneuil, France
- ⑦ Monsieur Laurence T. MALONEY, Professeur titulaire
Department of Psychology, Center for Neural Science
New York University, New York, USA
- ⑧ Monsieur Sudipta MUKHOPADHYAY, Professeur adjoint
Indian Institute of Technology
Kharagpur, India
- ⑨ Monsieur Ettore MERLO, Professeur titulaire
Ecole Polytechnique de Montréal
Département de génie informatique et de génie logiciel
Montréal (Québec)
- ⑩ Monsieur Christophe ROSENBERGER, Professeur titulaire
Laboratoire GREYC, Université de Caen Basse-Normandie
Caen , France

Rayonnement national et international

- Professeur invité à l'Université de Sherbrooke (09/2010-09/2011, 10/2008-12/2009)
- Délégation CNRS au laboratoire XLIM-SIC de l'Université de Poitiers (09/2010-09/2011)
- Professeur invité à l'Université du Texas à Austin (06/2009, 07-08/2008)
- Membre de la commission de spécialistes en Informatique de l'Université de Caen (06/2007 - 12/2008)

-
- Membre du comité d'évaluation de quatre thèses (dont une à l'Université Laval, Qc.) : L. Martel, A. Mayache, G. Lebrun et V. Rosselini.
 - Evalueur externe pour les demandes de subvention
 - 4 dossiers pour la Region Bretagne (2002, 2004, 2006 et 2008)
 - 1 dossier pour CRSNG-découverte (2005)
 - 1 dossier ANR (2010)
 - Membre de la commission de recrutement de PRAG à l'IUT Cherbourg Manche (2006)
 - Membre de la commission de recrutement d'ATER à l'IUT Cherbourg Manche (2002 - 2009)
 - Membre du conseil scientifique du laboratoire LUSAC de l'Université de Caen (2006 - 2008)
 - Orateur invité
 - Ecole d'Hiver sur les Images Numériques Couleur (Lille 2005, Poitiers 2007 et Toulon 2009)
Cette école (organisée sous l'égide du club EEA et du GdR-ISIS) était à destination des chercheurs issus du monde universitaire et académique et des ingénieurs issus des centres de R&D industriels désireux d'acquérir un savoir-faire ou de parfaire leurs connaissances sur la perception de la couleur, la formation, l'acquisition et la reproduction des images numériques couleur, ainsi que les traitements des images numériques couleur.
 - Action spécifique
 - *Responsable scientifique* du projet ANR « Qualité des Images de VidéosUveillance » (2010-2012) pour l'Université de Caen-Basse Normandie
 - *Participation* au projet ANR « Fouille de Grandes Images Microscopiques » (2007-2010)
 - *Participation* au projet ANR « Apprentissage Statistique par une Architecture Profonde » (2009-2012)
 - *Participation* au projet CRITIC « Industrialisation d'un ensemble de composants logiciels pour la cytopathologie » (2006-2009)
 - **Animation** de l'action « protocoles et évaluation de la qualité perceptuelle » du thème B du GdR en Information-Signal-Image-Vision (2006-2011)
 - *Participation* à l'AS 163 du RTP 25 du département STIC du CNRS « Traitements perceptuels pour l'image numérique couleur » (2003-2004)
 - *Participation* à l'AS « Fouille des données image », 2003.

- *Participation* au projet « Traitement et Organisation de grandes masses de données images : applications en histopathologie cancérologique ». (2003 - 2006)
- Comité de rédaction, de programme
 - *Editeur invité* pour l'issue spéciale : « Processing and Analysis of High-Dimensional Masses of Image and Signal Data » de la revue EURASIP Journal of Applied Signal Processing (publication automne 09)
 - *Editeur invité* pour l'issue spéciale : « Machine Learning in Image Processing » de la revue EURASIP Journal of Applied Signal Processing (publication 05/08)
 - *Editeur en Chef* de l'issue spéciale « Color Image Processing and Analysis for Machine Vision » de la revue IJRA en février 2005.
 - *Editeur associé* de la revue International Journal of Robotics and Automation (09/03 - ...)
 - *Membre du Comité International de Programme* de la conférence annuelle IEEE « International Conference on Image Processing Theory, Tools and Applications » (2010 - ...)
 - *Membre du Comité International de Programme* de la conférence annuelle IS&T « European Conference on Colour in Graphics, Imaging, and Vision » (03/2007 - ...)
 - *Membre du Comité International de Programme* de la conférence annuelle « Visualization, Imaging, and Image Processing » (03/2003 - ...)
 - *Membre du Comité de Programme* du congrès « Compression et Représentation des Signaux Audiovisuels » (05/2005 - ...)
- Organisation de conférences
 - *Membre du comité d'organisation* de la conférence « High Performance Computing & Simulation », Caen, France, juin 2010
 - *Membre du comité d'organisation* de la conférence « International Conference in Image and Signal Processing », Cherbourg, France, juillet 2008
 - *Co-président* de la conférence « COmpression et REprésentation des Signaux Audiovisuel » (CORESA), Caen, France, novembre 2006
 - *Co-président* de la session spéciale « Color Image Processing and Image Analysis for Machine Vision » organisée lors de la conférence internationale Complex Systems, Intelligence and Modern Techonology Applications qui a eu lieu à Cherbourg en septembre 2004.
 - *Co-organisateur* de la « journée de l'image couleur : indexation et perception » depuis juin 2002 sous l'égide du GdR-ISIS 720. Depuis 2002, cette journée est pérenne et est reconduite annuellement.
- Evalueur externe (comité de lecture)
 - *Revues internationales* : IEEE transactions on Image Processing, IEEE transactions

on Information Technology in Biomedicine, IEEE transactions on Multimedia, Journal of Real-Time Imaging, Computer Vision and Image Understanding.

