

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Mohamed Khider – BISKRA

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

Département d'informatique

N° d'ordre :



THÈSE

Présentée en vue de l'obtention
du diplôme de Doctorat en Sciences

Option : Informatique

Titre

**Métaheuristiques pour l'évaluation objective de
la qualité d'images et de vidéos**

Par

Naima MERZOUGUI

Soutenue publiquement, le 23 /06/2022, devant le jury composé de :

Mohamed Chaouki BABAHENINI	Professeur	Université de Biskra	Président
Leila DJEROU	Professeur	Université de Biskra	Rapporteur
Abdelhamid DJEFFAL	Professeur	Université de Biskra	Examineur
Ali BEHLOUL	Professeur	Université de Batna 2	Examineur
Allaoua HEMMAK	Maitre de Conférences A	Université de Msila	Examineur
Amine KHALDI	Maitre de Conférences A	Université d'Ouargla	Examineur

Année universitaire : **2021-2022**

Dédicaces

Je dédie ce travail accompagné d'un profond amour :

A la mémoire de mon père, c'est grâce à sa sagesse et la confiance qu'il m'a accordé que j'ai pu tracer mon chemin dans cette vie.

Ma mère, pour son amour, ses encouragements et ses sacrifices.

A mon époux, pour son continuel soutien et ses encouragements qui m'ont toujours aidé pour avancer.

A mes petites filles "Oumaima" et "Sarra".

A ma sœur Ghalia et son époux Mustafa qui m'a apporté leurs conseils et leurs soutiens.

A mes sœurs et mes frères, merci pour vos encouragements et votre aide incontournable.

A toute la famille.

Remerciements

Tout d'abord, je remercie ALLAH, qui m'a éclairé le chemin et m'a aidé de mener à bien ce travail.

Je tiens à remercier grandement mon rapporteur Pr.Djerou Leila de m'avoir proposé un tel intéressant sujet, m'ouvrant ainsi les portes sur un domaine de recherche assez vivant, et je tiens tout particulièrement à la remercier pour son aide et ses précieux conseils tout au long de ces années.

Que les membres du jury : Pr.BABAHENINI, Pr.DJEFFAL, Pr.BEHLOUL, Dr.HEMMAK et Dr.KHALDI trouvent ici mes vifs remerciements d'avoir accepté évaluer ce travail et pour le temps qu'ils ont consacré pour la lecture du mémoire.

J'aimerais remercier tout particulièrement mon mari Dr.Mhamed, ma soeur Dr.Ghalia et son époux Pr.Mostafa qu'ils étaient avec moi tout en long de ces années de travail, avec ses discussions, ses conseils, ainsi que ses soutiens moraux.

Finalement, mes remerciements à tous ceux qui m'ont encouragé.

Résumé

Le travail présenté dans cette thèse s'inscrit dans le domaine d'évaluation objective de la qualité d'images et de vidéos et plus particulièrement sur l'évaluation de qualité avec référence complète. Il s'agit d'un domaine de recherche actif qui s'est imposé avec l'évolution rapide des technologies de l'informatique, des communications et des réseaux. Malgré la richesse des travaux dans ce domaine, et l'abondance des méthodes objectives d'évaluation (les métriques), le problème d'évaluation de qualité d'image reste toujours posé. En effet, les métriques présentées sont généralement spécifiques à un type particulier de dégradation, sensibles aux données bruitées par plusieurs types de dégradations, d'où l'idée de s'orienter vers de nouvelles approches recommandant la fusion de différentes métriques pour quantifier la qualité de l'image en se basant sur les techniques d'apprentissage et les métaheuristiques d'optimisation.

Dans cette thèse, nous avons développé de nouvelles méthodes basées sur la programmation génétique multigène (MGGP) comme une technique d'apprentissage évolutive inspirée de l'évolution naturelle, permettant, à partir d'un ensemble d'apprentissage, de trouver de bons modèles de combinaison de métriques de qualité d'image, en optimisant simultanément deux objectifs concurrents : l'adéquation de modèles par rapport à leur corrélation avec les scores subjectifs de divers types de distorsions dans les images, et la complexité de leur structure. En raison de la croissance explosive des applications orientées écran, nous avons exploité les avantages offerts par MGGP afin de développer une méthode générique pour l'évaluation objective des images de contenu d'écran. Les résultats d'expérimentation menée sur les grands benchmarks, ont montré les performances supérieures des méthodes proposées par rapport aux mesures de pointe, y compris d'autres approches de fusion récemment publiées.

Mots clés : Métriques d'évaluation objective, Métaheuristiques d'optimisation, Optimisation multiobjective, Qualité visuelle.

Abstract

The work presented in this thesis, fits into the field of objective assessment of the quality of images and videos and more specifically on the assessment of quality with full reference. This is an active research area that has established itself with the rapid development of computer, communications and networking technologies. Despite the wealth of work in this field, and the abundance of objective evaluation methods (metrics), the problem of image quality evaluation remains. Indeed, the metrics presented are generally specific to a particular type of degradation, sensitive to data noisy by several types of degradation; hence the idea of moving towards new approaches calling on the merging of different metrics to quantify the quality image based on learning techniques and optimization meta-heuristics.

In this thesis, we have developed new methods based on multigene genetic programming (MGGP) as an evolutionary learning technique inspired by natural evolution, allowing, from a learning set, to find good models combination of image quality metrics, simultaneously optimizing two competing objectives: the suitability of models for their correlation with the subjective scores of various types of distortions in the images, and the complexity of their structure. Due to the explosive growth of screen-oriented applications, we have exploited the advantages offered by MGGP to develop a generic method for the objective evaluation of images of screen content. The results of experiments carried out on the large benchmarks, showed the superior performance of the proposed methods compared to advanced measurements, including other recently published fusion approaches.

Keywords: Objective evaluation metrics, Optimization metaheuristics, Multiobjective optimization, Visual quality.

الملخص

يقع العمل المقدم في هذه الأطروحة ضمن مجال التقييم الموضوعي لجودة الصور ومقاطع الفيديو وبشكل أكثر تحديداً على تقييم الجودة مع المرجع الكامل. هذا مجال بحث نشط أثبت نفسه مع التطور السريع لتقنيات الكمبيوتر والاتصالات والشبكات. على الرغم من كثرة العمل في هذا المجال، ووفرة طرق التقييم الموضوعية (المقاييس)، تظل مشكلة تقييم جودة الصورة قائمة. في الواقع، المقاييس المقدمة الخاصة بشكل عام بنوع معين من التدهور، حساسة للبيانات المزعجة من عدة أنواع من التدهور، ومن هنا جاءت فكرة الانتقال نحو مناهج جديدة تدعو إلى دمج مقاييس مختلفة لتحديد جودة الصورة تستند إلى تقنيات التعلم والاستدلال الأمثل.

في هذه الأطروحة، قمنا بتطوير طرق جديدة تعتمد على البرمجة الحينية متعددة الحينات (MMGP) كأسلوب تعلم تطوري مستوحى من التطور الطبيعي، مما يسمح، من مجموعة التعلم، بالعثور على نماذج جيدة تتمثل في مزيج من مقاييس جودة الصورة، وفي نفس الوقت تحسين هدفين متنافسين: ملاءمة النماذج لعلاقتها بالدرجات الذاتية لأنواع مختلفة من التشوهات في الصور، ومدى تعقيد بنيتها. نظرًا للنمو الهائل هذه الآلات الموجهة نحو الشاشة، فقد استغلنا المزايا التي توفرها MMGP لتطوير طريقة عامة للتقييم الموضوعي لصور محتوى الشاشة. أظهرت النتائج التجريبية، التي أجريت على معايير كبيرة، الأداء المتفوق للطرق المقترحة مقارنة بالقياسات المتقدمة، بما في ذلك طرق الاندماج الأخرى المنشورة مؤخرًا.

الكلمات الرئيسية: مقاييس التقييم الموضوعي، خوارزميات التحسين، التحسين متعدد الأهداف، الجودة المرئية.

Articles et conférences

- N. Merzougui, L.Djerou, “Multi-measures fusion based on multi-objective genetic programming for full-reference image quality assessment”, In the fifth edition of the International Conference on the Image and Signal Processing and their Applications (ISPA'17) (2017).
- N.Merzougui, L.Djerou, Multi-gene Genetic Programming based Predictive Models for Full-reference Image Quality Assessment, Journal of Imaging Science and Technology. 65 (2021) 60409-1-60409–13.

Liste des matières

Dédicaces	1
Remerciements	2
Résumé	3
Abstract	4
الملخص	5
Articles et conférences	6
Liste des matières	7
Liste des Figures	10
Liste des Tableaux	11
Liste des abréviations	12
Introduction générale	15
1 Contexte de la thèse	15
2 Objectifs de la Thèse	16
3 Contributions	17
4 Organisation de la thèse	17
1 Qualité d'image et de vidéo	19
1.1 Introduction	19
1.2 Qualité d'image	19
1.2.1 Évaluation de la qualité d'image	21
1.2.2 Méthodes existantes pour l'évaluation de la qualité	22
1) Évaluation subjective de la qualité	22
2) Évaluation objective de la qualité d'image	23
1.2.3 Survey des métriques de qualité d'image existantes	24
1) Méthodes Full Référence classique	24

2)	Méthodes par fusion des caractéristiques	27
3)	Méthodes par fusion de métriques	28
1.2.4	Les bases d'images populaires pour l'IQA	31
1.3	Qualité de la vidéo	34
1.3.1	Survey des métriques VQA	36
1.3.2	Les bases populaires de vidéos pour VQA	38
1.4	Domaines d'application d'I/VQA	39
1.5	Conclusion	40
2	Optimisation et Métaheuristique	41
2.1	Introduction	41
2.2	Problème d'optimisation	41
2.3	Vocabulaire et définitions	43
2.4	Algorithmes d'optimisation	45
2.5	Classification des métaheuristiques	46
2.5.1	Métaheuristique à base de solution unique	47
1)	Les méthodes de la descente (Hill Climbing)	47
2)	Le Recuit Simulé	48
3)	La méthode GRASP	50
4)	La Recherche Tabou	51
2.5.2	Métaheuristiques basée sur la population	52
1)	Les algorithmes évolutionnaires	52
2)	Les algorithmes en essaim	55
2.5.3	Les algorithmes immunitaires artificiels	57
2.6	Comparaison des métaheuristiques	58
2.7	Conclusion	60
3	Programmation Génétique MultiGène	62
3.1	Introduction	62
3.2	Régression symbolique avec la Programmation génétique	62
3.3	Programmation Génétique	63
3.3.1	Population initiale	64
3.3.2	Reproduction	65
3.3.3	Croisement	65
3.3.4	Mutation	65
3.4	Programmation Génétique Multigène (MGGP)	67
3.4.1	Croisement de haut niveau	70
3.4.2	Croisement de bas niveau	70
3.5	Application de la programmation génétique en traitement d'images	71
3.6	Conclusion	72

4	MGGP pour l'évaluation de qualité d'image	73
4.1	Introduction	73
4.2	Principe de l'approche	73
4.2.1	Ensemble de données d'entraînement	75
4.2.2	Solutions	76
4.2.3	Les Fonctions Objective	77
4.2.4	Paramètres MGGP	79
4.2.5	Sélection du modèle prédictif	80
4.3	Évaluation expérimentale	82
4.3.1	Évaluation des performances de prédiction	82
4.3.2	Évaluation de la signification statistique	87
4.3.3	Comparaisons des performances des mesures obtenues avec les mesures de fusion IQA	88
4.3.4	Test croisé de bases de données	89
4.3.5	Influence des métriques IQA sur la fusion	90
4.4	Conclusion	91
5	Evaluation de la qualité des images du contenu de l'écran	93
5.1	Introduction	93
5.2	Image de contenu d'écran SCI	93
5.3	Bases d'image de contenu d'écran	94
5.4	Etat de l'art sur les métriques d'IQA des SCIs	96
5.5	La programmation génétique et les images de contenu d'écran (GP-SCIQ)	97
5.6	Évaluation expérimentale	98
5.6.1	Évaluation des performances de prédiction	98
5.6.2	Comparaisons de performances avec des méthodes SCI sur SIQAD	101
5.6.3	Évaluation croisée	101
5.6.4	Importance statistique	102
5.7	Conclusion	103
	Conclusion Générale	104
	Références Bibliographiques	106

Liste des Figures

Figure 1.1:	Exemple de dégradation d'une image par différents types de distorsions	21
Figure 1.2:	Les images de référence des différentes bases	33
Figure 2.1:	Front de Pareto pour un problème de minimisation de deux objectifs	43
Figure 2.2:	Différence entre un optimum global et des optima locaux	44
Figure 3.1:	Une formule mathématique représentée sous forme d'arbre	63
Figure 3.2:	Une opération de croisement en GP	66
Figure 3.3:	Une opération de mutation en GP	66
Figure 3.4:	Exemple typique du modèle MGGP	68
Figure 3.5:	Un organigramme typique pour une procédure de programmation génétique multi-gène	69
Figure 4.1:	Organigramme de l'approche proposée	74
Figure 4.2:	Structure simple dans l'ensemble de formation	76
Figure 4.3:	Population de modèle évoluée, en termes de performances et de complexité du modèle, pour les cinq ensembles d'entraînement (cas : LIVE)	79
Figure 4.4:	Corrélations des métriques objectives obtenues et les trois meilleures mesures IQA de pointe avec les valeurs subjectives pour chaque ensemble de données (Chaque point de données représente une image de test)	85
Figure 4.5:	Fréquence d'entrée dans les modèles Pareto pour chaque sous-groupe	91
Figure 5.1:	Images de référence de la base SCID	95
Figure 5.2:	Images de référence de la base SIQAD	96
Figure 5.3:	Corrélations entre l'évaluation subjective et le GPSCIQ obtenu et les sept métriques IQA de pointe pour l'ensemble des images de SIQAD	99

Liste des Tableaux

Tableau 1.1:	Caractéristiques des bases d’images	34
Tableau 1.2:	Caractéristiques des bases VQA	39
Tableau 2.1:	Comparaison des métaheuristiques basées sur la population . . .	59
Tableau 4.1:	Les métriques FR-IQA utilisées	76
Tableau 4.2:	Paramètres de l’algorithme MGGP	80
Tableau 4.3:	Comparaison des performances de l’approche proposée avec les dix meilleurs modèles d’IQA	86
Tableau 4.4:	Evaluation de la signification statistique des base (a) LIVE, (b) CSIQ, (c) TID2008, (d) TID2013, (e) IVC and (f) MDID	87
Tableau 4.5:	Comparaison des mesures obtenues avec d’autres mesures de fusion IQA basées sur les valeurs SRCC	89
Tableau 4.6:	Performances SRCC de FR-MGGP entre bases de données . . .	90
Tableau 5.1:	Comparaison des performances du modèle proposé avec les sept métriques utilisées dans la fusion	100
Tableau 5.2:	Résultats expérimentaux de la méthode proposée et d’autres méthodes actuelles sur la base de données SIQAD	101
Tableau 5.3:	Évaluation croisée (Apprentissage sur SIQAD et test sur SCID) .	102
Tableau 5.4:	Tests de signification statistique sur SIQAD et SCID	103

Liste des abréviations

ACO	Ant Colony Optimization
AWN	Additive White Noise
BVQA	Block-level multi-metric fusion for Video Quality Assessment
CC	Contrast Change
CCA	Canonical Correlation Analysis
CNNM	Combined Neural Network Metric
COHERENSI	Chaos of Harmonics/phase in Error as a REference-based Novel Similarity Index
CS	Cuckoo search
CSF	Contrast Sensitivity Function
CSIQ	Content-Based Strategies of Image and Video Quality Assessment
DIVIINE	Distortion Identification-based Image Verity and INtegrity Evaluation
DMOS	Differential Mean Opinion Score
DOG	Difference-Of-Gaussian
FDD	First Digit Distribution
FPSO	Flexible PSO
FR	Full-Reference
FR-MGGP	Full Reference metric based Multi Gene Genetic Programming
FSIMc	Feature SIMilarity Metric Color
GA	Genetic Algorithm
GB	Gaussian Blur
GE	Grammatical Evolution
GEP	Gene Expression Programming
GFM	Gabor Feature-based Model
G-IQM	Global-IQM
GM	Gradient Magnitude
GN	Gaussian Noise
GP	Genetic Programming