→ *Revue nationale* : Traitement du signal

→ *Conférences internationales* : International Conference on Visualization Imaging and Image Processing (depuis 2003), Conference on Color in Graphics, Imaging and Vision (depuis 2006), International Conference on Computing, Communication and Control Technologies (2002,2003,2004)

→ *Conférence nationale* : COmpression et REprésentation des Signaux Audiovisuel (depuis 2005)

Projets de recherche

- projet Franco-Québécois Samuel de Champlain « Clafiger » (2011-2012) : 12 000 euros
Je suis responsable scientifique pour l'équipe française.
- projet ANR QuIAVU ANR-08-SECU-007-04 (2008-2011) : 125 000 euros
Je suis responsable scientifique pour l'Université de Caen Basse-Normandie.
- fond Région Basse-Normandie de démarrage de projet (2009) : 6 000 euros
Ce fond m'a permis de faire un séjour de deux mois chez le Pr. Al Bovik à l'Université du Texas à Austin, ce qui s'est traduit par une co-supervision de thèse.
- fond Région Basse-Normandie d'accueil jeune chercheur (2002) : 10 000 euros

PUBLICATIONS

Thèse

- [1] C. Charrier, *Vers l'optimisation statistique et perceptuelle de la qualité pour la compression des images couleur par quantification vectorielle*. PhD thesis, Université Jean Monnet, Saint-Etienne, France, 1998.

Chapitre de Livres

- [1] A. Saadane, M. C. Larabi, and C. Charrier, *Imagerie Numérique Couleur : avancées et perspectives.*, ch. Couleur et Compression d'images. à paraître.
- [2] M. C. Larabi, A. Saadane, and C. Charrier, *Imagerie Numérique Couleur : avancées et perspectives.*, ch. Approches d'évaluation de la qualité. à paraître.
- [3] O. Lezoray, G. Lebrun, C. Meurie, C. Charrier, A. Elmoataz, M. Lecluse, and H. Elie, *Handbook of Research on Advanced Techniques in Diagnostic Imaging and Biomedical Applications*, ch. Machine learning in morphological segmentation of microscopic images, pp. 320–334. Themis P. Exarchos, Athanasios Papadopoulos, Dimitrios I. Fotiadis, information science reference ed., 2009.
- [4] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot, *Encyclopedia of Artificial Intelligence*, ch. an EA multi-model selection for SVM Multiclass schemes, pp. 520–525. Juan R. Rabuñal, Julian Dorado & Alejandro Pazos, information science reference ed., 2008.

Articles de revues internationales avec comité de lecture

- [1] M. Saad, A. C. Bovik, and C. Charrier, "No-reference quality assessment of images by modeling of natural DCT statistics," *IEEE Transactions on Image Processing*, 2011.
- [2] M. Saad, A. C. Bovik, and C. Charrier, "A DCT statistics-based blind image quality index," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 583–586, 2010.
- [3] O. Lezoray and C. Charrier, "Color image segmentation using morphological clustering and fusion with automatic scale delection," *Pattern Recognition Letters*, pp. 397–405, 2009.
- [4] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot, "Tabu search model selection for SVM," *International Journal of Neural Systems*, vol. 18, 2008. 19–31.

- [5] C. Charrier, L. T. Maloney, H. Cherifi, and K. Knoblauch, "Difference scaling of image quality in compression-degraded images," *Journal of the Optical Science of America*, vol. 24, no. 11, pp. 3418–3426, 2007.
- [6] C. Charrier, G. Lebrun, and O. Lezoray, "Evidential segmentation of microscopic color images with pixel classification posterior probabilities," *Journal of Multimedia*, vol. 2, no. 3, pp. 57–65, 2007.
- [7] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "A fast and efficient segmentation scheme for cell microscopic image," *Cellular and Molecular Biology, special issue on Biomedical Signal and Image Processing*, vol. 2, no. 53, pp. 51–61, 2007.
- [8] C. Meurie, O. Lezoray, C. Charrier, and A. Elmoataz, "Combination of multiple pixel classifiers for microscopic image segmentation," *International Journal of Robotics and Automation, special Issue on Colour Image Processing and Analysis for Machine Vision*, vol. 20, pp. 63–69, 2005.
- [9] G.-Y. Delisle, Z. Sebbani, C. Charrier, and F. Cote, "A novel approach to complex target recognition using RCS wavelet decomposition," *IEEE Antennas and Propagation Magazine*, vol. 47, pp. 35–55, Feb. 2005.
- [10] C. Charrier and C. Cherifi, "A classified vector quantization scheme for color image coding perceptually tuned," *Image Processing and Communications, an International Journal*, 2002. papier invité.
- [11] A. Trémeau and C. Charrier, "Influence of chromatic changes on the perception of color image quality," *Color Research and Application*, vol. 25, pp. 200–213, June 2000.