GP-SCIQ	Genetic Programming for Screen Content Image Quality
GRASP	Greedy Randomized Adaptive Search Procedure
GREED	GeneRalizEd Entropic Differences
GSGP	Geometric Semantic Genetic Programming
GSM	Gaussian Scale Mixture
GSS	Gradient Similarity Score
HaarPSI	Haar waveletbased Perceptual Similarity Index
HVS	Human Visual System
IFC	Information Fidelity Criterion
IFS	Independent Feature Similarity
IGM	Internal Generative Mechanism
IQA	Image Quality Assessment
IQM-D	Image Quality Metric-Distortion
ITU	International Télécommunications Union
IVC	Image and Video Communication
IWSSIM	Information Weighted SSIM
JPEG	Joint Photographic Experts Group
KRCC	Kendall Rank order Correlation Coefficient
LAF	Locally adaptive fusion
LC	Layer segmentation-backed Coding
LCSIM	Linearly Combined Similarity Measure
LIVE	Laboratory for Image and Video Engineering
MAD	Most Apparent Distortion
MB	Motion Blur
MDID	Multiply Distorted Image Database
MGGP	Multi Gene Genetic Program
MJPEG	Motion JPEG compression
MOO	Multi-Objective Optimization
MOVIE	MOtion-based Video Integrity Evaluation
MPC	monogenic phase congruency
MSSIM	Mean Structural Similarity Index
MS-UNIQUE	Multi-model UNsupervised Image Quality Estimation
MTD	Multiple Type Distortion
NQM	Noise Quality Metric
NR	No-Reference
NSGA-II	Non-dominated Sorting Genetic Algorithm
PCC	Pearson linear Correlation Coefficient
PerSIM	Perceptual SIMilarity index
PPSO	Parallel PSO

PSNR	Peak Signal to Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
RDO	Rate-Distortion Optimization
RFSIM	Riesz-transform based Feature SIMilarity metric
RLCM	Robust Linearized Combined Metric
RMSE	Root Mean Square Error
RR	Reduced Reference
RVSIM	Riesz transform and Visual contrast sensitivity-based feature SIMilarity index
SCDM	Structural Contrast Distortion Metric
SCI	Screen Content Image
SCID	Screen Content Image Database
SCQI	Structural Contrast-Quality Index
SFF	Sparse Feature Fidelity
SIQAD	Screen Image Quality Assessment Database
SQMS	Saliency-guided Quality Measure of SCIs
SRCC	Spearman Rank order Correlation Coefficient
SR-SIM	Spectral Residual SIM
STD	Simple Type Distortion
STGP	Strongly-typed Genetic Programming
ST-VMAF	Spatio Temporal Video Multi-method Fusion
SVM	Support Vector Machines
SVQI	Structural Variation based Quality Index
SVR	Support Vector Regression
TE	Transmission Errors
TID 2008	Tampere Image Database release in 2008
VCGS	Visual saliency with Color appearance and Gradient Similarity
VIDEVAL	VIDeo quality EVALuator
VIF	Visual Information Fidelity
VMAF	Video Multi-method Fusion
VQA	Video quality Assesment
VQA-MGGP	Video Quality Assesment metric based Multi Gene Genetic Programming
VQEG	Visual Quality Experts Group
VQM	Video Quality Metric
VSI	Visual Saliency-Induced Index
VSSIM	Video Structural SIMilarity

Introduction générale

1 Contexte de la thèse

Les différents maillons de la chaîne de traitement d'images ou de vidéos sont autant de sources de distorsions pouvant altérer leur qualité visuelle. L'évaluation de la qualité d'images et de vidéos est un des besoins majeurs des concepteurs de systèmes de traitement d'images ou de vidéos pour garantir qu'ils satisfont à une qualité de service donnée, de sorte que le niveau de qualité de l'expérience des utilisateurs est acceptable (Okarma, 2019; Zhai and Min, 2020). Par conséquent, pour maintenir et contrôler la qualité de l'image et de la vidéo, il est important pour les méthodes d'évaluation d'être en mesure d'identifier et de quantifier la dégradation de la qualité de l'image et de la vidéo.

Avec l'évolution rapide des technologies de l'informatique (Chandler, 2013), des communications et des réseaux, la recherche dans le domaine d'évaluation de la qualité d'images et de vidéos a connu une richesse des travaux et une abondance des méthodes d'évaluation qui sont classées en deux catégories : les méthodes subjectives et les méthodes objectives. Les méthodes subjectives sont les plus proches de la réalité mais elles sont chronophages et demandent des conditions expérimentales de visualisation bien précises. D'où la nécessité de développer des méthodes objectives permettant d'évaluer la qualité visuelle de façon automatique. Ces méthodes objectives sont appelées des métriques de qualité visuelle qui se distinguent par la disponibilité d'une image de référence, d'où on trouve : les méthodes de référence complète (FR : Full-Reference), de référence réduite (RR : Reduced Reference) et méthodes sans référence (NR : No-Reference). Les méthodes FR nécessitent un accès complet à la l'image de référence pour évaluer la qualité de l'image (Chandler, 2013).

En réponse à l'universalité et aux bonnes performances des méthodes FR (les métriques d'évaluation), un grand nombre de différentes métriques FR ont été proposées dans la littérature. Cependant, l'évaluation de la qualité des images soumises à divers types de

distorsions reste l'un des problèmes les plus difficiles en vision par ordinateur et en analyse d'images (Okarma, 2019), car il n'y a pas une seule métrique capable de quantifier la dégradation de l'image subi par différents types de distorsions et mesurer sa qualité. Pour cette raison, de nombreuses études récentes (Khaing et al., 2019; Varga, 2020) ont adopté des nouvelles stratégies qui impliquent la fusion de diverses métriques pour quantifier la qualité de l'image. Cependant, trouver la meilleure combinaison de métriques conduisant à une meilleure corrélation avec les scores subjectifs est un problème d'optimisation qui nécessite une connaissance sur le contenu d'image, d'où l'idée de s'orienter vers les techniques d'apprentissage et les métaheuristiques d'optimisation qui ont la possibilité de traiter des données imprécises et incomplètes qui sont deux caractéristiques inhérentes à la donnée image.

La programmation génétique multigène (MGGP : Multi-Gene Genetic Programming) (Gandomi and Alavi, 2012) est une technique d'apprentissage évolutive inspirée de l'évolution naturelle permettant de trouver un ensemble de programmes informatiques (des solutions) capables de traiter un problème posé par l'utilisateur, en optimisant généralement deux types de fonctions objectives ; une dépend de la complexité de la solution et l'autre sur sa qualité dont le calcul de sa valeur se fonde sur une base d'apprentissage du problème traité.

2 Objectifs de la Thèse

Les images et les vidéos ont un rôle important dans les médias, mais leur utilisation dépend de leur qualité. Elles sont affectées par une grande variété de déformations à travers les différentes étapes de la chaîne de distribution (acquisition, stockage, transmission et livraison à l'utilisateur). Malgré l'abondance des métriques dans la littérature, il n'y a pas une seule métrique capable de quantifier la dégradation de l'image par différents types de distorsions et mesurer sa qualité. Dans ce contexte, les objectifs de cette thèse sont :

- Faire un état de l'art sur les axes de recherche du domaine d'évaluation de qualité d'image et de vidéo et aussi les méthodes d'évaluation existantes dans la littérature.
- Faire un état de l'art sur les métaheuristiques d'optimisation pour éclaircir leurs caractéristiques qui nous aident à développer des méthodes d'évaluation de qualité d'image et de vidéos soumises à divers types de distorsion.
- Étudier la possibilité de la métaheuristique MGGP pour prédire les meilleures combinaisons de métriques permettant d'évaluer la qualité des images, sans connaître leurs types de dégradation.
- Proposer des méthodes pour évaluer la qualité des images de contenu d'écran qui est un axe de recherche très actif avec le progrès récent des technologies de multimédias.

3 Contributions

Les principales contributions de cette thèse sont résumées comme suis :

- Proposition d'une approche d'évaluation objective de la qualité d'image (IQA) permettant de prédire et de mesurer la qualité d'image contenant divers types de distorsions, à partir d'un large ensemble de métriques disponibles, en utilisant la méthode de programmation génétique multigène. A partir d'un grand nombre d'IQM disponibles et sans aucune connaissance a priori sur leurs natures, la MGGP permet de trouver le nombre "optimal" d'IQM appropriées, sous une formulation explicite de leur combinaison. Comme la méthode MGGP est une technique d'apprentissage évolutive inspirée de l'évolution naturelle, elle va exploiter un vaste espace de recherche composé de toutes les combinaisons possibles de scores objectifs de ces IQM et trouver les meilleur modèles de régression décrivant la somme pondérée des scores objectifs du IQM, en terme de leur simplicité et leur performance, sur les données d'apprentissage. Six plus populaires bases de données d'images (à savoir LIVE, CSIQ, TID2008, TID2013, IVC et MDID) sont utilisées pour apprendre et tester les modèles prédictifs, selon la validation croisée k-fold et les stratégies d'ensemble de données croisées.
- Proposition d'une meilleure composition des métriques de qualité des images de contenu d'écran qui comprennent à la fois des textes et des images et même des graphiques, en exploitant aussi les possibilités offertes par la MGGP. En utilisant une base d'apprentissage contenant des images de contenu d'écran, soumises par divers types de distorsion, la MGGP permet de trouver une meilleure combinaison de métriques de qualité avec un score objectif, qui est linéairement corrélé avec la perception subjective de divers types de distorsions, dans ces images. Pour valider la robustesse et la capacité de généralisation de notre modèle de combinaison, nous avons entraîné le système sur la base SIQAD, et le testé sur la base SCID.

4 Organisation de la thèse

Après l'introduction générale, le reste de la thèse est présenté comme suit :

Dans le chapitre 1 : Nous présenterons les principaux concepts de l'évaluation de la qualité d'image et de vidéo. Où nous commencerons par la définition de qualité d'image et d'évaluation de la qualité, puis nous mettrons l'accent sur les différentes méthodes d'évaluation des deux familles (évaluation subjective et évaluation objective). Nous présenterons ensuite un état de l'art sur les métriques objectives de qualité d'images existantes ainsi que les bases d'images les plus populaires. En ce qui concerne le cas de la qualité de vidéo,

nous présenterons un Survey sur les métriques de VQA et les caractéristiques de quelques bases de VQA les plus populaires.

Dans le chapitre 2 : Une taxinomie des Métaheuristiques d'optimisation sera dressée. Nous distinguerons les deux types : les méthodes basées sur une solution unique et les méthodes basées sur une population de solutions. Nous présenterons quelques algorithmes les plus courants dans la littérature dans les deux types. Cela nous permettra de chercher les points faibles de quelques algorithmes afin que nous puissions introduire le mécanisme d'optimisation adéquat.

Dans le chapitre 3 : Nous nous intéressons à la programmation génétique et plus particulièrement sur la programmation génétique multigène qui a rencontré un succès remarquable depuis sa création, grâce à ses hautes performances. Elle présente l'avantage d'être efficace, précise et d'avoir un grand potentiel.

Dans le chapitre 4 : Il sera consacré à expliquer notre approche proposée pour mettre en évidence une nouvelle méthode d'évaluation de qualité d'image en utilisant une métaheuristique d'optimisation nommée Programmation Génétique MultiGène (MGGP). Nous exposerons les différents résultats obtenus après l'expérimentation faites sur six bases d'images, accompagnés par des comparaisons et des commentaires pour mettre en valeur notre approche proposée.

Dans le chapitre 5 : Une extension de l'approche proposée pour le cas des images de contenu d'écran est présentée dans ce chapitre, tout en exposerons les résultats obtenus ainsi que une comparaison avec d'autre métriques sur les bases spécifiées pour les images de contenu d'écran.

Nous terminerons cette thèse par une conclusion générale et quelques perspectives ouvrant la porte à d'autres travaux de recherche.

Chapitre 1

Qualité d'image et de vidéo

1.1 Introduction

L'image et la vidéo sont deux types de contenus visuels à transmettre par les systèmes de communication visuelle. Cependant, elles peuvent subir des dégradations ce qui nécessite une évaluation de la qualité de l'image et de la vidéo, et qui est considéré comme un axe de recherche très important dans les domaines du traitement d'image et de la vision par ordinateur. Comment évaluer la qualité d'une image et/ou d'un vidéo de manière automatique, précise et efficace par un algorithme objectif est l'objectif des chercheurs depuis des années.

Ce chapitre présente les bases théoriques et les techniques d'évaluation de la qualité d'image et de vidéo, ainsi qu'un état de l'art sur les métriques et les bases d'images et de vidéo les plus couramment utilisées.

1.2 Qualité d'image

La qualité d'une image fait référence à la quantité de dégradation présente dans une image. Une image de haute qualité est toujours souhaitable. Par exemple, les images prises par l'appareil photo peuvent être de qualité variable. Il peut y avoir présence de bruit ou d'autres distorsions dans une image. Les distorsions peuvent survenir en raison de l'acquisition, de la compression, du stockage et de la décompression d'une image, du mouvement de l'appareil photo lors de la capture d'image ou de l'ajout de bruit à une image. La qualité d'une image ne peut être déterminée par seulement quelques paramètres tels que la luminosité, le contraste ou la netteté. Une image nette peut contenir du bruit de sel et de poivre. Pour plus de détails sur les type de distorsions dans le système de communication d'images, les lecteurs intéressés sont renvoyés aux publications de (Hemami and Reibman, 2010; Punchihewa and Bailey, 2002).

Nous discuterons brièvement quelques distorsions les plus apparus :

- **Contraste** : Le contraste est la différence d'intensité du signal ou de l'image affiché entre deux zones d'intérêt. Une image à fort contraste présente une plus grande différence entre les nuances de gris affichées mais une plus petite gamme de gris. Une image à faible contraste présente une différence faible (c'est-à-dire qu'il est plus difficile de distinguer les différentes zones) mais une plus grande gamme de gris.
- **Bruit** : Le bruit est constitué des variations aléatoires ou structurées au sein d'une image qui ne correspondent pas aux variations d'atténuation des rayons X de l'objet. Le spectre de puissance du bruit est la meilleure métrique du bruit qui mesure le contenu en fréquence spatiale du bruit. Le bruit quantique est principalement responsable du bruit de l'image, et le nombre de quanta de rayons X utilisés pour former l'image détermine le bruit quantique.
- **Le flou gaussien** (également appelé lissage gaussien) est le résultat du flou d'une image par une fonction gaussienne. Il s'agit d'un effet largement utilisé dans les logiciels graphiques, généralement pour réduire le bruit de l'image et atténuer les détails. L'effet visuel de cette technique de flou est un flou doux ressemblant à celui de la visualisation de l'image à travers un écran clair.
- **La compression d'image (JPEG et JPEG2000)** fait partie de la compression de données. Elle consiste à lire une suite de symboles binaires représentant une image ou une séquence d'image et à les transformer en code, afin de réduire la quantité de données nécessaires pour représenter une image. L'objectif de la norme JPEG2000 est de compléter les performances de la norme JPEG (JPEG : Joint Photographic Experts Group) mais sans la remplacer. Le JPEG2000 permet un taux de compression élevé avec une très faible dégradation appréciable de la qualité de l'image. Il comporte une version sans perte et une version avec perte de la compression d'image. Le compresseur d'images JPEG2000 est basé sur les techniques de codage par ondelettes/sous-bandes et emprunte des idées au schéma de codage par blocs embarqué avec une troncature optimisée.

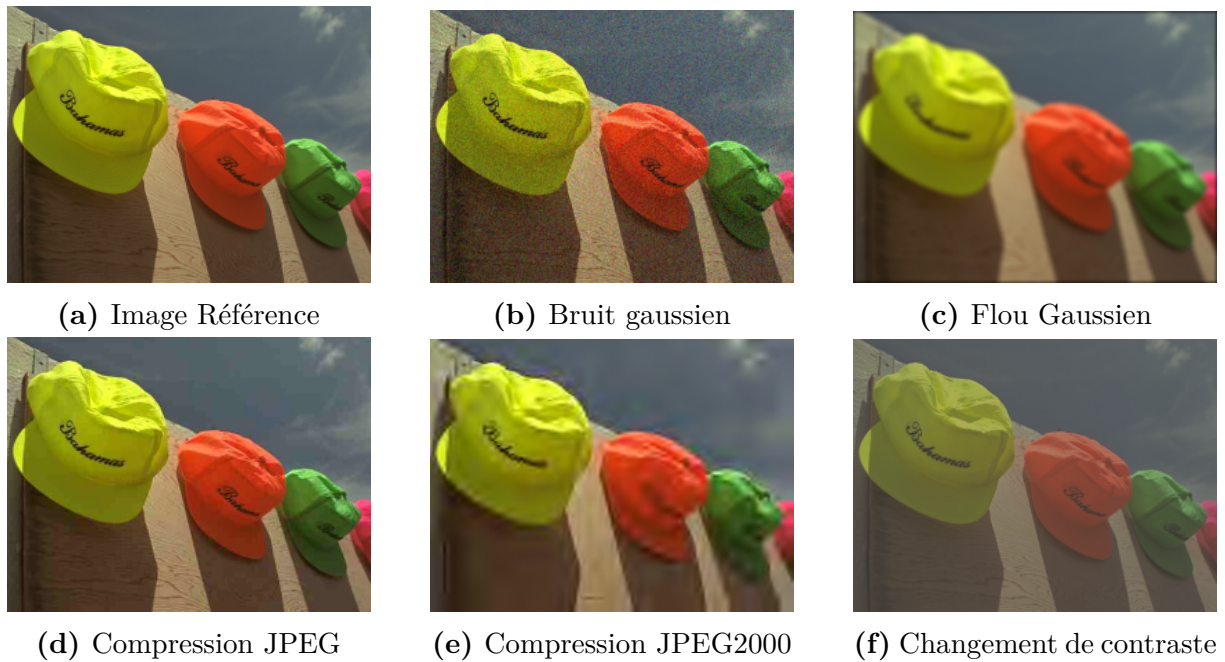


Figure 1.1. - Exemple de dégradation d'une image par différents types de distorsions

Par conséquent, une procédure standard est nécessaire pour évaluer la qualité d'une image quel que soit le type de distorsion qui l'a affectée.