Articles de revues nationales avec comité de lecture

- [1] M. El-Abed, B. Hemery, C. Charrier, and C. Rosenberger, "évaluation de la qualité de données biométriques," *RNTI journal, special issue on "Qualité des Données et des Connaissances"*, 2011.
- [2] O. Lezoray and C. Charrier, "Segmentation d'images couleur par coalescence non supervisée d'histogrammes 2D et fusion de régions selon la théorie de dempster-shafer," *Traitement du Signal*, vol. 21, no. 6, pp. 605–622, 2004.
- [3] C. Charrier and H. Cherifi, "Construction statistique du dictionnaire sous contraintes psychophysiques," *Traitement du Signal*, vol. 16, no. 2, pp. 105–119, 1999.

Article de conférences internationales avec actes et comité de lecture

- [1] C. Charrier, K. Knoblauch, L. T. Maloney, and A. C. Bovik, "Calibrating MS-SSIM for compression distortions using MLDS," in *IEEE International Conference on Image Proc. (ICIP)*, (Brussels, Belgium), 2011. accepted, to be presented.
- [2] M. Saad, A. C. Bovik, and C. Charrier, "DCT-statistics model-based blind image quality assessment," in *IEEE International Conference on Image Proc. (ICIP)*, (Brussels, Belgium), 2011. accepted, to be presented.
- [3] M. El-Abed, R. Giot, B. Hemery, C. Charrier, and C. Rosenberger, "A SVM-based model for the evaluation of biometric sample quality," in *IEEE Workshop on Computing Intelligence in Biometrics and Identity Management*, (Paris, France), 2011.
- [4] M. E. Abed, R. Giot, C. Charrier, and C. Rosenberger, "Evaluation of biometric systems : An svm-based quality index," in *Conference on Information Security NISK 2010*, Nov. 22-24 2010.
- [5] M. Saad, A. C. Bovik, and C. Charrier, "Natural DCT statistics approach to no-reference image quality assessment," in *IEEE International Conference on Image Proc. (ICIP)*, (Hong-Kong, PRC), 2010.
- [6] A. L. Caron, P. M. Jodoin, and C. Charrier, "Search strategies for image multi-distortion estimation," in *IEEE International Conference on Pattern Recognition (ICPR)*, (Istanbul, Turkey), 2010.
- [7] C. Charrier and O. Lezoray, "Color vq-based image compression by manifold learning," in *International Conference on Image and Signal Processing (ICISP)*, (Trois-Rivière, QC, Canada), 2010.
- [8] C. Charrier, K. Knoblauch, A. K. Moorthy, A. C. Bovik, and L. T. Maloney, "Comparison of image quality assessment algorithms on compressed images," in *SPIE, Image Quality and System Performance VII*, (San-Jose, California), Jan. 2010.
- [9] C. Charrier, "On the design of color image quality assessment using both achromatic and chromatic contrast representation," in *6th IEEE International Symposium on Image and Signal Processing and Analysis*, (Salzburg, Austria), Sept. 2009.
- [10] O. Lezoray, C. Charrier, and A. Elmoataz, "Learning complete lattices for manifold mathematical morphology," in *International Symposium on Mathematical Morphology*, (Groningen, The Netherlands), Aug. 2009.
- [11] O. Lezoray and C. Charrier, "Rank transformation and manifold learning for multivariate mathematical morphology," in *19th European Signal Processing Conference*, (Glasgow, Scotland), Aug. 2009.

- [12] C. Charrier, G. Lebrun, and O. Lezoray, "Image quality assessment with manifold and machine learning," in *SPIE, Image Quality and System Performance VI*, vol. 7242, (San-Jose, California), Feb. 2009.
- [13] C. Charrier and O. Lezoray, "Multiresolution image VQ compression by color codebook reordering," in *International Conference Computer Graphics, Imaging and Visualization (CGIV)*, (Barcelona, Spain), Sept. 2008.
- [14] C. Charrier, G. Lebrun, and O. Lezoray, "Color image quality assessment using a reduced-reference image obtained from a classification process," in *SPIE, Image Quality and System Performance V*, vol. 3025, (San-Jose, California), Feb. 2008.
- [15] C. Charrier, G. Lebrun, and O. Lezoray, "Selection of features by a machine learning expert to design a color image quality metric," in *Third Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, (Scottsdale, Arizona), pp. 113–119, 2007.
- [16] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot, "Speed-up LOO-CV with SVM classifier," in *7th Int. conf. on Intelligent Data Eng. and Automated Learning (IDEAL)*, vol. LNCS 4224, (Burgos, Spain), pp. 99–107, Sept. 2006.
- [17] G. Lebrun, O. Lezoray, C. Charrier, and H. Cardot, "A new model selection method for SVM," in *7th Int. conf. on Intelligent Data Eng. and Automated Learning (IDEAL)*, vol. LNCS 4224, (Burgos, Spain), pp. 108–115, Sept. 2006.
- [18] C. Charrier, G. Lebrun, and O. Lezoray, "A machine learning-based color image quality metric," in *IS&T Third European Conference on Colour in Graphics, Imaging and Vision (CGIV)*, (Leeds, UK), pp. 155–159, June 2006.
- [19] C. Charrier, G. Lebrun, and O. Lezoray, "Fusion of SVM-based microscopic color images through colorimetric transformation," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. II, (Toulouse, France), pp. 1112–1115, May 2006.
- [20] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "Fast pixel classification by SVM using vector quantization, tabu search and hybrid color space," in *International Conference on Computer Analysis of Images and Patterns* (Springer, ed.), vol. Lecture Notes in Computer Science 3691, pp. 685–692, 2005.
- [21] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "A fast and efficient segmentation scheme for cell microscopic image," in *World Congress of Cellular and Molecular Biology*, (Poitiers, France), pp. 55–61, Oct. 2005.
- [22] C. Charrier and M.-C. Larabi, "A psychovisual color image quality metric integrating a color contrast sensitivity function," in *International Conference on Imaging : Technology and Applications for the 21st Century*, (Beijing, P.R. of China), May 2005.
- [23] C. Charrier and T. Eude, "A psychovisual color image quality metric integrating both intra and inter channel masking model," in *13th European Signal Processing Conference (EU-SIPCO)*, (Vienna, Austria), pp. 2123–2126, Sept. 2004.

- [24] G. Lebrun, C. Charrier, and H. Cardot, "SVM training time reduction using vector quantization," in *International Conference on Pattern Recognition (ICPR)*, (Cambridge, UK), Aug. 2004.
- [25] C. Charrier and A.-L. Jusselme, "Color space combination and approximation for vector quantization," in *International Symposium on Image/Video Communications over fixed and mobile networks*, (Brest, France), pp. 231–234, July 2004.
- [26] C. Charrier and O. Lezoray, "Influence of JPEG2000 standard compression on color image segmentation quality," in *IAPR International Conference on Image and Signal Processing*, vol. 2, (Agadir, Morocco), pp. 389–396, June 2003.
- [27] C. Charrier and T. Eude, "A new visual masking tool for JPEG2000," in *Data Compression Conference (DCC)*, (Snowbird, Utah), pp. 267–275, 2003.
- [28] T. Eude and C. Charrier, "A masking model for scalar non fixed deadzone quantization in lossy JPEG2000 compression," in *Picture Coding Symposium (PCS)*, (Saint-Malo, France), pp. 129–134, Apr. 2003.
- [29] C. Charrier and T. Eude, "Visual masking model for perceptual color image compression," in *SPIE International Congress on High-Speed Photography and Photonics*, vol. 4048, (Beaune, France), pp. 713–718, Sept. 2002.
- [30] C. Charrier, T. Eude, and H. Cherifi, "Edge classification of color images based on psychophysical judgments," in *SPIE Electronic Imaging 2001*, (San-Jose, CA), Jan. 2001.
- [31] C. Charrier and H. Cherifi, "A vector quantization scheme for color images based on a perceptual preclassification of edges," in *IASTED Signal Processing and Communications International Conference 2000*, (Marbella, Spain), pp. 184–189, 19-22 Sept. 2000.
- [32] C. Charrier and G. Y. Delisle, "Determination of characteristic features of a RCS using wavelet analysis," in *IEEE APS International Symposium*, vol. 4, (Salt Lake City, Utah), pp. 2166–2169, July 2000.
- [33] C. Charrier, T. Eude, and H. Cherifi, "Métriques d'évaluation de la qualité des images monochromes compressées," in *International Symposium on Image/Video Communications over Fixed and Mobile Networks*, vol. 1, (Rabat, Morocco), pp. 170–179, Apr. 2000.
- [34] C. Charrier and G. Y. Delisle, "Radar targets characterization using wavelets pyramidal decomposition," in *ICCEA'99*, (Beijing – China), pp. 467–470, Nov. 1999.
- [35] C. Charrier and H. Cherifi, "A fast vector quantization algorithm preserving color image quality," in *SPIE, Very High Resolution and Image Quality III*, (San-Jose, California), Jan. 1998.
- [36] C. Charrier and H. Cherifi, "Subjective measure of edge degradation for vector quantized color images," in *EUSIPCO. Image Quality Evaluation*, vol. II, (Island of Rhodes, Greece), pp. 1052–1056, Sept. 1998.

- [37] C. Charrier, H. Cherifi, and B. Laget, "Codebook generation using the SEM algorithm," in *IASTED International Conference on Signal and Image Processing*, (New Orleans, Louisiana), pp. 251–256, Dec. 1997.
- [38] C. Charrier, K. Knoblauch, and H. Cherifi, "VQ-coded image quality optimized by color space selection," in *Picture Coding Symposium*, vol. 143, (Berlin, Germany), pp. 195–200, Sept. 1997.
- [39] C. Charrier, K. Knoblauch, and H. Cherifi, "Perceptual distortion analysis of color image VQ-based coding," in *SPIE, Very High Resolution and Image Quality II*, vol. 3025, (San-Jose, California), pp. 134–143, Feb. 1997.
- [40] C. Charrier and H. Cherifi, "Minimization of color image entropy through colorimetric transformation," in *Seventh IEEE Digital Signal Processing Workshop*, (Loen, Norway), pp. 17–20, Sept. 1996.