1.2.1 Évaluation de la qualité d'image

L'évaluation de la qualité d'image a été une tâche ardue dans le domaine du traitement d'image sans réponse satisfaisante jusqu'à présent. L'évaluation de la qualité d'image tente de quantifier une qualité visuelle, une quantité de déformations dans une image donnée. Ces modifications sont une composante incontournable de tout traitement d'image numérique. La méthode correcte pour évaluer la qualité visuelle des images perçue par l'homme est l'évaluation par les êtres humains (évaluation subjective). Malheureusement, ce processus est fastidieux, prend du temps et ne peut être appliqué dans des applications en temps réel. Par conséquent, il existe une demande pour une technique informatisée (évaluation objective) permettant de concevoir la qualité visuelle perçue par l'homme aussi étroitement que possible.

Il est difficile de déterminer lequel de deux types de distorsion différents influence le plus la qualité de l'image, ou quel dispositif d'imagerie est le meilleur, même s'il est facile de juger subjectivement laquelle des deux images est la meilleure. Des tentatives sont faites pour construire des algorithmes qui modélisent les estimations humaines de la qualité de l'image.

Chandler dans (Chandler, 2013) énumère les problèmes rencontrés par les développeurs

de tels algorithmes dans son article. Il s'agit notamment de la variété des distorsions possibles, l'interaction entre la distorsion et le matériau, des changements géométriques et des changements dus à l'amélioration de l'image.

1.2.2 Méthodes existantes pour l'évaluation de la qualité

Comme nous l'avons vu précédemment ; il existe deux grandes catégories de méthodes pour évaluer la qualité d'image : les méthodes subjectives et les méthodes objectives. L'évaluation subjective est effectuée par des observateurs humains, tandis que l'évaluation objective ne fait pas intervenir d'observateurs. Cependant, elle implique l'utilisation des méthodes informatiques qui peuvent prédire automatiquement la qualité d'image perçue.

1) Évaluation subjective de la qualité

Les observateurs humains sont généralement reconnus comme les évaluateurs les plus précis de la qualité d'image. Différentes méthodes d'évaluation subjective ont été définies par l'ITU (BT, 2002). Il existe principalement trois grandes familles : les tests à simple stimulus permettant de juger la qualité de l'image sans aucune information relative à l'image originale. Les tests à double stimulus où l'observateur est invité à juger de la similitude entre deux images et les tests comparatifs où l'objectif principal est de comparer deux ou plusieurs stimuli. Les observateurs doivent juger la qualité sur une échelle de un à cinq, où cinq correspond à la meilleure qualité (similaire au jugement par catégorie). Il est également courant d'avoir une échelle de qualité descriptive en cinq points (très mauvaise, mauvaise, passable, bonne, excellente).

Il convient toutefois de noter que chaque méthode présente des avantages et des inconvénients, et que les utilisateurs doivent donc choisir une méthode appropriée en fonction de leur propre environnement d'application. Par exemple, la méthode à double stimulus s'avère plus stable que la méthode à stimulus unique pour évaluer les petites dégradations car les observateurs détectent plus facilement les dégradations en présence d'images de référence. En revanche, la méthode à stimulus unique présente un intérêt pratique dans les cas où aucune référence n'est disponible.

Après évaluation de la qualité par des sujets humains, un score unique - le score d'opinion moyen (MOS) - représentant la qualité perçue d'une image est obtenu en regroupant les scores subjectives individuels. Le score d'opinion moyen (MOS) est défini par l'Union internationale des télécommunications (UIT) pour la qualité audio, mais il a également été largement utilisé pour la QI (Ninassi et al., 2008; Yamsang and Udomhunsakul, 2009). Le MOS est donné par la formule suivante :

$$MOS(i) = \frac{1}{N} \sum_{j=1}^N Note_i(j) \quad (1)$$

Où N est le nombre d'observateurs et $Note_i(j)$ le score de l'observateur j pour l'image dégradée i .

Alternativement, le score final peut également être interprété comme un score d'opinion moyen différentiel (DMOS), qui représente la différence de MOS entre l'image déformée et sa référence correspondante.

Les évaluations de la qualité produites à partir d'expériences subjectives sont généralement acceptées comme étant la «vérité» de la prédiction de la qualité. Cependant, le fait que ces évaluations subjectives doivent impliquer des observateurs humains les rend coûteuses, chronophage (=de longue durée) et impossibles à déployer dans la plupart des applications du monde réel.

Un modèle automatique d'IQA capable de fournir des mesures de qualité d'image de manière objective est préférable. Cependant, les scores subjectifs obtenus restent utiles pour concevoir et valider des mesures de qualité objectives.

2) Évaluation objective de la qualité d'image

L'évaluation objective de la qualité présente plusieurs avantages par rapport à l'évaluation subjective : tout d'abord, elle nécessite moins de ressources, elle est dans la plupart des cas plus économique, plus rapide, elle permet de surveiller les systèmes d'imagerie en temps réel. Elle est facile à utiliser pour l'optimisation des systèmes de traitement d'image, son utilisation nécessite peu ou ne nécessite pas de connaissances, et elle fournit généralement des résultats plus cohérents que les observateurs.

Les méthodes objectives d'IQA sont généralement classées en trois catégories (Mohammadi et al., 2014; Zhai and Min, 2020) :

- **Méthodes sans référence (NR-IQA)** (Ahmed et al., 2016) : il s'agit de méthodes qui évaluent la qualité des images sans avoir accès à des images de référence. Elles sont particulièrement adaptées à l'évaluation des performances des caméras car pour ce type de système, aucune référence n'est disponible. Parmi les métriques sans référence les plus utilisées, nous pouvons citer DIIVINE (Moorthy and Bovik, 2011), BLIINDS II (Saad et al., 2012) ou BRISQUE (Mittal et al., 2012).
- **Méthodes à référence réduite (RR-IQA)** (Rehman and Wang, 2012) : seules certaines parties des informations de l'image de référence sont nécessaires pour les modèles RR-IQA. Un ensemble de paramètres liés à la perception visuelle de la

qualité de l'image et sensibles à diverses distorsions sont d'abord identifiés et sélectionnés à partir de l'image de référence. Avec l'image déformée, ces paramètres sont ensuite utilisés pour prédire sa qualité. Les méthodes RR-IQA constituent un outil pratique pour évaluer la qualité de la communication visuelle en temps réel sur des réseaux filaires ou sans fil. Parmi les métriques à référence réduite les plus utilisées, citons WNISM (Wang et al., 2006), RR-SSIM (Rehman and Wang, 2012) ou RRED (Soundararajan and Bovik, 2011).

- **Méthodes de référence complète (FR-IQA)** (Phadikar et al., 2018) : les métriques de référence complète utilisent la référence et l'image déformée. Elles sont faciles à calculer en temps réel et sont en corrélation avec l'appréciation subjective humaine. Et elles sont particulièrement adaptées à l'évaluation de la compression car le but de ces algorithmes est de maintenir autant que possible l'intégrité d'une image de référence. Exemples de modèles FR-IQA : FSIM (Moorthy and Bovik, 2011), PSNR (Sheikh et al., 2006) et VIF (Sheikh and Bovik, 2006).

Ainsi, nous pouvons noter qu'en réponse à l'importance de la métrique de référence complète, un grand nombre de métriques d'évaluation de la qualité des images ont été proposées.

1.2.3 Survey des métriques de qualité d'image existantes

De nombreuses métriques FR-IQA ont été proposées dans la littérature (Phadikar et al., 2018). Pour mieux comprendre l'état de l'art des métriques de qualité d'image (QI), qui peut être utilisé pour améliorer ou développer de nouvelles métriques de QI mieux corrélées à la perception ; dans cette section, nous effectuons une enquête sur des métriques de qualité d'image de référence complète.

1) Méthodes Full Référence classique

- PSNR (Peak Signal to Noise Ratio) (Sheikh et al., 2006) : est basée sur l'hypothèse que le système visuel humain (HVS) est hautement adapté pour extraire les informations structurelles des images. La qualité de l'image est évaluée en calculant l'erreur de pixel globale, et l'erreur moyenne est utilisée comme résultat final d'évaluation.
- NQM (Noise Quality Metric) (Damera-Venkata et al., 2000) : imite le système visuel humain en tenant compte de la sensibilité au contraste, de la luminance locale, de l'interaction du contraste entre les fréquences spatiales et des effets de masquage du contraste.
- SSIM (Structural Similarity Index) (Wang et al., 2004) : estime la qualité de l'image en combinant les écarts entre l'image déformée et l'image de référence en termes de contraste, de luminance et de similarités structurelles.

- MSSIM (Mean Structural Similarity Index) (Wang et al., 2003) : est une méthode de similarité structurelle multi-échelle, qui diffère de la SSIM en intégrant les variations dans la résolution de l'image et les conditions de visualisation.
- IWSSIM (Information Weighted SSIM) (Wang and Li, 2011) : diffère de la méthode SSIM en ce qu'elle prend en compte le poids du contenu de l'information.
- VIF (Visual Information Fidelity) (Sheikh and Bovik, 2006) : quantifie la perte d'information de l'image par le processus de distorsion et explore la relation entre l'information d'image et la qualité visuelle.
- IFC (Information Fidelity Criterion) (Sheikh et al., 2005) : est basé sur les statistiques des scènes naturelles. Il utilise des modèles de sources de signaux et de distorsions.
- FSIM, FSIMc (Feature SIMilarity Metric) (Zhang et al., 2011b) : est basée sur le système visuel humain (HVS). FSIM reprend l'idée similaire de SSIM et évalue la qualité en calculant la similarité entre l'image déformée et l'image de référence en termes d'amplitude de gradient et de congruence de phase. FSIM a ensuite été étendue à FSIMc en incorporant les caractéristiques chromatiques de l'image.
- RFSIM (Riesz-transform based Feature SIMilarity metric) (Lin et al., 2010) : est basé sur le fait que le système de vision humaine (HVS) perçoit une image principalement en terme de ses caractéristiques de bas niveau. Les coefficients de transformation de Riesz du 1er ordre et du 2ème ordre de l'image sont considérés comme des caractéristiques de l'image, tandis qu'un masque de caractéristiques est défini comme les emplacements des bords de l'image. L'indice de similarité entre l'image de référence et l'image déformée est mesuré en comparant les deux cartes de caractéristiques à des emplacements clés marqués par le masque de caractéristiques.
- SR-SIM (Spectral Residual SIM) (Zhang and Li, 2012) : est basé sur un modèle de saillance visuelle spécifique, la saillance visuelle résiduelle spectrale. SR-SIM est conçu sur la base de l'hypothèse selon laquelle la carte de saillance visuelle d'une image est étroitement liée à la qualité perçue.
- VSI (Visual Saliency-Induced Index) (Zhang et al., 2014) : est basé sur l'hypothèse que la carte de saillance visuelle (SV) d'une image est étroitement liée à sa qualité perceptuelle. Dans le VSI, la carte de saillance visuelle est explorée en deux étapes. A l'étape du calcul de la carte de qualité locale, la carte VS est considérée comme une image ; à l'étape de la mise en commun des niveaux de qualité, elle est utilisée comme une fonction de pondération pour caractériser l'importance d'une région d'image locale.
- IFS (Independent Feature Similarity) (Chang et al., 2015) : permet de prédire efficacement la qualité d'une image présentant une distorsion des couleurs. Elle est

basée sur des caractéristiques indépendantes obtenues à partir d'un détecteur de caractéristiques entraîné sur des échantillons d'images naturelles.

- SFF (Sparse Feature Fidelity) (Chang et al., 2013) : est basée sur les caractéristiques éparses acquises par un détecteur de caractéristiques. Deux stratégies sont conçues pour simuler les propriétés de la perception visuelle : l'attention visuelle et le seuil visuel. La SFF reflète également les propriétés chromatiques du système HVS et est très efficace pour l'IQA couleur.
- GSM (Gaussian Scale Mixture) (Liu et al., 2012) : utilise la similarité de gradient pour mesurer le changement de contraste et de structure dans les images. Les effets des changements de luminance, de contraste et de structure sont intégrés via une méthode adaptative pour obtenir le score global de qualité de l'image.
- PerSIM (Perceptual SIMilarity index) (Temel and AlRegib, 2015) : est une métrique de modélisation des caractéristiques du système visuel et de la similarité de chrominance dans le domaine chromatique perceptuellement uniforme. Des caractéristiques laplaciennes et gaussiennes sont obtenues pour modéliser les cellules ganglionnaires de la rétine dans le système visuel humain, et la similarité des couleurs est également calculée. Dans cette métrique, une approche multi-résolution est suivie pour imiter la nature hiérarchique du système visuel humain.
- UNIQUE (UNsupervised Image Quality Estimation) (Temel et al., 2016) : est basée sur l'approche d'apprentissage non supervisé. Une transformation de l'espace colorimétrique, une soustraction de la moyenne et une opération de blanchiment sont utilisées pour améliorer le caractère descriptif des images en réduisant la redondance spatiale, un décodeur linéaire est utilisé pour obtenir des représentations éparses et une étape de seuillage est utilisée pour formuler des mécanismes de suppression dans un système visuel.
- MSUNIQUE (Multi-model and Sharpness-weighted UNsupervised Image Quality Estimation) (Prabhushankar et al., 2017) : est une extension de l'estimateur de qualité UNIQUE. En analysant le réseau d'apprentissage utilisé et en développant manuellement un système de pondération qui capture la netteté. Ceci est effectué dans les blocs de pré-traitement et de post-traitement en améliorant l'information acquise à partir des données, en analysant les caractéristiques des bords des filtres appris afin que leurs réponses soient pondérées en fonction de la théorie de l'évaluation de la qualité.
- COHERENSI (Chaos of Harmonics/phase in Error as a REference-based Novel Similarity Index) (Hegazy and Al Regib, 2014) : est basé sur l'analyse du chaos harmonique et l'analyse du chaos de phase. Ces composants produisent deux cartes de caractéristiques qui sont pondérées, additionnées, mises à l'échelle logarithmique et moyennées pour obtenir un score de qualité.

- RVSIM (Riesz Transform and Visual contrast sensitivity-based feature SIMilarity index) (Yang et al., 2018) : est basé sur la sensibilité au contraste et la sensibilité au contraste visuel. Il combine la transformée de Riesz et la sensibilité au contraste visuel. En exploitant la fonction CSF (Contrast Sensitivity Function) pour attribuer les poids des différentes bandes de fréquence. En même temps, la similarité GM (Gradient Magnitude) est introduite pour obtenir la matrice de similarité de gradient. Ensuite, la matrice MPC (Monogenic Phase Congruency) est utilisée pour construire la fonction de regroupement et obtenir l'index RVSIM.
- MAD (Most Apparent Distortion) (Chandler, 2010) : elle fonctionne selon la pré-misse que l'HVS exécute deux stratégies distinctes lorsqu'il évalue la qualité de l'image : Le masquage local de la luminance et du contraste évalue les images de haute qualité, tandis que les statistiques locales des composantes de fréquence spatiale évaluent les images de basse qualité.

Malgré la disponibilité de toutes ces métriques, il n'existe pas de métrique unique capable de mesurer la qualité de l'image pour tous les types de dégradations. Par conséquent de nombreuses études ont récemment adopté une nouvelle stratégie faisant appel à différentes techniques de fusion pour quantifier la qualité de l'image. En fonction des informations qu'ils utilisent, les algorithmes de fusion IQA peuvent être classés en deux catégories : la fusion de caractéristiques ou la fusion de métriques.

2) Méthodes par fusion des caractéristiques

Dans cette catégorie, les algorithmes de fusion IQA visent à prédire la qualité d'image à partir de caractéristiques extraites des images ; après avoir calculé un ensemble de différentes caractéristiques, une combinaison de certains éléments de cet ensemble est effectuée de différentes manières pour produire le score de qualité des images. Par exemple :

- Dans (Pei and Chen, 2015) : Une méthode FR-IQA est proposé en apprenant un modèle à partir du processus de la rétine ; une combinaison non linéaire de caractéristiques extraites de plusieurs bandes de fréquences gaussiennes (DOG : Difference-Of-Gaussian), qui imitent le système visuel humain (HVS), et des techniques d'apprentissage automatique sont utilisées pour émuler le mécanisme du cerveau humain sont appliquées. L'étape DOG et l'étape de régression par forêt aléatoire (Random Forest) sont mises en œuvre en parallèle, ce qui permet un calcul plus rapide.
- Dans (Al-Bandawi and Deng, 2018), les auteurs ont développé une technique IQA spécifique sans référence qui utilise des caractéristiques dérivées de la conformité de la distribution du premier chiffre (First Digit Distribution : FDD) des images naturelles dans le domaine de la transformation avec la loi de Benford ; une com-

binaison non linéaire de ces caractéristiques est entraînée à l'aide d'une régression par processus gaussien avec un noyau quadratique rationnel.

- Dans (Zhang and Hou, 2019), les auteurs ont conçu une méthode pour évaluer la qualité de l'image de diffraction cellulaire, par une combinaison de luminance, de contraste, de représentation profonde et de matrice de cooccurrence des niveaux de gris. Ils utilisent K-means et SVM pour prétraiter l'image de diffraction cellulaire et DNN pour extraire la représentation profonde de chaque image, et à la fin ils fusionnent les quatre caractéristiques.
- Récemment, un modèle FR-IQA basé sur la similarité été introduit sans procédure d'apprentissage (VCGS : Visual saliency with Color appearance and Gradient Similarity) (Shi and Lin, 2020), qui combine trois caractéristiques : la saillance visuelle, la structure et la chrominance.

3) Méthodes par fusion de métriques

Dans la deuxième catégorie, les algorithmes de fusion IQA sont basés sur la combinaison de deux métriques ou plus. La combinaison de métriques est effectuée dans différentes stratégies. La qualité d'image prédictive est le résultat des valeurs de métriques combinées, en fonction de la stratégie utilisée. Par exemple :

- Dans (Liu and Yang, 2008), la méthode CCA (Canonical Correlation Analysis) est utilisée pour la fusion des quatre métriques SNR, SSIM, VIF et VSNR. La métrique de fusion augmente la précision et la robustesse des prévisions pour différents types de distorsion. Une caractéristique intéressante de cette métrique est qu'elle fournit un réglage flexible des paramètres qui exploite efficacement la connaissance préalable de la base de données de test.
- Les auteurs dans (Chetouani et al., 2010) ont proposés une métrique globale de qualité d'image appelée G-IQM (Global-IQM). Le schéma de fusion proposé est divisé en deux étapes. La première étape consiste à améliorer les performances de l'IQM pour des distorsions spécifiques (IQM-D) (IQM-Distorsion). Ensuite, tous les IQM-D sont combinés à l'aide d'un réseau d'apprentissage ANN en un seul indice global, le GIQM.
- La régression régularisée a été utilisée pour combiner jusqu'à sept métriques IQA dans (Lahouhou et al., 2010). Tout en appliquant des techniques de sélection de variables pour trouver automatiquement les indicateurs les plus pertinents, les trois métriques (SSIM, PSNR et MSE) contribuent le plus.
- Dans (Okarma, 2010) les scores de MSSIM, VIF et R-SVD ont été combinés de manière non linéaire. Où toutes les images ont été converties en niveaux de gris avant les calculs. Le principal avantage de cette approche est la forte corrélation

linéaire avec les scores subjectifs sans mappage non linéaire supplémentaire. La métrique proposée est définie comme suit :

$$\text{CQM} = (\text{MS-SSIM})^a * (\text{VIF})^b * (\text{R-SVD})^c \quad (2)$$

Cette métrique a été améliorée en 2012 (Okarma, 2012) en remplaçant la métrique R-SVD par la métrique FSIMc et a proposé la métrique CISI. Le CISI est considérée comme l'approche la plus universelle indépendamment du contenu de l'image et des types de distorsions de l'image par rapport à CQM.

Une version étendue de l'approche précédente basée sur la combinaison pondérée de quatre mesures : MS-SSIM, VIF, RFSIM et FSIM est proposée dans (Okarma, 2013). Par la formule :

$$\text{EHIS} = (\text{MS-SSIM})^a * (\text{VIF})^b * (\text{WFSIMc})^c * (\text{RFSIM})^d \quad (3)$$

En raison de sa corrélation hautement linéaire avec les évaluations subjectives de la qualité, son application ne nécessite aucune opération de mappage supplémentaire menant directement à une prédiction précise de la qualité de l'image, quel que soit le type de distorsion.

La vérification de cette approche pour des images présentant des distorsions multiples a été présentée dans l'article (Okarma, 2014).

- Peng et Li (Peng and Li, 2012) ont à leur tour présenté une approche basée sur un mélange conditionnel bayésien de modèle d'expert. Dans cette approche, un classifieur par la machine à vecteurs de support (SVM) a été utilisé pour prédire le type de distorsion, puis le SSIM, le VSNR et le VIF ont été fusionnés avec une régression par k plus proches voisins.
- Un autre mécanisme a été utilisé pour combiner des mesures de qualité objectives présenté dans (Wu et al., 2013). Les mérites des métriques IQA existantes sont intégrés au guide du mécanisme génératif interne (IGM). L'adoption d'un algorithme de prédiction autorégressif pour décomposer une scène d'entrée en deux parties, la partie prédite avec le contenu visuel et la partie désordonnée avec le contenu résiduel. Les distorsions sur la partie prédite dégradent les informations visuelles primaires et des métriques de similarité structurelle sont utilisées pour mesurer sa dégradation. Enfin, en fonction du déploiement de l'énergie sonore sur les deux parties, le score de qualité global est obtenu par la combinaison des deux résultats d'évaluation.
- La fusion localement adaptative (LAF) proposée dans (Barri et al., 2014) est spécifiquement conçue pour la prédiction de la qualité perceptuelle des images ou des vidéos. Une mesure de qualité objective basée sur LAF est construite en deux étapes. La première étape consiste en une sélection de mesures de qualité objective dont la

portée est limitée, c'est-à-dire qu'elles ne sont fiables que pour des contenus et des classes de distorsion spécifiques. La deuxième étape consiste à combiner les mesures objectives de qualité à portée limitée sélectionnées via la pondération adaptative, où les facteurs de pondération sont déterminés par une formation sur une base de données d'évaluation subjective de la qualité.

L'implémentation implique une mesure NR et deux mesures RR, sur quatre types de distorsion standard considérés dans la plupart des bases de données d'évaluation subjective de la qualité des images : le flou gaussien, la compression JPEG, la compression JPEG2000 et le bruit blanc.

- Dans (Lukin et al., 2015), le réseau de neurones a été utilisé pour combiner les six mesures IQA : FSIMc, PSNR-HMA, PSNR-HVS, SFF, SR-SIM et VIF. Ces mesures ont été utilisées en tant que données d'apprentissage comme entrées réseau, pour le cas des images de la base de données TID2013.

La moitié des 3000 images de la base de données TID2013 sélectionnées au hasard ont été utilisées au stade de l'apprentissage, tandis que l'autre moitié a été exploitée pour évaluer la qualité des métriques HVS en se basant sur un réseau de neurones.

- Dans (Oszust, 2016a,b) Un problème d'optimisation de la conception d'une approche de la décision de fusion de 18 mesures IQA (une combinaison non linéaire) est défini et résolu à l'aide d'un algorithme génétique, qui est également chargé de sélectionner les mesures IQA appropriées, sauf que dans (Oszust, 2016b) c'est la somme pondérée au lieu du produit pondéré dans (Oszust, 2016a). Les mêmes auteurs dans (Oszust, 2016c) utilisent la régression linéaire multiple des opinions fournies par les mesures IQA sélectionnées génétiquement.
- Une autre métrique est proposée dans (Borhen Eddine et al., 2016). Les auteurs ont utilisé le schéma d'optimisation de l'essaim de particules pour sélectionner un ensemble de métriques IQA pertinentes, puis une stratégie de fusion basée sur la régression par vecteurs de support est adoptée pour dériver l'indice global de qualité d'image.
- Dans l'article (Ieremeiev et al., 2018), une approche robuste pour la conception des métriques de référence complète est proposée, basée sur la combinaison de plusieurs métriques de référence complète existantes. Une linéarisation préliminaire de la dépendance du MOS par rapport aux métriques des composants est effectuée afin de compenser les défauts de chaque composant. La métrique a été conçue pour trois configurations : médiane de trois estimations de MOS, médiane de cinq estimations, moyenne rognée en alpha de cinq estimations. Les aspects d'ajustement et de linéarisation ont été pris en compte.
- Un réseau de neurones artificiels a été utilisé pour combiner différentes métriques sans référence dans (Ieremeiev et al., 2019), où plusieurs types de tel réseaux avec

différentes configurations et l'influence de divers facteurs sur la précision finale de ces métriques ont été étudiés. Différentes métriques avec différentes combinaisons ont été conçues.

1.2.4 Les bases d'images populaires pour l'IQA

Une fois qu'une solution algorithmique d'évaluation de la qualité est établie, il nécessite de mesurer les erreurs de prédiction, ce qui implique que nous disposons d'une vérité terrain. Cette vérité terrain est généralement accessible sous la forme de valeurs MOS ou DMOS. Actuellement, plusieurs bases de données de qualité des images ont été développées récemment à ces fins de conception et de validation. Ces bases de données contiennent généralement les images utilisées dans les expériences et leurs scores de qualité associés notés par les participants.

Parmi les bases de données couramment utilisées dans les travaux sur IQA, nous pouvons citer :

- **Base de données LIVE** (Wang et al., 2004) : Cette base de données a été développée à l'Université du Texas à Austin, aux États-Unis (USA). Elle est composée de 982 images avec 29 images de référence en couleurs et de leurs images dégradées utilisant les 5 types de distorsions suivants : JPEG2000, JPEG, bruit blanc, flou gaussien, et erreurs de bits, avec 5 à 7 niveaux de distorsion. Entre 20 et 29 observateurs ont participé aux expériences. La qualité perçue de chaque image de cette base de données est fournie et représentée par un score DMOS.
- **Base de données CSIQ** (Chandler, 2010) : Cette base de données a été développée à l'université Oklahoma State University, États-Unis, avec 35 observateurs. Elle comprend 866 images dégradées. Les six distorsions utilisées dans cette base sont : la compression JPEG, la compression JPEG2000, le décalage de contraste global (CONTRAST), le bruit Gaussien rose (PN), le bruit Gaussien blanc (WN) et le flou Gaussien (GBLUR) ont été appliquées à 30 images de référence de 4 à 5 niveaux de dégradation. Le DMOS de chaque image représente la qualité perçue.
- **Base de données TID2008** (Ponomarenko et al., 2009) : Cette base de données a été développée à l'université Tampere University of Technology, Finland. Elle contient 17 types de distorsions : (a) bruit gaussien additif ; (b) Bruit additif dans les canaux de couleur ; (c) bruit spatialement corrélé ; (d) bruit masqué ; (e) bruit à haute fréquence ; (f) bruit impulsionnel ; (g) bruit de quantification ; (h) flou gaussien ; (i) débruitage d'image ; (j) compression JPEG ; (k) compression JPEG2000 ; (l) erreurs de transmission JPEG ; (m) erreurs de transmission JPEG2000 ; (n) bruit de motif de non-excentricité ; (o) distorsions locales par blocs d'intensité différente ; (p) décalage moyen (décalage d'intensité) ; (q) changement de contraste. Chaque distorsion a été

appliquée à 25 images de référence à 4 niveaux de dégradation, résultant en 1700 images dégradées (25 images de référence x 17 types de distorsions x 4 niveaux de distorsions). Les scores ont été recueillis auprès de 838 observateurs, et un MOS a été calculé pour chaque image.

- **Base de données IVC** (Ninassi et al., 2006) : Cette base contient 10 images de référence, 235 images ont été générées à partir de 4 types de dégradations différentes : JPEG, JPEG2000, LAR coding et le flou. Avec 5 niveaux d'appréciation et 15 observateurs ont été impliqués dans les expériences afin de calculer le DMOS de chaque image.

Ces bases de données sont développées sous la contrainte que les images sont soumises à un seul type de distorsion. Étant donné que les images peuvent être soumises à plusieurs types de distorsion dans des scénarios de vie plus réalistes, des efforts récents ont été faits pour développer des bases de données d'évaluation subjective des images à distorsion multiple. Voici quelques exemples de bases de données d'images à distorsions multiples :

- **Base de données TID2013** (Ponomarenko et al., 2015) : Cette base est une extension de la version de la base de données TID2008. Les mêmes 25 images de référence ont été soumises à 24 types de distorsion avec 5 niveaux de dégradation pour produire 3000 images dégradées. Sept types de distorsions supplémentaires ont été inclus dans la base de données : bruit multiplicatif, bruit de confort, changement de saturation des couleurs, erreur de quantification des couleurs, aberrations chromatiques, compression avec perte et erreur d'échantillonnage. Au total 985 personnes ont participé aux expériences. Les valeurs MOS de toutes les images ont été acquises via une évaluation par comparaison appariée.
- **Base de données MDID** (Sun et al., 2017) : Cette base de données a été développée à l'Université Shenzhen Tsinghua University, China. Elle est constituée de 20 images de référence. Chaque image a d'abord été floutée ou on contraste modifié, puis compressée par JPEG ou JPEG2000 et en fin soumise à un bruit supplémentaire. Il existe 4 niveaux de dégradation pour chaque distorsion, ce qui donne un total de 1600 images dégradées. Les scores DMOS ont été recueillis auprès de 192 observateurs.

La Figure 1.2 présente des images de référence des différentes bases.



Figure 1.2. - Les images de référence des différentes bases

Le tableau 1.1 résume les informations de ces bases de données IQA. où STD : Distorsion Type Simple. MTD : Distorsion Type Multiple.

Tableau 1.1. - Caractéristiques des bases d’images

	LIVE	IVC	CSIQ	TID2008	TID2013	MDID
Nombre d’images de référence	29	10	30	25	25	20
Nombre d’images de distorsion	779	235	866	1700	3000	1600
Types de distorsion	5	4	6	17	24	5
Niveaux de distorsion	5	5	4-5	4	5	4
Type de base de données	STD	STD	STD	STD	MTD	MTD
Nombre de dist. dans une image	1	1	1	1	1-2	1-4
Format des notes subjectives	DMOS	DMOS	DMOS	MOS	MOS	MOS
Intervalle des notes subjectives	0-100	1-5	0-1	0-9	0-9	0-8
Format d’image	BMP	BMP	PNG	BMP	BMP	BMP
Année de publication	2006	2006	2010	2008	2013	2017

1.3 Qualité de la vidéo

Avec le développement rapide des technologies de communication, la vidéo a joué un rôle important dans les systèmes de communication multimédia. Par conséquent, il est crucial de maintenir la qualité de la vidéo à un niveau acceptable dans divers environnements d’application tels que les communications visuelles en réseau.

Dans ce contexte, l’évaluation de la qualité vidéo (VQA : Video Quality Assessment) joue un rôle essentiel dans le domaine du traitement vidéo, notamment l’acquisition vidéo, le filtrage vidéo lors de la récupération, la compression vidéo, la restauration vidéo et l’amélioration vidéo (LI et al., 2019).

De nombreux facteurs peuvent affecter et/ou altérer la qualité des vidéos (Winkler, 2005). La plupart des normes contemporaines de codage vidéo utilisent la compensation de mouvement et des schémas de codage par blocs pour la compression. En conséquence, la vidéo décodée souffre d’un ou plusieurs artefacts de compression, tels que le blocage, le flou, l’altération des couleurs, l’oscillation amorties (Ringing), les faux bords, le mouvement irrégulier, l’inadéquation de la chrominance et le reflet. Les erreurs de transmission telles que les paquets endommagés ou perdus peuvent dégrader davantage la qualité vidéo. En outre, les étapes de pré ou de post-traitement dans le système de transmission vidéo, telles que la conversion de domaine (analogique à numérique ou vice-versa), la conversion de la fréquence d’images dégradent également la vidéo. Il est donc nécessaire d’évaluer la

qualité de la vidéo.