Article de conférences nationales avec actes et comité de lecture

- [1] C. Charrier, K. Knoblauch, L. T. Maloney, and A. C. Bovik, "Calibration de la métrique MS-SSIM pour les distorsions de compression à l'aide d'une échelle perceptive des différences.," in *GRETSI*, (Bordeaux, France), 2011.
- [2] M. El-Abed, B. Hemery, C. Charrier, and C. Rosenberger, "Un modèle SVM pour l'évaluation de la qualité des données biométriques.," in *GRETSI*, (Bordeaux, France), 2011.
- [3] C. Charrier, O. Lezoray, and G. Lebrun, "Mesure de qualité des images par combinaison de classifieurs," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Dijon, France), p. 22è colloque GRETSI Traitement du Signal et des Images, Sept. 2009.
- [4] G. Lebrun, C. Charrier, and O. Lezoray, "Schéma d'apprentissage basé sur la sélection automatique d'exemples pertinents," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Montpellier, France), Nov. 2007.
- [5] C. Charrier and O. Lezoray, "Segmentation d'images couleur par la théorie de l'evidence," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Rennes, France), pp. 291–296, Nov. 2005.
- [6] G. Lebrun, C. Charrier, O. Lezoray, C. Meurie, and H. Cardot, "Classification rapide de pixels par SVM dans un espace couleur hybride," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Rennes, France), pp. 301–306, Nov. 2005.
- [7] G. Lebrun, C. Charrier, O. Lezoray, and H. Cardot, "Construction de fonctions de décision performantes et de complexités réduites avec des SVM," in *Rencontres Jeunes Chercheurs en Intelligence Artificielle*, (Nice, France), pp. 125–128, Sept. 2005.

- [8] C. Charrier, G. Lebrun, and O. Lezoray, "Fusion de segmentations d'images microscopique par SVM dans différents espaces couleur selon la théorie de l'évidence," in *20è colloque GRETSI Traitement du Signal et des Images*, (Louvain-La-Neuve, Belgique), pp. 231–234, Sept. 2005.
- [9] G. Lebrun, C. Charrier, and O. Lezoray, "Réduction du temps d'apprentissage des SVM par quantification vectorielle," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Lille, France), pp. 223–226, May 2004.
- [10] C. Charrier and T. Eude, "Une métrique psychovisuelle d'évaluation de la qualité des images couleur intégrant les effets de masquage inter-canal," in *19è colloque GRETSI Traitement du Signal et des Images*, vol. 2, (Paris, France), pp. 131–135, Sept. 2003.
- [11] C. Charrier and A.-L. Joussetme, "Quantification vectorielle d'images couleur par la théorie de l'évidence," in *COmpression et REprésentation des Signaux Audiovisuels (CORESA)*, (Lyon, France), pp. 227–230, Jan. 2003.
- [12] C. Charrier and H. Cherifi, "Détection de contours couleur par coopération de modèles statistiques et de mesures psychophysiques de perception," in *GRETSI'01*, (CNES, Toulouse, France), Sept. 2001.
- [13] C. Charrier, H. Cherifi, and G. Y. Delisle, "Fonction psychométrique pour la mesure de la qualité des images couleur lors de transformations colorimétriques," in *CORESA 2000*, (Poitiers, France), pp. 36–43, 19-20 Oct. 2000.
- [14] C. Charrier and H. Cherifi, "Construction du dictionnaire par coopération de modèles statistiques et perceptuels pour la quantification vectorielle d'images couleur," in *GRETSI'99*, vol. 2, (Vannes), Sept. 1999.
- [15] C. Charrier and H. Cherifi, "Évaluation de métriques de qualité couleur pour la compression par quantification vectorielle," in *CORESA'99*, (Sophia Antipolis), pp. 81–88, 14-15 juin 1999.
- [16] C. Charrier and H. Cherifi, "Influence de la dégradation des contours sur la perception de la qualité des images couleur compressées," in *CORESA'98*, (Lannion), pp. 92–98, 9-10 juin 1998.

Séminaire

- [1] C. Charrier, "De la représentation numérique des couleurs." Actes de l'Ecole d'Hiver sur l'Image Numérique Couleur, Jan. 2009.
- [2] C. Charrier, "Les protocoles d'évaluation subjective de la qualité des images." Actes de l'Ecole d'Hiver sur l'Image Numérique Couleur, Jan. 2007.
- [3] C. Charrier, "évaluations qualitatives et quantitatives de la qualité d'une image couleur." Actes de l'Ecole d'Hiver sur l'Image Numérique Couleur, Jan. 2005.

Rapports

- [1] C. Charrier and G. Y. Delisle, "Modélisation of near-field radar cross section of air targets for proximity fuzing," tech. rep., Centre de Recherche pour la Défense, Valcartier, Québec, Canada, Nov. 2003. 49 pages.