Les méthodes d'évaluation de la qualité vidéo (VQA) ont été largement étudiées (Seshadri-nathan et al., 2010) comme un ensemble d'outils de plus en plus important. Ces méthodes VQA peuvent être divisées en deux catégories étroitement liées : les études subjectives de la qualité vidéo et la modélisation objective de la qualité vidéo. En fait, il existe un lien très fort entre les études subjectives et les méthodes objectives d'évaluation de la qualité. L'évaluation subjective de la qualité de vidéo est différente de celle des images en raison de la réaction du système visuel humain et de ses propriétés telles que la durée de fixation de l'œil et la mémoire visuelle à court terme.

Les Recommandation ITU-T P.912¹ (ITU : International Télécommunications Union) définissent certaines procédures les plus couramment utilisées pour l'évaluation subjective de la qualité ; à savoir les tests à stimulus unique, les tests à double stimulus et les tests comparatifs ainsi qu'une méthode spécifiée au cas des vidéos nommée SAMVIQ (Subjective Assessment Methodology for Video Quality). Il s'agit d'une méthode alternative qui est un mélange de méthodes à stimulus unique, à double stimulus et de comparaison, c'est à dire une méthode d'échelle de qualité continue multi-stimulus avec des références explicites et cachées. Chaque méthode dépend de l'intention et du problème de recherche. Les faiblesses et les forces de ces différentes méthodologies ont été comparées par de nombreux chercheurs (Zhang et al., 2018; Leszczuk and Janowski, 2021).

Les notes de qualité vidéo perceptuelle obtenues par les évaluateurs sont moyennées pour obtenir le score d'opinion moyen (MOS) ou bien le score d'opinion moyen différentiel (DMOS).

Les méthodes subjectives d'évaluation de la qualité vidéo sont capables de mesurer de manière fiable la qualité vidéo perçue par le système visuel humain (HVS). Cependant, avec les données volumineuses entrant dans Internet et le système de communication, l'évaluation subjective pose un défi. Elle est non seulement longue et coûteuse, elle n'est pas en mesure de fournir une mesure instantanée de la qualité vidéo. C'est pourquoi l'évaluation objective de la qualité est la voie à suivre.

L'évaluation objective de la qualité vidéo est également plus difficile que celle de la qualité de l'image, car il faut tenir en compte des variations temporelles de la vidéo.

L'évaluation objective VQA, comme dans le cas des IQA, peuvent être classés en trois

1. Pour accéder à la recommandation, saisissez l'URL : <https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=14828&lang=fr>

catégories (Chikkerur et al., 2011) : les modèles à référence complète (FR) qui ont accès à une référence parfaite de haute qualité, les modèles à référence réduite (RR) qui utilisent des informations de référence limitées et modèles sans référence (NR) qui utilisent uniquement la version déformée pour la prédiction de la qualité.

1.3.1 Survey des métriques VQA

La littérature sur le FR-VQA a considérablement mûri car une multitude d’approches ont été proposées. Par exemple des types de méthodes pour VQA qui étendent les métriques FR-IQA (Zhang et al., 2011b; Wang et al., 2004, 2003) pour une application sur des vidéos, en prédisant la qualité de chaque image vidéo et en utilisant un schéma de regroupement temporel approprié. Bien que cette procédure puisse être peu coûteuse en termes de calcul, ses performances sont limitées car les informations utiles sur la qualité temporelle ne sont pas utilisées efficacement.

En plus de ces métriques, il existe d’autres métriques spécifiques à la vidéo comme la métrique de qualité vidéo (VQM) (Pinson and Wolf, 2004). Il s’agit de l’une des premières méthodes VQA, qui utilise les pertes de gradients de luminance spatiale, ainsi que des caractéristiques basées sur le produit du contraste de luminance et du mouvement. Son extension VQM-VFD (Pinson et al., 2014) est particulièrement efficace pour capturer les retards de trame et atteint des performances compétitives sur la base de données LIVE-mobile (Moorthy et al., 2012).

La métrique MOVIE (MOtion-based Video Integrity Evaluation) (Seshadrinathan and Bovik, 2009) et la métrique VSSIM (Video Structural SIMilarity) (Seshadrinathan and Bovik, 2007) utilisent l’idée de réglage du mouvement en suivant les artefacts perceptible le long des trajectoires de mouvement pour mesurer la qualité vidéo.

Les modèles proposés dans (Ortiz-Jaramillo et al., 2014), (Manasa and Channappayya, 2016) utilisent les caractéristiques du flux optique pour mesurer la qualité vidéo.

La métrique VMAF (Video Multi-method Fusion) (Li et al., 2016) développé par Netflix est un prédicteur de qualité largement utilisé basé sur des métriques IQA/VQA existants où les caractéristiques sont combinées à l’aide d’un régresseur de vecteur de support (SVR). Cependant, elle évalue la qualité globale de la trame, ce qui n’est pas idéal dans un environnement dans lequel l’estimation de la qualité au niveau des blocs est requise (par exemple : environnement d’optimisation débit-distorsion RDO). Cette métrique a été améliorée dans (Bampis et al., 2018) en ajoutant des caractéristiques supplémentaires pertinentes sur le plan perceptif qui expriment mieux les aspects temporels de la qualité

vidéo ; la nouvelle métrique est nommée ST-VMAF.

Dans (Tu et al., 2021), les auteurs utilisent une stratégie consistant à sélectionner 60 caractéristiques les plus performantes parmi 763 extraites des modèles NR-VQA efficaces existants afin de créer un nouveau modèle basé sur la fusion appelé VIDEVAL (VIDeo quality EVALuator). Le modèle est obtenu par une régression symbolique à l'aide de l'algorithme SVR (Support Vector Regression). Le modèle est adapté pour effectuer l'optimisation débit-distorsion. Trois familles différentes de prédicteurs ont été conçues, qui utilisent une combinaison linéaire et non linéaire et avec un nombre différents d'entrées.

La récente popularité de l'apprentissage en profondeur a conduit à une variété de modèles basés sur CNN (Kim et al., 2018; Becker et al., 2019; Xu et al., 2020a) qui atteignent des performances compétitives sur les bases de données VQA existantes. Notez que tous ces modèles nécessitent que les vidéos de référence et déformées aient le même nombre d'images en synchronisation temporelle ainsi que la même résolution spatiale.

Une nouvelle métrique appelée GREED (GeneRalizEd Entropic Difference) été proposé dans (Madhusudana et al., 2021b). Elle est basée sur l'utilisation de classes très générales de filtres temporels qui sont déployés à plusieurs échelles, ainsi que sur l'utilisation de l'apprentissage sur les données pour obtenir une prédiction de qualité perceptuelle considérablement améliorée.

Les mêmes auteurs de cette métrique ont développé une autre métrique (Madhusudana et al., 2021a), basée sur la combinaison des caractéristiques spatiales de VMAF (Li et al., 2016) avec les caractéristiques spatio-temporelles de GREED (Madhusudana et al., 2021b). Ce modèle exploite les avantages des deux modèles et il est intéressant lorsque les vidéos de référence et déformées ont des fréquences d'images différentes, sans prétraitement temporel supplémentaire et peut capturer efficacement d'autres artefacts que les artefacts de fréquence d'images et de compression.

Les auteurs dans (Izima et al., 2021) ont passé en revue les applications récentes des techniques d'apprentissage qui utilisent des mesures de qualité de livraison pour la prédiction de la qualité vidéo. Les résultats de l'enquête révèlent que les algorithmes d'apprentissage traditionnels sont les modèles les plus largement adoptés pour résoudre les problèmes de prédiction de la qualité vidéo. Cette famille d'algorithmes a beaucoup de potentiel car ils sont bien compris, faciles à déployer et ont des exigences de calcul plus faibles que les techniques d'apprentissage en profondeur.

Récemment, les auteurs dans (Ganesh and Patil, 2022) ont proposé une méthode FR-VQA,

ils appliquent huit algorithmes d'optimisation pour sélectionner les métriques classiques de qualité les plus performantes parmi 16 métriques. La régression par RMSE est ensuite utilisée pour construire le modèle prédictif.

1.3.2 Les bases populaires de vidéos pour VQA

Les chercheurs ont consacré des efforts considérables au développement de bases de données vidéo VQA de haute qualité. En fait, la plupart de ces bases de données sont utilisées comme données de «vérité terrain» pour des métriques objectives d'évaluation de la qualité. Ainsi, ces bases de données aideront à la validation des algorithmes objectifs d'évaluation de la qualité vidéo.

Un nombre croissant de bases de données sont disponibles dans le domaine public, au point qu'il est devenu difficile de choisir celle qui convient le mieux. La comparaison de ces bases à l'aide des mêmes critères est utile pour les développeurs de modèles, qui peuvent décider en connaissance de cause quelles bases de données sont les mieux adaptées à leurs besoins.

Le Tableau 1.2 résume quelques bases de données publiques populaires de la VQA.

Tableau 1.2. - Caractéristiques des bases VQA

Base de Vidéos	#Vidéos référence	#Vidéos dégradé	Résolution	Durée	Fréquence d'images	Format	Types de distorsion
LIVE VQA (Seshadrinathan et al., 2010)	10	150	768*432	10 sec.	25, 50	Yuv420p8b	H.264, MPEG-2, TE
EPFL (Simone et al., 2010)	12	144	352*288 704*576	10 sec.	30	Yuv420p8b	H.264, TE
IVPL (Zhang et al., 2011a)	10	128	1920*1088	10 sec.	25	Yuv420	DIRAC, H.264, IPPL, MPEG2
LIVE Mobile (Moorthy et al., 2012)	10	200	1280*720	15 sec.	30	Yuv420p8b	H.264, TE, RA
CSIQ-VQA (Vu and Chandler, 2014)	12	216	832*480	10 sec.	24, 25, 30, 50, 60	Yuv420p8b	H.264, H.265, MJPEG, WC, TE, AWN
LIVE-HTTP (Chen et al., 2014)	8	15	1280*720	300 sec.	30	Yuv420p8b	QoE
NFLX (Li et al., 2016)	34	300	1920*1080	6 sec.	24, 25, 30	Yuv420p8b	H.264 and scaling
LIVE VQC (Sinno and Bovik, 2018)	-	585	1080*240	10	19,30	MP4	Diverse distorsions

TE : erreurs de transmission, RA : adaptation de débit, MJPEG : compression motion JPEG, WC : compression par ondelettes, AWN : bruit blanc additif, QoE : adaptation de débit et/ou rebuffering. yuv420p8b : YUV 420 planaire, profondeur 8 bits.

1.4 Domaines d'application d'I/VQA

En tant que forme d'information interactive, les images jouent un rôle de plus en plus important dans le domaine du multimédia. L'importance des informations véhiculées par les images est fortement liée à la qualité de l'image. En général, plus la qualité de l'image est élevée, plus les gens peuvent recevoir et percevoir d'informations en regardant l'image. Actuellement, les méthodes IQA devient de plus en plus importante dans le domaine

du traitement d'image et de la vision par ordinateur. Elles sont largement utilisées dans différents domaines d'applications. A titre d'exemples mais sans s'y limiter :

La distribution de sessions de streaming vidéo pour les utilisateurs fixes et mobiles avec prise en charge de la qualité de service et de la qualité d'expérience (Ghani and Ajrash, 2018), dans les applications d'imagerie médicale (Ma et al., 2020), dans les applications de reconnaissance de visage (Dihin et al., 2020), les jeux vidéo et l'infographie, ainsi que les images de synthèse (Piórkowski et al., 2017), l'évaluation des surfaces imprimées en 3D (Fastowicz and Okarma, 2018), La recherche sur le sauvetage sous-marin et le suivi biométrique dans les piscines (Lei et al., 2022)...etc.

Avec le développement des appareils numériques et des services multimédias, les vidéos émergentes sont nombreuses donc l'évaluation de la qualité vidéo devient critique. Il existe diverses applications pour l'évaluation de la qualité vidéo, certains d'entre eux sont les suivants :

Les vidéos tridimensionnelle (3D) utilisés par exemple dans les industries du cinéma et de la radiodiffusion (Appina and Channappayya, 2018). Les applications médicales telles que la téléchirurgie (Lévêque et al., 2017). Dans la technologie de la réalité virtuelle, nous pouvons noter les vidéos panoramiques (Zhang et al., 2018) et les vidéos à points de vue libres (Ling and Le Callet, 2017). En outre, nous pouvons énumérer les applications typiques des caméras de surveillance (Uhrina et al., 2021), Le traitement vidéo/image à 360° (Xu et al., 2020b), les services de streaming vidéo qui consomment actuellement la majorité du trafic Internet (Wang et al., 2021)...etc.

1.5 Conclusion

Le problème de l'évaluation de la qualité multimédia est important pour de nombreux domaines d'application différents. Cependant, les images et les vidéos sont affectées par une grande variété de déformations tout au long des différentes étapes de la chaîne de distribution ; ce qui nécessite une évaluation de la qualité. Dans ce chapitre, nous avons expliqué la qualité d'image et de vidéo et les types de distorsions les plus survenues. Nous avons également abordé les différents types d'évaluation de la qualité et les méthodes existantes ainsi que les différentes bases d'images et de vidéos les plus utilisées. Une recherche a été menée sur les différents domaines d'application de qualité d'images et de vidéos.

Différentes approches et algorithmes métaheuristiques sont utilisés pour prédire la qualité d'image et de vidéo. Dans le chapitre suivant, nous présenterons les concepts de base liés aux métaheuristiques ainsi que les principaux algorithmes.

Chapitre 2

Optimisation et Métaheuristique

2.1 Introduction

L'optimisation est une branche des mathématiques appliquées et de l'analyse numérique. La majorité des problèmes d'ingénierie, de science, d'économie et de vie peuvent être formulés comme un problème d'optimisation. Les métaheuristiques sont largement reconnues comme une approche efficace pour de nombreux problèmes d'optimisation difficiles. Comme le résume Talbi (Talbi, 2009), «*l'optimisation est partout ; les problèmes d'optimisation sont souvent complexes ; alors les métaheuristiques sont partout*».

Dans ce chapitre nous présenterons le problème d'optimisation, ainsi que les algorithmes permettant de résoudre ce problème tout en mettant l'accent sur les métaheuristiques. La classification adoptée dans ce chapitre distingue les métaheuristiques basées sur une solution unique et les métaheuristiques basées sur la population.

2.2 Problème d'optimisation

Le mot optimisation signifie rechercher le meilleur moyen de réaliser une activité. Il peut également être défini comme “*Un algorithme d'optimisation est un processus qui est mis en œuvre de manière itérative en comparant et en modifiant des solutions de test jusqu'à obtention d'une solution optimale*” (Chong and Zak, 2004).

- De manière informelle, un problème d'optimisation consiste à trouver la ou les solutions qui maximisent ou minimisent un critère donné diversement appelé fonction objectif, fonction de coût ou fonction de fitness. Une solution peut être tenue d'obéir à certaines contraintes, alors qu'elle est généralement appelée solution réalisable.
- Formellement, une fonction est optimisée lorsqu'on trouve les arguments x permettant de générer le résultat minimum (ou maximum) possible pour la fonction donnée. L'optimisation d'une fonction mathématique est utile lorsque cette fonction modé-

lise un processus d'intérêt. La solution de l'optimisation est alors décrite comme suit :

$$x^* = \operatorname{arg}_{x \in S} \min f(x) \quad (1)$$

En fait, le problème de maximisation est inclus dans la formulation de la minimisation car :

$$\max\{f(x)/x \in S\} = \min\{-f(x)/x \in S\} \quad (2)$$

Dans la notation précédente, S est utilisé pour regrouper toutes les solutions possibles du problème d'optimisation en tant qu'ensemble de solutions. Les arguments x sont appelés solution, et la fonction f est utilisée pour mesurer sa «qualité».

Lorsqu'un problème d'optimisation n'implique qu'une seule fonction objectif, il s'agit d'une optimisation mono-objectif. La plupart des problèmes d'ingénierie nécessitent que le concepteur optimise un certain nombre d'objectifs contradictoires. Les objectifs sont en conflit les uns avec les autres si l'amélioration d'un objectif entraîne la détérioration d'un autre. Les problèmes d'optimisation multi-objectifs sont également appelés problèmes d'optimisation multicritères ou vectoriels.