Table des figures

1.1	Exemple de test comparatif n-naire. L'observateur doit classer la qualité des images selon un ordre croissant de qualité.	6
1.2	Exemple de test de mesure absolue de la qualité. L'observateur doit noter la qualité de l'image présentée en faisant varier le curseur et valide son choix en cliquant sur le bouton.	6
1.3	Schéma conducteur utilisé lors du développement d'une métrique de qualité avec référence.	11
2.1	Schéma des différentes couches de cellules constituant la rétine.	20
2.2	Modification d'un histogramme des coefficients DCT dans le cas d'une image originale non dégradée (a) relativement à celui de la même image originale dégradée par flou gaussien (b).	23
2.3	Synopsis de la métrique de qualité sans référence agnostique basée sur une extraction de caractéristiques statistiques des scènes naturelles.	24
2.4	Exemple des cartes de dégradations obtenues sur trois images tests (a) dont une est exempte de dégradation (à gauche). L'image du centre a été dégradée par application d'un bruit blanc tandis que l'image de droite a subi une compression JPEG à 0.15312 bpp. Les sous-figures (b) correspondent à la carte d'erreur du Kurtosis tandis que les sous-figures (c) correspondent à la carte d'erreur du contraste. Les sous-figures (d) présentent les cartes de distorsion liées au maximum d'entropie, alors que les sous-figures (e) présentent les cartes de distorsion liées à la variance d'entropie.	29
2.5	Scores obtenus avec la méthode de mesure de qualité sans référence G-BLIINDS et L-BLIINDS ainsi qu'avec le PSNR. Les images sont extraites de la base TID2008. De gauche à droite les dégradations subies sont un bruit haute-fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit du à une insertion de blocs de couleur et d'intensité différente. les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellent)	33
2.6	Synopsis de la méthode d'évaluation de la qualité des données biométriques	34
2.7	Classes prédites et réelles par la méthode développée pour une image de visage et une image d'empreinte digitale originales ayant subies une dégradation par bruit gaussien, une opération de redimensionnement de l'image et un flou gaussien. Les classes sont définies dans le tableau 2.3	37

2.8	Valeurs de l'EER (en %) sur chacune des bases testées permettant de mesurer l'effet des dégradations sur la performance globale du système biométrique utilisé.	38
2.9	Synopsis de la métrique de qualité sans référence agnostique G-BLIINDS2 basée sur une modélisation statistique des scènes naturelles.	40
2.10	Illustration de l'évolution des distributions d'une gaussienne généralisée 2D en fonction des valeurs du paramètre γ .	41
2.11	Disposition des coefficients DCT d'un bloc de taille 7×7 en sous-bandes fréquentielles (a) et sous-bandes orientées (b).	43
2.12	Exemple des cartes de dégradations obtenues sur trois images tests (a) dont une est exempte de dégradation (à gauche). L'image du centre a été dégradée par application d'un bruit blanc tandis que l'image de droite a subi une compression JPEG à 0.15312 bpp. Les sous-figures (b) correspondent à la carte d'erreur du coefficient γ tandis que les sous-figures (c) correspondent à la carte d'erreur du coefficient ζ . Les sous-figures (d) présentent les cartes de distorsion liées au coefficient ρ , alors que les sous-figures (e) présentent les cartes de distorsion liées au coefficient ξ .	45
2.13	Scores obtenus avec la méthode de mesure de qualité sans référence G-BLIINDS et BIQI ainsi qu'avec l'indice MS-SSIM et le PSNR, toutes deux nécessitant une image de référence. Les images sont extraites de la base TID2008. De gauche à droite les dégradations subies sont un bruit haute-fréquence, une compression JPEG, une erreur de transmission JPEG2000 et un bruit du à une insertion de blocs de couleur et d'intensité différente. les valeurs MOS sont calculées sur une échelle de 0 (très mauvaise qualité) à 9 (qualité excellente).	49
3.1	Illustration des limitations de l'utilisation de la métrique de Minkowski.	52
3.2	Synopsis de la méthode d'évaluation de la qualité basée sur une approche d'apprentissage statistique, nommée MLIQM— <i>Machine Learning based Image Quality Metric</i> .	53
3.3	Synopsis du schéma proposé où p_i représente la probabilité <i>a posteriori</i> calculée dans l'espace couleur i .	68
3.4	Pixels incohérents obtenus par intersections des diverses cartes de segmentation en fonction d'une image originale.	71
3.5	Manifolds SSIM obtenus pour les trois images tests. Chaque point dans l'espace (l, c, s) correspond à une distorsion spécifique (α, β) . Ces manifolds correspondent à une combinaison de bruit gaussien et de flou gaussien (d), de bruit gaussien et de flou moyen (e) et de bruit poivre et sel et de flou gaussien (f).	74
3.6	Synopsis de la procédure d'estimation des distorsions.	76

3.7	Fonction d'erreur $E(\cdot)$ pour l'image <i>bikes</i> avec et sans utilisation d'un patch bicubique. Le minimum est à $(\frac{\alpha_{MAX}}{2}, \frac{\beta_{MAX}}{2}) = (10, 0.005)$	77
3.8	Manifold SSIM associé à l'image <i>cemetery</i> . Ce manifold peut être vu comme un patch paramétrique sur lequel chaque point est défini par un couple (α, β)	78
3.9	Comparaison de la performance et de la précision des algorithmes de recherche utilisés lors de l'estimation de bruit gaussien et de flou gaussien. Les graphiques supérieurs et intermédiaires présentent la moyenne et l'écart-type des erreurs d'estimation de la quantité de bruit et de flou. Les graphiques inférieurs présentent le temps moyen d'exécution des méthode Matlab. Les méthodes testées sont les suivantes : BF – Estimation par force brute; SI – Estimation par simplex; NR – Estimation par l'algorithme de Newton-Raphson; NU – Estimation par l'algorithme newUOA; PBF – Estimation par force brute avec patch; PSI – Estimation par simplex avec patch; RPSI – Estimation selon le raffinement du patch avec simplex; HMF – Estimation par utilisation d'un filtre homomorphique; PSF – Estimation par utilisation de la fonction d'étalement du point; VAR – Estimation par calcul de la variance du bruit; EVAR – estimation par l'algorithme de Buckley.	82
3.10	Distribution de l'erreur des estimations sur l'ensemble de l'espace de recherche pour l'image <i>bikes</i> en utilisant l'algorithme de recherche par force brute. L'estimation a été calculée sur un treillis régulier de taille 15×15 sur lequel la dynamique des valeurs est $(0, 0)$ à $(\alpha_{MAX}, \beta_{MAX})$	83
3.11	Exemple de résultats obtenus avec la méthode de de recherche brute utilisant des patches. Soit une image originale I (a) et ses versions dégradées présentées à gauche pour (b) et (c) avec les paramètres $(\alpha^* = 16, \beta^* = 7)$ pour une dégradation par bruit gaussien combinée à un flou gaussien et $(\alpha^* = 13, \beta^* = 10\%)$ pour une dégradation par bruit poivre & sel combinée à un flou par filtre moyenneur, respectivement. Les images de droites (dans (b) et (c)) présentent les versions dégradées de l'image I en se basant sur les paramètres estimés $(\hat{\alpha}, \hat{\beta})$	84
4.1	Échantillon des scores prédits pour les métriques algorithmique d'évaluation de la qualité q_1 et q_2	88
4.2	Exemple d'une épreuve présentée lors du test psychophysique associée à la technique MLDS. Les stimuli sont arrangés en deux paires d'images (i, j) et (k, l) . Pour chacune des paires affichées, le stimulus affiché à droite correspond à l'image la moins dégradée. L'observateur humain doit alors décider si la variation de qualité entre les stimuli i et j est plus grande ou plus petite que la variation de qualité entre les stimuli k et l . Dans cet exemple, la majorité des observateurs devraient juger que la paire affichée en haut exhibe la plus grande variation de qualité.	92

4.3	Résultats obtenus pour chacune des images de la base. Les points noirs représentent les valeurs MLDS et la courbe noire correspond aux scores MS-SSIM.	93
4.4	Résultats obtenus pour chacune des images de la base. Les points noirs représentent les valeurs MLDS et la courbe noire correspond à chacun des trois facteurs utilisés dans le calcul de l'indice MS-SSIM.	94
4.5	Évolution de la valeur de l'attribut de structure à divers niveaux de décomposition pour cinq images (<i>imga</i> à <i>imge</i>). Les points noirs représentent les valeurs MLDS et la courbe noire correspond au facteur de distorsion des structures utilisés dans le calcul de l'indice MS-SSIM.	95
4.6	Résultats obtenus pour les images testées. Les points noirs représentent les valeurs MLDS, la courbe continue noire les valeurs MS-SSIM calculées avec les exposants originaux, et la courbe noire pointillée fait référence aux valeurs MS-SSIM calculées avec les nouveaux exposants du tableau 4.1.	98
A.1	Étapes d'un filtrage homomorphique.	117

Liste des tableaux

1.1	Notes utilisées permettant de quantifier la qualité (à gauche) ou la dégradation (à droite) d'une image selon l'échelle à cinq valeurs définie par l'ITU dans sa recommandation ITU-R BT500.	5
1.2	Caractéristiques des principales bases d'images disponibles.	14
2.1	Coefficients de corrélation de la métrique appliquée sur la base test LIVE pour divers niveaux de résolution.	31
2.2	Corrélation de Spearman (SROCC) entre les scores prédits par la méthode proposée selon un modèle gaussien (G-BLIINDS), un modèle laplacien (L-BLIINDS) et la mesure du PSNR sur la base de test.	32
2.3	Catégories de qualité retenues dans le cadre de l'évaluation de la qualité des données biométriques.	35
2.4	Valeurs du coefficient LCC calculé entre les cinq attributs et les altérations sur les bases de visages. Les valeurs mises en valeur correspondent à de fortes corrélations.	36
2.5	Précision des modèles SVM multiclassés sur les ensembles de test $(\text{TestS}_i)_{1 \leq i \leq 5}$ et sur l'ensemble $(\text{TestS})_{\text{all}}$. La base DB_2 contient des images d'empreintes digitales tandis que les quatre premières contiennent des images de visage.	37
2.6	Comparaison entre la méthode d'évaluation de la qualité proposée et la métrique NFIQ avec des valeurs du test de KS pour un intervalle de confiance égal à 95%	39
2.7	Coefficients de corrélation LCC et SROCC entre les scores prédits par chacun des paramètres développés et les valeurs DMOS subjectives pour la base de test issue de la base d'image LIVE.	47
2.8	Coefficients de corrélation calculés entre les scores prédits par les métriques G-BLIINDS2, BIQI, MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test LIVE.	47
2.9	Coefficients de corrélation calculés entre les scores prédits par les métriques G-BLIINDS2, BIQI, MS-SSIM et le PSNR et les scores obtenus par jugement humain sur la base test TID2008.	48
3.1	Transformation d'un problème de classification à cinq classes en 10 problèmes binaires obtenus lors de l'utilisation d'une approche UCU.	56