➤ Le problème d'optimisation multiobjectif

L'optimisation multi-objectifs (MOO) est une technique efficace pour trouver un ensemble de solutions qui définissent le meilleur compromis entre des objectifs concurrents tout en satisfaisant plusieurs critères. Formellement, le problème d'optimisation multiobjectif s'énonce comme suit (Chong and Zak, 2013) :

$$\operatorname{minimize} f(x) = \begin{bmatrix} f_1(x_1, x_2, \dots, x_n) \\ f_2(x_1, x_2, \dots, x_n) \\ \vdots \\ f_l(x_1, x_2, \dots, x_n) \end{bmatrix} ; x \in \Omega \quad (3)$$

Où : $f : \mathbb{R}^n \rightarrow \mathbb{R}^l$ et $\Omega \subset \mathbb{R}^n$.

Par exemple, l'ensemble des contraintes Ω peut prendre la forme :

$$\Omega = \{x : h(x) = 0, g(x) \leq 0\} \quad (4)$$

Où : $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $g : \mathbb{R}^n \rightarrow \mathbb{R}^p$, $m \leq n$.

La résolution de tels problèmes est très difficile par rapport à l'optimisation à objectif unique. En effet, pour les problèmes d'optimisation multi-objectifs, les objectifs étant généralement contradictoires, il n'existe pas une solution optimale mais un ensemble de

solutions de compromis. Cet ensemble de solutions de compromis est connu sous le nom de Pareto-optimal, comme l'illustre la figure 2.1.

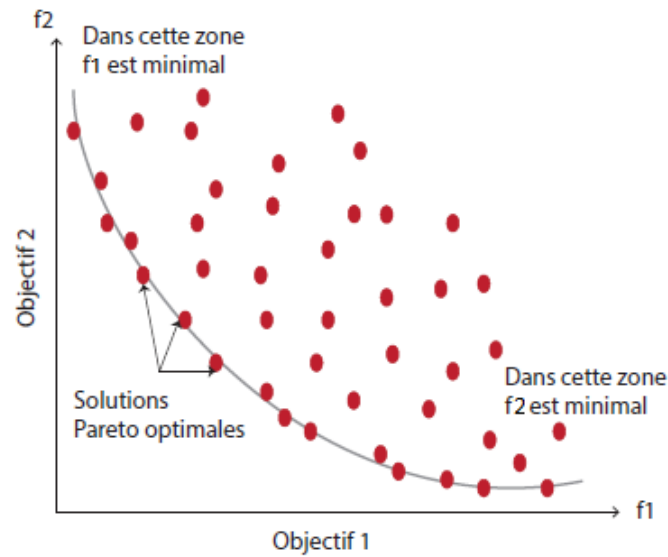


Figure 2.1. - Front de Pareto pour un problème de minimisation de deux objectifs

Une solution est de Pareto optimale s'il n'est pas possible d'améliorer un objectif donné sans détériorer au moins un autre objectif. Cet ensemble de solutions représente les solutions de compromis entre les différents objectifs conflictuels. L'objectif principal de la résolution d'un problème multi-objectifs est d'obtenir l'ensemble optimal de Pareto et, par conséquent, le front de Pareto (solutions non dominées). La première propriété garantit la génération de solutions de Pareto quasi-optimales, tandis que la deuxième propriété indique une bonne distribution des solutions obtenues autour du front optimal de Pareto, de sorte qu'aucune information précieuse n'est perdue (Emmerich and Deutz, 2018).

Les solutions placées sur le front de Pareto ne peuvent pas être comparées, aucune n'étant systématiquement meilleure que les autres sur tous les objectifs. Le choix de la solution finale est laissé au décideur. De nombreux algorithmes génétiques ont été développés pour résoudre ce type de problème, le plus populaire étant l'algorithme NSGA-II (Deb et al., 2002).

2.3 Vocabulaire et définitions

Definition 1 (Fonction objectif).

Appeler aussi fonction de coût ou critère d'optimisation. Il s'agit de la fonction que l'algorithme d'optimisation va essayer "d'optimiser" (rechercher un optimum); c'est à dire une fonction sur les variables de décision à maximiser ou à minimiser (Chong and Zak, 2013).

Definition 2 (Variables de décision).

Une quantité inconnue pour un modèle d'optimisation qui représente une décision à prendre. Il s'agit de contraintes qui peuvent être des variables continues (nombres réels) ou discrètes (valeurs d'un ensemble discret) (Talbi, 2009).

Definition 3 (Contrainte).

Toute restriction, exigence ou interaction qui limite les valeurs des variables de décision, également appelée contrainte générale (Chong and Zak, 2013).

Definition 4 (Solution optimale).

Pour un modèle d'optimisation avec une fonction objectif f , une solution $x \in \mathbb{R}^n$ est optimal si et seulement si pour toutes les solutions réalisables y , $f(x) \geq f(y)$ pour un problème de maximisation ou $f(x) \leq f(y)$ pour un problème de minimisation (Talbi, 2009).

Definition 5 (Optimum local/Optimum global).

Un **optimum local** d'un problème d'optimisation est une solution qui est optimale (maximum ou minimum) dans un ensemble voisin de solutions candidates. Cela contraste avec un **optimum global**, qui est la solution optimale parmi toutes les solutions possibles, et pas seulement celles qui se trouvent dans un voisinage particulier de valeurs (Talbi, 2009). Ces deux notions sont illustrées dans la Figure 2.2.

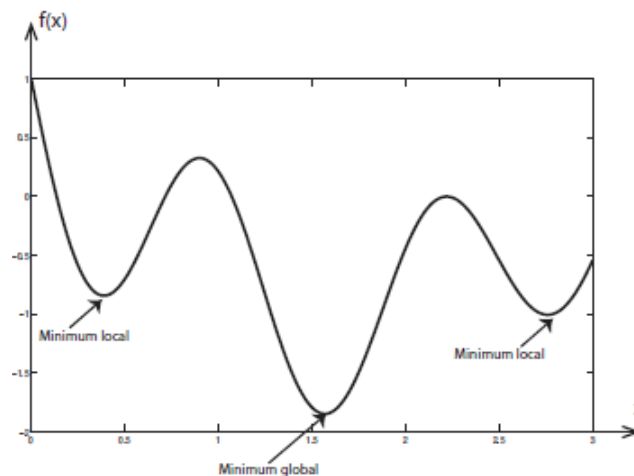


Figure 2.2. - Différence entre un optimum global et des optima locaux

2.4 Algorithmes d'optimisation

Plusieurs méthodes de résolution ont été proposées pour traiter les problèmes d'optimisation. Les méthodes existantes peuvent être classées en différentes classes selon leur efficacité, leur applicabilité et la complexité du problème. Dans ce qui suit nous nous concentrerons sur deux classes les heuristique et les métaheuristique :

➤ Heuristiques

Les méthodes heuristiques se sont révélées être un outil complet pour résoudre des problèmes d'optimisation difficiles ; elles fournissent un équilibre entre les "bonnes" solutions (relativement proches de l'optimum global) et des coûts et des délais abordables (Silver, 2004). Les principaux avantages des algorithmes heuristiques sont qu'ils sont (souvent) conceptuellement plus simples et (presque toujours) beaucoup moins coûteux en termes de calcul.

Parmi les méthodes heuristiques, il est possible de distinguer les heuristiques constructives (aussi appelées algorithmes glouttons) et les méthodes de recherche locale (Nazari-Heris et al., 2018). Cependant, les heuristiques sont généralement basées sur des caractéristiques spécifiques du problème à résoudre, ce qui rend leur conception et leur développement une tâche complexe. Pour résoudre cet inconvénient, les métaheuristiques apparaissent comme une avancée significative.

➤ Métaheuristique

Le mot «métaheuristique» est dérivé de deux mots grecs : «meta» signifie «au-delà» («à un niveau supérieur»), et heuristique signifie «trouver».

Le terme «métaheuristique» a été proposé pour la première fois par Glover (Glover, 1986). Le mot «métaheuristique» contient toutes les méthodes heuristiques qui montrent des preuves d'obtention de solutions de bonne qualité au problème d'intérêt dans un délai acceptable.

Osman et al. (Osman and Laporte, 1996), définit formellement la métaheuristique comme un processus de génération itératif, qui guide une heuristique subordonnée en combinant intelligemment différents concepts pour explorer et exploiter l'espace de recherche.

Selon (Zavala et al., 2014) *“les métaheuristiques sont une famille de méthodes d'optimisation non exactes pour trouver des solutions de haute qualité à des problèmes d'optimisation”*

complexes qui ne peuvent pas être résolus efficacement par des techniques exactes”.

Blum et Roli (Blum and Roli, 2003), ont examiné les propriétés fondamentales des métaheuristiques. Certaines de ces propriétés sont :

- Les métaheuristiques sont des stratégies qui «guident» le processus de recherche.
- L’objectif est d’explorer efficacement l’espace de recherche pour trouver des solutions (presque) optimales.
- Les techniques métaheuristiques vont des simples procédures de recherche locale à des processus d’apprentissage complexes.
- Les algorithmes métaheuristiques sont généralement approximatifs.
- Les concepts de base des métaheuristiques permettent une description au niveau abstrait.
- Les métaheuristiques ne sont pas destinées à un problème spécifique.
- Les métaheuristiques utilisent des connaissances spécifiques au domaine sous la forme d’heuristiques qui sont contrôlées par la stratégie de niveau supérieur.
- Les métaheuristiques avancées utilisent l’expérience de recherche pour guider la recherche.

Ainsi, nous pouvons résumer que les métaheuristiques ne sont pas spécifiques à un problème, mais elles peuvent utiliser des connaissances spécifiques au domaine sous la forme d’heuristiques contrôlées par une stratégie de niveau supérieur. Elles sont approximatives et généralement non déterministes, et elles sont généralement appliquées à des problèmes pour lesquels il n’existe pas d’heuristique satisfaisante spécifique au problème ; ou lorsqu’il n’est pas pratique de mettre en œuvre une telle méthode.

La métaheuristique est un domaine de recherche actif, avec une littérature considérable, une grande communauté de chercheurs et d’utilisateurs, et un large éventail d’applications. Une bibliographie sur les développements en métaheuristique a été présentée dans (Sotoudeh-Anvari and Hafezalkotob, 2018). Cet ouvrage répertorie et catégorise 112 articles sur les métaheuristiques. Et récemment le livre (Malik et al., 2021), rassemble des contributions sur l’état de l’art qui discute des développements récents des métaheuristiques et mettent en évidence certaines idées qui se sont avérées efficaces pour adapter une métaheuristique à un problème spécifique.

2.5 Classification des métaheuristiques

Dans la littérature, il existe plusieurs méthodes de métaheuristiques qui sont généralement classées en deux classes : les méthodes basées sur une solution unique et les

méthodes basées sur la population (Gendreau and Potvin, 2005). Dans la première classe de méthodes, la recherche invoque une solution unique à la fin de chaque itération à partir de laquelle la recherche commencera à l'itération suivante. D'autre part, les méthodes basées sur la population invoquent un ensemble de nombreuses solutions à la fin de chaque itération.

En général, les métaheuristiques à base de solution unique sont orientées vers l'exploitation de l'espace de recherche, tandis que les métaheuristiques à base de population de solutions sont orientées vers l'exploration et permettent une meilleure diversification de l'espace de recherche. Pour plus de détails sur les métaheuristique d'optimisation voir (Chopard and Tomassini, 2018). Ci-dessous, nous mettons en évidence les principes de quelques algorithmes pour les deux types de métaheuristique.

2.5.1 Métaheuristique à base de solution unique

Dans cette section, nous décrivons quelques métaheuristiques basées sur une solution unique, également appelées méthodes de trajectoire. Elles partent d'une seule solution initiale et s'en éloignent, décrivant une trajectoire dans l'espace de recherche.

1) Les méthodes de la descente (Hill Climbing)

Les algorithmes d'escalade (ou de descente) sont une famille de métaheuristiques de recherche locale simple (Papadimitriou, 1976). Comme avec de nombreuses méthodes de recherche locale, les algorithmes Hill Climbing ont tendance à ne conserver qu'un seul individu tout au long de la procédure de recherche. Bien qu'il en existe de nombreuses variétés, dont beaucoup ont été développées pour résoudre les problèmes rencontrés avec d'autres variantes, le fonctionnement général de ces algorithmes est assez similaire et a tendance à différer principalement dans la façon dont les solutions candidates sont échantillonnées. La descente, s'il s'agit d'une minimisation, à chaque étape de la recherche, cette méthode progresse vers une meilleure solution voisine. Elle s'arrête lorsque tous les candidats voisins sont pires que la solution actuelle ; c'est-à-dire lorsqu'un optimum local est atteint. Des expérimentations démontrent que la génération de voisins le long d'une seule dimension, c'est-à-dire directement perpendiculaire, s'avère être une stratégie plus efficace que ceux générés par perturbation sur plusieurs dimensions.

(Forrest and Mitchell, 1993) Compte tenu de cette description, les méthodologies d'escalade peuvent être considérées comme composées de deux phases principales : (i) une fonction escalier, qui définit le voisinage de la solution actuelle à rechercher et (ii) une stratégie d'acceptation déterminant quelles solutions doivent être acceptées comme nouvelle solution courante. Nous distinguons différents types de descente en fonction de la

stratégie de génération de la solution de départ et du parcours du voisinage : descente déterministe, descente stochastique et descente vers le premier meilleur.

L'extension des algorithmes de base à la recherche de voisinages locaux multiples fournit la diversification nécessaire pour éviter le problème de l'enfermement dans les optima locaux. Des approches telles que : l'escalade à redémarrage aléatoire, la recherche par faisceau local et la recherche par faisceau stochastique peuvent fournir une diversification grâce à la répétition du processus de recherche - accomplie par divers moyens, chacun ayant des propriétés différentes (Lourenço et al., 2003).

Le pseudo code de l'escalade simple est montré dans l'Algorithme suivant :

Algorithm 1 Algorithme Simple de Hill Climbing

```
SolutionActuelle ← Solution Aléatoire
Terminer = Faux                                     ▷ /*Score=Fonction objectif*/
Tant que Terminer == Faux Faire
    SolutionCandidate ← NextNeighbour(SolutionActuelle)
    Si Score(SolutionCandidate) < Score(SolutionActuelle) Alors
        SolutionActuelle ← SolutionCandidate
    Fin Si
    Si Score(SolutionActuelle) est suffisant Alors
        Terminer = vrai
    Fin Si
Fin Tant que
Renvoie la solution actuelle. (Retourner la meilleure solution découverte)
```

2) Le Recuit Simulé

L'algorithme de recuit simulé a été proposé par Kirkpatrick et al. (Kirkpatrick et al., 1983; Černý, 1985). Le recuit simulé est une méthode empirique inspirée d'un processus utilisé en métallurgie (appelé le recuit) où, afin d'atteindre des états de basse énergie d'un solide, celui-ci est chauffé à des températures élevées avant d'être laissé refroidir lentement.

La description classique du recuit simulé le présente comme un algorithme probabiliste, où un point évolue dans l'espace de recherche. Le recuit simulé basé sur l'algorithme de Metropolis (Hastings, 1970; Metropolis et al., 1953), qui permet de décrire l'évolution d'un système en thermodynamique.

Le principe de l'algorithme de Metropolis consiste à itérer les deux étapes suivantes :

- Évaluer la variation d'énergie associée à une transition élémentaire aléatoire de l'état courant i , d'énergie E_i , vers un nouvel état j , d'énergie E_j : $\Delta E_{ij} = E_j - E_i$;
- Accepter la transition vers le nouvel état avec une probabilité P_{ij} où :

$$\begin{cases} P_{ij}(T) = 1 & \text{si } \Delta E_{ij} \leq 0 \\ P_{ij}(T) = \exp(-\frac{\Delta E}{T}) & \text{si } \Delta E_{ij} > 0 \end{cases} \quad (5)$$

Cette procédure permet de sortir des minima locaux avec une forte probabilité si la température T est élevée et, lorsque l'algorithme atteint des températures basses, de conserver les états les plus probables.

L'algorithme de recuit simulé est résumé ci-dessous.

Algorithm 2 Algorithme du Recuit simulé

Déterminer une fonction aléatoire S

Choix des mécanismes de perturbation d'une configuration

Initialiser la température T

Tant que le critère d'arrêt n'est pas satisfait **Faire**

Tant que l'équilibre n'est pas atteint **Faire**

 Tirer une nouvelle configuration S'

 Appliquer la règle de Metropolis

Si $f(S') < f(S)$ **Alors**

$S_{min} = S'$

$f_{min} = f(S')$

Fin Si

Fin Tant que

 Décroître la température

Fin Tant que

Le recuit simulé de base dispose deux problèmes ; le réglage de plusieurs paramètres et le temps de calcul qui est très importants. Cependant, il existe des études qui se concentrent sur le réglage optimale des paramètres de l'algorithme (Courat et al., 1994). En outre, pour surmonter le problème du temps de calcul, plusieurs méthodes de parallélisation des calculs ont été introduites (Azencott, 1992). D'autre part, la méthode du recuit simulé a l'avantage d'être flexible vis-à-vis de l'évolution du problème et facile à mettre en oeuvre.