3.2	Transformation d'un problème de classification à cinq classes en quatre problèmes binaires obtenus lors de l'utilisation du schéma de décomposition OR.	57
3.3	Mesures de cohérence calculées sur la base de test.	58
3.4	Taux de reconnaissance pour la méthode UCU-VM (avec Vote Majoritaire) et pour la méthode UCU-DS (avec règle de combinaison de Dempster-Shafer).	61
3.5	Coefficients de corrélation obtenus entre les valeurs prédites par les quatre méthodes d'évaluation de la qualité standard et par la méthode développée et les valeurs subjectives DMOS en considérant la base de test TestC1.	64
3.6	Coefficients de corrélation linéaire (LCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.	65
3.7	Coefficients de corrélation de rang de Kendall (KROCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.	66
3.8	Coefficients de corrélation de rang de Spearman (SROCC) obtenus entre les méthodes d'évaluation de la qualité et les scores subjectifs de qualité DMOS en considérant la base de test TID2008.	66
3.9	Taux de reconnaissance moyen des pixels par classe.	72
3.10	Pourcentage des pixels classifiés correctement, incorrectement selon les vérités terrain.	72
3.11	Moyennes du taux de reconnaissance et de qualité selon le critère de Martin pour la segmentation par les SVMs, la combinaison des SVMs intégrant la distance euclidienne (d_1) et la distance de Mahalanobis (d_2).	72
4.1	Valeurs des 15 exposants et les intervalles de confiance (IC) à 95% utilisant une optimisation par algorithme génétique.	97
4.2	Coefficients de corrélation calculés sur la base d'images LIVE pour les valeurs originales des exposants utilisées par l'indice de qualité MS-SSIM et les nouvelles valeurs.	99
4.3	Coefficients de corrélation calculés sur la base d'images TID2008 pour les valeurs originales des exposants utilisées par l'indice de qualité MS-SSIM et les nouvelles valeurs.	100

Résumé :

L'ensemble des travaux présentés dans ce manuscrit concerne principalement la mesure de la qualité des images avec ou sans référence.

Dans une première partie, deux méthodes d'évaluation agnostique de la qualité des images sans référence ont été développées. L'idée inhérente à l'approche retenue réfère au principe des statistiques de scène naturelles (SSN). Le point commun des deux méthodes développées porte sur l'exploitation des statistiques calculées dans le domaine fréquentiel de la TCD, qui constitue notre cadre de travail. Dans un premier temps, le schéma développé intègre des caractéristiques « classiques » de distorsion induites par des propriétés du SVH telles que la sensibilité aux orientations des structures, aux variations de la netteté des structures des images lors de la vision fovéale et aux variations de contraste. Dans une seconde approche, l'idée appliquée a été de dériver un modèle généralisé des SSN utilisant les coefficients DCT locaux, et d'utiliser les paramètres du modèle comme autant de facteurs que nécessaire. Ces derniers sont utilisés pour prédire la qualité des images.

Dans une deuxième partie, une métrique d'évaluation de la qualité lorsque l'on dispose d'une image de référence a été développée en utilisant un schéma de classification multi-classes. Un processus de décomposition de type un-contre-un a été déployé. Un processus de combinaison sous raisonnement incertain est appliqué afin de prendre en compte les incertitudes de chacun des classificateurs binaires. Après classification, une fonction de régression par SVM est générée pour chacune des classes de qualité. Ceci permet d'optimiser la fonction associée à chacune des classes et de maximiser la corrélation des scores prédits avec les notes de qualité associées aux observateurs humains. Dans une extension de ces travaux, plusieurs attributs de qualité ont été détournés de leur utilisation classique afin de définir un processus d'estimation du niveau et du type de distorsion dans le cas de multi-dégradation. La méthode développée se focalise sur la relation bijective entre les paramètres de dégradation et le point de l'espace des attributs de qualité.

Finalement, l'amélioration des niveaux de corrélation des méthodes algorithmiques d'évaluation de la qualité avec les scores donnés par les observateurs humains a été étudiée. En utilisant les échelles de différence, l'observateur doit juger les différences perceptuelles entre deux couples d'images. Cette technique a été appliquée dans le cas des images dégradées selon le schéma de compression JPEG2000. Elle a permis ainsi d'évaluer l'impact du schéma de compression test sur la précision de la prédiction de l'indice MS-SSIM vis-à-vis des jugements humains et d'ajuster les valeurs de pondération des trois facteurs de distorsion intégrées dans l'indice MS-SSIM.