3) La méthode GRASP

La procédure de recherche gloutonne aléatoire adaptative (GRASP : Greedy Randomized Adaptive Search Procedure), proposée par Feo et Resende dans (Feo and Resende, 1995, 1989), est une métaheuristique à départs multiples, dépourvue de mémoire, fondée sur les algorithmes gloutons randomisés et les techniques de recherche de voisinage. Chaque itération de l'algorithme GRASP se compose de deux étapes appelées construction et recherche locale. L'étape de la construction est similaire à l'heuristique semi-gloutonne (semi-greedy heuristic) proposée par Hart et Shogan (Hart and Shogan, 1987). Elle génère une solution réalisable par l'application d'une procédure d'initialisation gloutonne. Dans la deuxième étape, cette solution est utilisée comme solution initiale de la procédure de recherche locale. Après un nombre donné d'itérations, l'algorithme GRASP se termine et la meilleure solution trouvée est conservée. Une étude bibliographique de la méthode GRASP a été fournie dans (Festa and Resende, 2009a,b). Récemment dans (Resende and Ribeiro, 2019) une étude reflétant des applications réussies de GRASP à un grand nombre de problèmes classiques d'optimisation combinatoire, ainsi qu'à des problèmes qui surviennent dans des situations réelles dans différents domaines de la science et de la technologie.

L'implémentation de la méthode GRASP est simple, ainsi que leur extensions constituent un ensemble d'outils qui peuvent être ajoutés à des métaheuristicues plus simples pour trouver de meilleures solutions.

Le pseudo code de l'algorithme GRASP de minimisation est le suivant (où : Seed est utilisé comme graine initiale pour le générateur de nombres pseudo-aléatoires) :

Algorithm 3 GRASP

```
Procédure GRASP(Max-Itérations, Seed)
  Lire Entrée();
  Pour k = 1 : Nombre maximum d'itérations Faire
    Solution ← Construction aléatoire Greedy (Seed);
    Si la Solution n'est pas faisable Alors
      Solution ← Réparation (solution);
    Fin Si
    Solution ← Recherche locale (Solution);
    Mise à jours (Solution, MeilleureSolution);
  Fin Pour
  Retourner MeilleureSolution;
Fin Procédure
```

4) La Recherche Tabou

Cette méthode a été développée par Glover (Glover, 1990, 1986; Glover and Laguna, 1997). Elle utilise une mémoire à court terme, appelée liste tabou, qui garde une trace des solutions les plus récemment visitées et interdit tout mouvement vers celle-ci, ce qui permet de surmonter les minima locaux de la fonction objectif. La liste tabou empêche les mouvements (solution) cycliques et empêche également les mouvements d'être inversés.

L'algorithme de recherche tabou commence avec une solution initiale (sélectionnée au hasard), il maintient une liste tabou qui contient un enregistrement des solutions précédemment visitées. Tout mouvement enregistré dans la liste tabou n'est pas autorisé. A chaque itération, de nouvelles solutions candidates sont générées. Une nouvelle solution est acceptée, conformément au contenu de la liste tabou, si elle est meilleure que la solution actuelle.

S'il n'existe pas de meilleure solution, le meilleur voisin est sélectionné pour remplacer la solution existante. Les pires solutions sont également acceptées pour sortir des optima locaux. Les mémoires à moyen et long terme peuvent également être utilisées pour améliorer l'intensification et la diversification (Talbi, 2009). La première stocke la meilleure solution obtenue lors de la recherche, et la seconde aide à explorer des zones inexplorées. Les critères d'aspiration sont une autre fonctionnalité qui permet de remplacer la liste tabou et d'autoriser les déplacements tabou, s'ils aboutissent à de meilleures solutions. L'algorithme se termine si les ensembles autorisés sont vides ou si la condition de terminaison est remplie. L'article (Prajapati et al., 2020) présente une enquête complète sur les variantes de l'algorithme de recherche tabou, leur principales caractéristiques et leur comportement.

Le pseudo-code de cette méthode est représenté dans l'algorithme suivant.

Algorithm 4 Algorithme du Recherche Tabou

Init $L = \Phi$ (liste des points tabous)

Init x (point de départ)

Init k_{fin} (nb d'itérations max total)

Init $k = 0, l = 0$

Répéter

$y = Voisinage(x) - L$ (on prend un voisinage sans point tabous)

$\Delta C = C(y) - C(x)$

$x = y$ avec $y \in Voisinage(x) - L$ tel que $\min C(y) - C(x)$

$L = L + \{x\}$

$k = k + 1$

Update L (on retire des mouvements tabou de L suivant des critères d'aspiration)

Jusqu'à $k = k_{fin}$

2.5.2 Métaheuristiques basée sur la population

Les métaheuristiques à base de solution unique visent l'intensification. Elles sont pourvues pour explorer les domaines prometteurs de l'espace de recherche. Cependant, il s'agit d'algorithmes de recherche locale et ils ne sont donc pas capables d'explorer efficacement l'espace de recherche. Cet inconvénient est surmonté par les méthodes basées sur la population. Elles itèrent sur un ensemble de solutions, ce qui conduit à une meilleure couverture et exploration de l'espace de recherche.

Les métaheuristiques basées sur la population traitent d'un ensemble (c'est-à-dire une population) de solutions plutôt qu'une solution unique. Elles améliorent, au fil des itérations, une population de solutions. L'intérêt de ces méthodes est d'utiliser la population comme facteur de diversité (pour plus de détails, veuillez consulter l'ouvrage (Patel et al., 2019)).

La majorité des métaheuristiques de ce type sont basées sur les principes de l'évolution biologique. En particulier, elles simulent diverses métaphores biologiques qui diffèrent par la nature des schémas de représentation (structure, composants, etc.). Il existe trois principaux paradigmes : les algorithmes évolutionnaires, en essaim et immunitaire (Abdel-Basset et al., 2018).

1) Les algorithmes évolutionnaires

Les algorithmes évolutionnaires simulent la progression biologique de l'évolution au niveau cellulaire en utilisant des opérateurs de sélection, de croisement, de mutation et de reproduction pour générer des solutions candidates de plus en plus performantes. Les

algorithmes les plus populaires sont :

➤ **Algorithme génétique**

L'algorithme génétique (AG) a été inventé par John Holland et ses collègues au début des années 1970, inspiré par la théorie de Darwin. L'idée derrière AG est de modéliser l'évolution naturelle en utilisant l'héritage génétique avec la théorie de Darwin.

Les algorithmes génétiques sont les méthodes les plus connues et les plus robustes. L'étape la plus importante dans la conception d'un AG est de représenter la solution avec le chromosome. Un chromosome doit être capable de refléter les caractéristiques d'un problème et de les représenter correctement, et de produire une solution plus appropriée pour la fonction objectif, à travers un processus évolutif, par un opérateur génétique. Pour résoudre un problème d'optimisation, l'algorithme génétique commence par la représentation chromosomique d'un ensemble de paramètres, qui doit être codée sous forme de chaîne de taille finie sur un alphabet de longueur finie. Habituellement, les chromosomes sont des chaînes de 0 et de 1.

Chaque chromosome renvoie en fait à une solution codée possible. Un ensemble de tels chromosomes dans une génération est appelé une population, dont la taille peut être constante ou peut varier d'une génération à l'autre.

La fonction fitness/objectif est choisie en fonction du problème à résoudre, de sorte que les chaînes (solutions possibles) représentant les bons points dans l'espace de recherche aient une valeur fitness élevée. Les opérateurs génétiques fréquemment utilisés sont la sélection, le croisement et la mutation. Ils sont appliqués à une population de chromosomes pour produire potentiellement une nouvelle progéniture.

L'AG émule le processus biologique d'évolution des chromosomes en définissant les opérateurs suivants : sélection, croisement et mutation. Les chromosomes sont traités comme des solutions candidates pour un problème donné et sont évalués en fonction de leur fitness. La sélection des parents pour la reproduction est un processus influent pour la génération de nouvelles solutions. Il existe différents schémas de sélection tels que la sélection par roulette, la sélection par tournoi, la sélection par rang linéaire et la sélection par troncature, etc.

Pendant la phase de croisement, les parties de deux chromosomes sélectionnés sont permuées. Les parties des chromosomes peuvent être échangées de différentes manières, telles que le croisement à un, deux et uniforme, etc.

Dans la mutation, des parties de chromosomes sont modifiées de manière aléatoire afin d'échapper aux optima locaux. Cependant, les meilleurs chromosomes peuvent être perdus lors de la création de nouveaux chromosomes. Pour éviter cela, l'élitisme est utilisé pour copier le meilleur chromosome (ou quelques meilleurs chromosomes) dans la nouvelle population.

Algorithm 5 Pseudo code de l'algorithme génétique

```
Générer une population aléatoire initiale P d'individus
gen = 0
Tant que gen < GenMax Faire
    Évaluer la fitness de chaque individu
    Q = Selection(P)
    IntPop = Croisement(Q)
    PopFils = Mutation(IntPop)
    Évaluer la fitness de PopFils
    new_Pop = Remplacer les mauvaises individus par les meilleurs nouveaux
    gen = gen + 1
Fin Tant que
```

➤ Programmation génétique

La programmation génétique (GP) (Koza, 1992) est un algorithme d'optimisation évolutionnaire qui utilise les principes de la théorie de l'évolution de Darwinienne et des opérateurs d'inspiration biologique afin de créer des solutions de haut niveau à des problèmes particuliers. La GP est connue comme un exemple de recherche réussie en intelligence artificielle et en apprentissage automatique, et a été étudiée dans une grande variété de domaines de connaissance, tels que l'exploration de données, les tâches d'optimisation, etc.

La GP est lié à GA, où elle génère automatiquement des programmes informatiques. Chaque programme informatique individuel est une solution au problème d'optimisation.

Néanmoins, il existe des différences fondamentales entre la GP et les algorithmes génétiques (AG) standard ; les algorithmes génétiques utilisent des chromosomes qui codent les points de conception dans un format séquentiel par contre la programmation génétique représente les individus en utilisant plutôt des arbres (de gènes) qui représentent mieux les fonctions mathématiques, les programmes, les arbres de décision et d'autres structures hiérarchiques. Un arbre est composé de nœuds terminaux et de fonctions ; les

nœuds terminaux représentent une valeur constante, tandis que les nœuds de fonction sont les opérateurs mathématiques appliqués aux nœuds terminaux afin d'évaluer les arbres.

Semblables aux algorithmes génétiques, au cours du processus évolutif des programmes génétiques, plusieurs opérations sont effectuées sur la population actuelle pour produire un nouvel ensemble d'individus plus adaptés, telles que la sélection, la reproduction, la mutation et le croisement.

La mise en œuvre de la programmation génétique est identique à celle des algorithmes génétiques. Cependant, il faut généralement faire plus attention lors de la mise en œuvre des routines de croisement et de mutation, en particulier lorsqu'il s'agit de déterminer quels types de nœuds peuvent être générés et que seuls des arbres syntaxiquement corrects sont produits. En particulier, l'objectif de l'algorithme GP est de sélectionner, par recombinaison de gènes, le programme le plus prépondérant résolvant un dilemme donné.

2) Les algorithmes en essaim

Ils imitent le comportement collectif des agents d'une communauté, tels que les oiseaux et les insectes. Ils dépendent principalement du principe de décentralisation, c'est-à-dire que les solutions candidates sont mises à jour grâce à une interaction locale entre elles et avec leur environnement. Les algorithmes les plus populaires :

➤ Les colonies de fourmis

L'optimisation par colonies de fourmis (ACO : Ant Colony Optimization), conçue par Dorigo (Dorigo and Gambardella, 1997) s'inspire comme son nom l'indique du comportement des fourmis lorsqu'elles cherchent de la nourriture et optimisent le chemin entre leur nid et la nourriture trouvée (Deneubourg et al., 1990). En effet, les fourmis utilisent leur environnement pour communiquer entre elles, elles déposent des phéromones sur le sol pour indiquer aux autres fourmis le chemin qu'elles ont parcouru pour atteindre la nourriture. Ainsi, les autres fourmis peuvent suivre la piste des phéromones pour trouver la source de nourriture. Cependant, il s'avère que les phéromones s'évaporent avec le temps, par conséquent ce sont les chemins les plus courts qui conserveront une plus grande concentration de phéromones.

C'est comme cela que les fourmis trouvent naturellement le chemin le plus court vers leur nourriture depuis leurs abris. L'ACO reprend la notion de système multi-agents dans lequel chaque agent est représenté par une fourmi. Cela peut être utilisé par exemple pour parcourir un graphe : une fourmi parcourt le graphe de manière aléatoire, mais avec une probabilité plus élevée de suivre une arête du graphe, en fonction de la quantité de

phéromones déposée dessus. Lorsque le graphe est entièrement parcouru, elle laisse sur le chemin qu'elle a pris une quantité de phéromones proportionnelle à la longueur de ce chemin.

L'algorithme ACO de base se compose de quatre étapes principales selon l'algorithme suivant :

Algorithm 6 Algorithme de base d'ACO

Initialisation des paramètres et les variables de phéromones ;

Tant que la condition de terminaison n'est pas atteinte **Faire**

 Construire des solutions de fourmis ;

 Méthode de recherche locale ;

 Mise à jour globale sur les phéromones ;

Fin Tant que

Dans (Akhtar, 2019), un bref aperçu des développements récents des algorithmes ACO en termes d'applications et de développements algorithmiques.

➤ Algorithmes à essaim de particules

L'optimisation par essaim de particules (PSO) est une technique de calcul évolutive pour résoudre des problèmes d'optimisation globale développée par Eberhart et Kennedy (Eberhart and Kennedy, 1995). Il s'agit d'une technique de calcul fonctionnant grâce à l'amélioration individuelle ainsi que par la coopération et la compétition de la population, qui sont basées sur la simulation de modèles sociaux simplifiés, tels que le vol des oiseaux, le banc de poissons et la théorie de l'essaim (Injeti and Thunuguntla, 2020; Ye et al., 2009). Étant donné que PSO a fait preuve d'une excellente flexibilité et fonctionnalité dans la résolution de divers problèmes d'optimisation complexes (Elsheikh and Abd Elaziz, 2019).

Chaque individu, ou particule, de la population garde une trace de sa position actuelle, de sa vitesse et de la meilleure position qu'il a vue jusqu'à présent. La quantité de mouvement permet à un individu d'accumuler de la vitesse dans une direction favorable, indépendamment des perturbations locales.

À chaque itération, chaque individu est accéléré en même temps vers la meilleure position qu'il a vue et vers la meilleure position trouvée jusqu'à présent par tout individu. L'accélération est pondérée par un terme aléatoire, des nombres aléatoires séparés étant générés pour chaque accélération. Le pseudo-code de l'algorithme PSO est donné ci-dessous :

Algorithm 7 Pseudo Code de l'algorithme PSO

Initialiser les paramètres de l'algorithme

Initialiser un particule

Évaluez les valeurs de la fonction objective de toutes les particules

Pour $i = 1$ à F **Faire**

 Pour chaque ensemble de particules (la meilleure position jamais rencontrée par chaque particule) et sa valeur de fitness correspondante

 Pour l'ensemble de particules complet (meilleure position globale) et sa valeur de fitness correspondante atteinte par l'essaim

 Mettez à jour la vitesse et la position de chaque particule

 Évaluez les valeurs de fitness de toutes les particules pour la position mise à jour

 Pour chaque particule, comparez sa valeur de fitness actuelle avec la meilleure valeur de fitness qu'elle a atteinte jusqu'à présent. Si la valeur actuelle est meilleure, choisissez la position de mise à jour

 Déterminer la meilleure particule de la population actuelle. Si la valeur de fitness actuelle est la meilleure, alors choisissez la position mise à jour

 Utiliser la nouvelle population générée pour une nouvelle exécution de l'algorithme

Fin

Fin Pour

A la terminaison, la valeur optimale du problème considéré contenant la meilleure solution trouvée

De nombreuses variantes améliorées sont proposées, par exemple, le PSO parallèle (PPSO) (Ma et al., 2016), des coefficients d'accélération variables dans le temps PSO (TVACPSO) (Jordehi, 2016), PSO leader amélioré (ELPSO) (Jordehi, 2018), PSO à convergence garantie (GCPSO) (Nunes et al., 2018), PSO flexible (FPSO) (Ebrahimi et al., 2019).

2.5.3 Les algorithmes immunitaires artificiels

Ils s'inspirent de l'immunologie théorique et des fonctions, principes et modèles immunitaires observés. Ils peuvent être considérés comme une autre variante algorithmique des algorithmes évolutionnaires. Lorsqu'ils sont appliqués à l'optimisation, les anticorps représentent des solutions candidates qui ont évolué de manière itérative en répétant les opérateurs de clonage, de mutation et de sélection. L'antigène représente une fonction objective et les bonnes solutions sont stockées dans une cellule mémoire. Presque toutes les métaheuristiques de cette catégorie dépendent des principes de sélection clonale.

➤ Algorithme de sélection clonale (CLONALG)

Brent (Burnet, 1959) a introduit pour la première fois la théorie de la sélection clonale en 1959 afin de discuter de la réponse fondamentale du système immunitaire adaptatif (lymphocytes) à un stimulus antigénique. Il a confirmé que seules les cellules capables d'identifier un antigène prolifèrent, tandis que celles qui n'identifient pas d'antigène sont contre-sélectionnées. CLONALG est la version d'optimisation des paradigmes de sélection clonale car le système immunitaire naturel peut avoir des objectifs multiples ou contradictoires et n'a aucune raison de développer une réponse optimale. Comme mentionné précédemment, une population d'anticorps (solutions candidates) est générée de manière aléatoire et évaluée. Ensuite, les anticorps présentant des affinités plus élevées sont clonés afin de générer plus d'anticorps contre l'antigène. Ces anticorps non clonés sont remplacés par de nouveaux. Pour acquérir une immunité, les meilleures solutions sont conservées dans une cellule mémoire. Le CLONALG (LN and FJ, 2002) est l'un des algorithmes les plus populaires, son pseudo code est le suivant :

Algorithm 8 Pseudo code de CLONALG

```
Générer la population initiale au hasard ;
Évaluez chaque solution ;
Tant que le critère de terminaison n'est pas satisfait Faire
    Évaluez chaque solution ;
    Sélectionnez les meilleurs anticorps ;
    Cloner ;
    Subir une mutation ;
    Gardez les meilleurs anticorps actuels ;
    Remplacer les anticorps non clonés par de nouveaux ;
Fin Tant que
```

2.6 Comparaison des métaheuristiques

La comparaison des métaheuristiques entre elles est une étape importante dans le développement des métaheuristiques afin de déterminer où se situe une nouvelle approche par rapport aux autres en termes de diverses mesures de performance et d'adéquation à certaines situations. Basant sur (Silberholz et al., 2019), plusieurs considérations importantes doivent être prises en compte afin de comparer efficacement les métaheuristiques entre elles de manière significative.

Cependant, nous pouvons résumer les avantages et les inconvénients spécifiques à quelque métaheuristiques à base de population, que nous avons extraites des articles référencés.

Le Tableau 2.1 présente une telle comparaison :

Tableau 2.1. - Comparaison des métaheuristiques basées sur la population

Algorithme	Avantages	Inconvénients
Algorithme génétique	<ul style="list-style-type: none"> • Facile à mettre en œuvre. • la capacité de gérer des types aléatoires d'objectifs et de contraintes. • Il peut être utilisé indépendamment pour résoudre un problème donné. • Adaptation pour le traitement de problèmes avec des contraintes et des fonctions objectifs non linéaires ou discontinues. • Il utilise des opérateurs simples et peut être utilisé pour résoudre des problèmes d'une grande complexité de calcul. 	<ul style="list-style-type: none"> • Il a une probabilité d'être piégé dans les optima locaux. • N'a pas de critères de terminaison standard, ni de méthode standard pour ajuster ses paramètres. • Temps élevé, notamment pour les problèmes avec un grand nombre de variables.
Programmation génétique	<ul style="list-style-type: none"> • La compréhensibilité, • La flexibilité, • La diversité de l'espace de recherche, • Et la flexibilité de la fonction de Fitness 	<ul style="list-style-type: none"> • Elle nécessite de grandes données d'apprentissage. • En raison de la nature stochastique de la GP, elle ne garantit pas une solution exacte.

<p>Optimisation des colonies de fourmis</p>	<ul style="list-style-type: none"> • Le processus de construction de l'ACO est intrinsèquement parallèle. • Il permet d'éviter une convergence prématurée. • Il convient aux applications dynamiques. 	<ul style="list-style-type: none"> • La distribution de probabilité change avec l'itération. • La convergence soit garantie mais le temps de convergence n'est pas défini. • Il est difficile d'analyser le comportement de manière théorique.
<p>Optimisation des essaims de particules</p>	<ul style="list-style-type: none"> • La simplicité. • Il est utile dans la recherche scientifique et en ingénierie. 	<ul style="list-style-type: none"> • Toutes les solutions convergent prématurément et perdent par conséquent la diversité de la population. • Il souffre d'un optimisme partiel.
<p>Algorithme de sélection clonale</p>	<ul style="list-style-type: none"> • Peut être très efficaces pour échapper aux optima locaux dans les problèmes d'optimisation. 	<ul style="list-style-type: none"> • Temps d'exécutions considérablement lent pendant la phase d'exploitation.

2.7 Conclusion

Dans ce chapitre, nous avons présenté le problème d'optimisation, et les algorithmes de résolution. Nous avons insisté sur les méthodes de métaheuristique pour les deux types : les méthodes basées sur une solution unique et les méthodes basées sur une population de solutions. En outre, nous avons souligné quelques exemples les plus courants dans la littérature pour les deux types.

Parmi les métaheuristicues existantes, la méthode de programmation génétique, et qui selon la littérature, présente de nombreux avantages à savoir la compréhensibilité, la flexibilité, la diversité de l'espace de recherche, la réduction du temps de test ou d'exé-

cution, et la flexibilité de la fonction de Fitness. Cette méthode a montré d'excellentes performances pour les problèmes liés à l'optimisation et à la régression.

Dans le chapitre suivant, nous allons développer le principe de cette méthode et plus particulièrement sa variante robuste : la Programmation Génétique Multi Gène.

Chapitre 3

Programmation Génétique MultiGène

3.1 Introduction

La programmation génétique GP est une technique d'optimisation inspirée de la théorie de l'évolution de Darwin. Il permet de découvrir et d'optimiser des modèles mathématiques et informatiques complets qui décrivent le mieux certains phénomènes souhaités en imitant le processus d'évolution de la nature. La GP possède plusieurs variantes, dont sa version robuste la Programmation Génétique Multi-Gène MGGP est introduite dans ce chapitre car elle constitue une partie centrale de cette thèse.

Dans ce chapitre, nous présenterons tout d'abord la régression symbolique qui est une méthode bien étudiée en programmation génétique pour découvrir des modèles mathématiques de forme libre à partir de données observées. Ensuite, nous exposerons les concepts de base de la programmation génétique et nous détaillerons sa variante MGGP. Enfin nous discuterons la relation des GP avec le traitement d'image.

3.2 Régression symbolique avec la Programmation génétique

La régression symbolique a été introduite pour la première fois par John R. Koza dans son livre (Koza, 1992). C'est une méthodologie permettant de générer automatiquement des modèles mathématiques qui décrivent les relations sur des données donné, et pas seulement un ensemble de coefficients d'un modèle prédéfini (comme dans la régression classique). Il s'agit d'une nette différence avec les techniques de régression classique dans lesquelles un modèle spécifique est prédéfini. L'objectif de la régression symbolique est de déterminer lesquelles des entrées sont les plus efficaces pour prédire les sorties et

d'identifier la relation entrée-sortie.

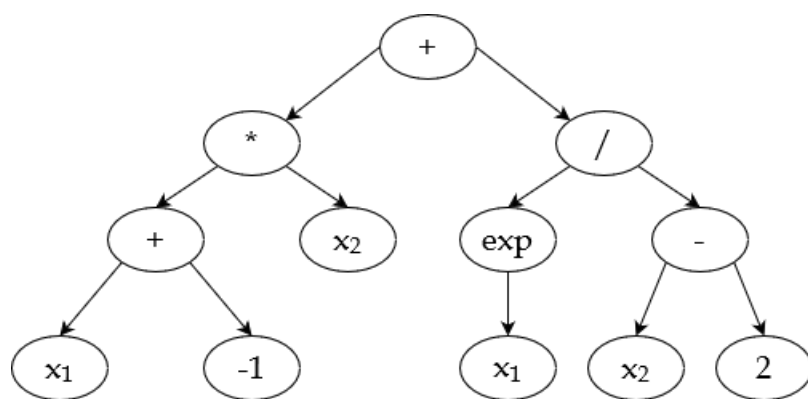
La régression symbolique basée sur la GP ne nécessite pas la forme et la taille prédéfinies des solutions; ce qui présente un énorme avantage pour le concepteur, car il supprime le besoin d'intuition spéculative de la part du concepteur pour choisir le modèle de régression approprié en fonction de la nature des données. La régression symbolique essay d'aboutir à des modèles de boîte blanche clairs à interpréter, il a un fort pouvoir expressif. Toutes ces propriétés font de GP l'approche la plus appropriée à la régression symbolique (Gustafson et al., 2005).

3.3 Programmation Génétique

La Programmation Génétique (GP) a été introduite pour la première fois par Mr. Koza (Koza, 1992). C'est une technique d'apprentissage automatique qui est une branche de l'intelligence artificielle. GP est aussi l'algorithme génétique appliqué aux programmes informatiques permettant de résoudre automatiquement divers problèmes complexes d'optimisation et de recherche.

L'article (Ahvanooy et al., 2019) passe en revue la littérature existante concernant la GP et leur applications dans différents domaines scientifiques et vise à fournir une compréhension facile des différents types de la GP.

La GP fournit une solution sous la forme d'une structure arborescente ou sous la forme d'une équation compacte utilisant l'ensemble de données donné. La Figure 3.1 montre un exemple d'arbre simple représentant une formule mathématique.



$$\Rightarrow ((x_1 - 1) * x_2) + (e^{x_1} / (x_2 - 2))$$

Figure 3.1. - Une formule mathématique représentée sous forme d'arbre

L'idée derrière la programmation génétique est de faire évoluer une population de solutions, c'est à dire une population des programmes informatiques. Chaque programme informatique est traité comme un individu avec sa propre structure génétique, tout comme une personne à son propre code génétique qui la sépare des autres personnes de la population. Par conséquent, dans la première étape, la population initiale est créée ; l'initialisation est un processus aléatoire et il existe plusieurs stratégies pour optimiser ce processus. Lorsque la population est créée, les individus doivent être déplacés vers la population pour la prochaine génération. Pour ce faire, la fitness de chaque solution doit être calculée puis référencée aux valeurs de fitness présentes dans la population. Ce faisant, les individus sélectionnés à l'étape suivante ont une meilleure fitness. Une fois les candidats sélectionnés, les opérateurs génétiques (sélectionnés avec la probabilité spécifiée en paramètre) leur sont appliqués. Il peut s'agir d'une simple réplique ou d'une sorte de variation (mélange de 2 individus ou modification de l'individu). Après avoir généré un individu, il est ajouté à la population pour la génération suivante. Le processus est répété jusqu'à ce que certaines des conditions d'arrêt soient remplies, comme la recherche d'une solution acceptable ou le dépassement du nombre spécifié de générations. Un pseudo code de la GP est présenté ci-dessous :

Algorithm 9 Pseudo code de la GP

Créer aléatoirement une population initiale de programmes

Répéter

 Calculer la fitness pour chaque programme dans une population

 Sélectionner un ou deux programme(s) parmi la population avec une probabilité basée sur la capacité de participer à des opérations génétiques

 Créer de nouveaux programmes en appliquant des opérations génétiques avec des probabilités spécifiées

Jusqu'à ce qu'une solution acceptable soit trouvée ou qu'une autre condition d'arrêt soit remplie (par exemple, atteindre un nombre maximal de générations)

Renvoie le meilleur individu

Une brève description des divers mécanismes évolutifs dans GP est présentée ci-dessous :

3.3.1 Population initiale

Dans la première étape de la programmation génétique, un certain nombre d'arbres sont générés en sélectionnant de manière aléatoire des fonctions et terminaux définis par l'utilisateur. Ces arbres forment la population initiale.

3.3.2 Reproduction

Dans la deuxième étape du GP, une proportion de la population initiale est sélectionnée et copiée à la génération suivante et cette procédure est appelée reproduction ; le mécanisme de reproduction ne produit pas de nouvelle population. Comme avec la plupart des algorithmes évolutionnaires, les opérateurs génétiques en GP sont appliqués à des individus sélectionnés de manière probabiliste en fonction de leur fitness. Autrement dit, les meilleurs individus sont plus susceptibles d'avoir plus de programmes enfants que les mauvaises individus. Ainsi, l'arbre GP avec une bonne fitness entre dans le creux d'accouplement et les autres meurent. Il existe différents opérateurs de sélection mais la sélection par tournoi est le mécanisme de sélection le plus largement utilisé en raison de sa grande efficacité et de sa facilité de mise en œuvre.

➤ Sélection par Tournoi

Dans cette stratégie, les gènes existants (parents) sont choisis au hasard dans la population la plus large, et les gènes sélectionnés se font concurrence (en fonction de la taille du tournoi). Le gène par la plus grande fitness est attribué comme l'un de la population de la prochaine génération. Cette stratégie peut contrôler facilement la pression de sélection en modifiant la taille du tournoi de sorte que si la taille du tournoi est supérieure à celle des gènes faibles, ils ont alors moins de chances d'être choisis. Il offre également la possibilité de choisir tous les gènes et conserve la diversité, bien que la préservation de la diversité puisse réduire la vitesse de convergence.

3.3.3 Croisement

Pour l'opération de croisement, deux arbres GP (Parent1 et Parent2) sont sélectionnés au hasard dans la population du pool d'accouplement. Un nœud de chaque arbre est sélectionné de manière aléatoire, les sous-arbres situés sous les nœuds sélectionnés sont échangés et deux descendants (Enfant1 et Enfant2) sont générés (Koza, 1992). Un exemple de fonctionnement de croisement est illustré à la Figure 3.2.

3.3.4 Mutation

Contrairement à l'opération de croisement, la mutation se concentre sur un parent unique et la modifie pour former le nouvel enfant. La mutation signifie l'évolution d'une structure complètement nouvelle déterminée au hasard au nœud sélectionné (c'est-à-dire que le sous-arbre du nœud choisi est supprimé et qu'un nouveau est créé au hasard, en "poussant" au point de mutation choisi). Ce changement aléatoire ou l'apparition de gènes, comme le montre la figure 3.3, permet à certains programmes de la nouvelle génération

de s'éloigner du chemin évolutif de la population générale afin que différents programmes puissent être examinés.

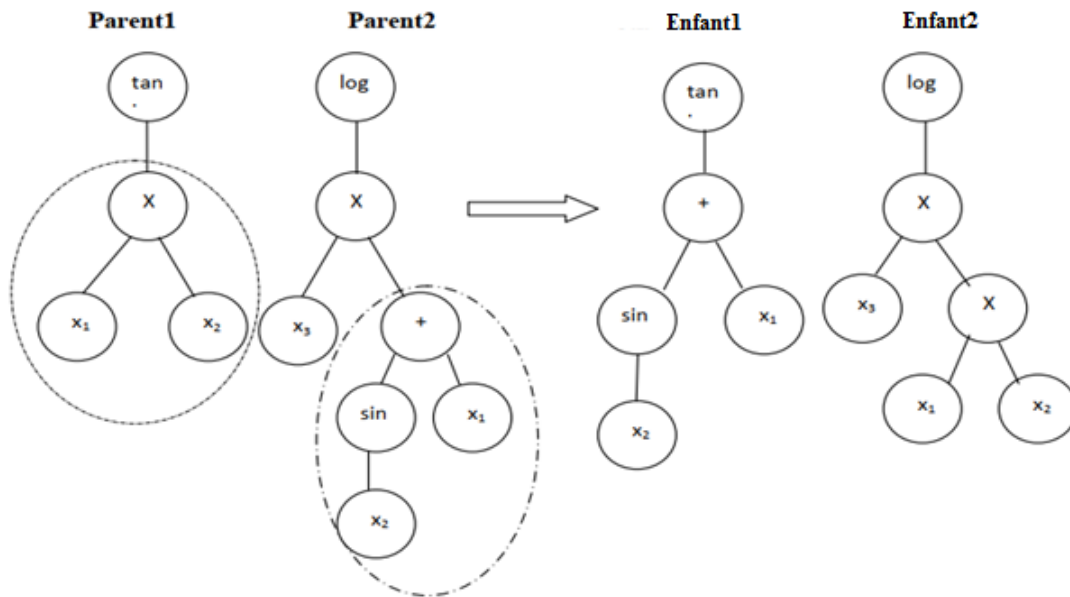


Figure 3.2. - Une opération de croisement en GP

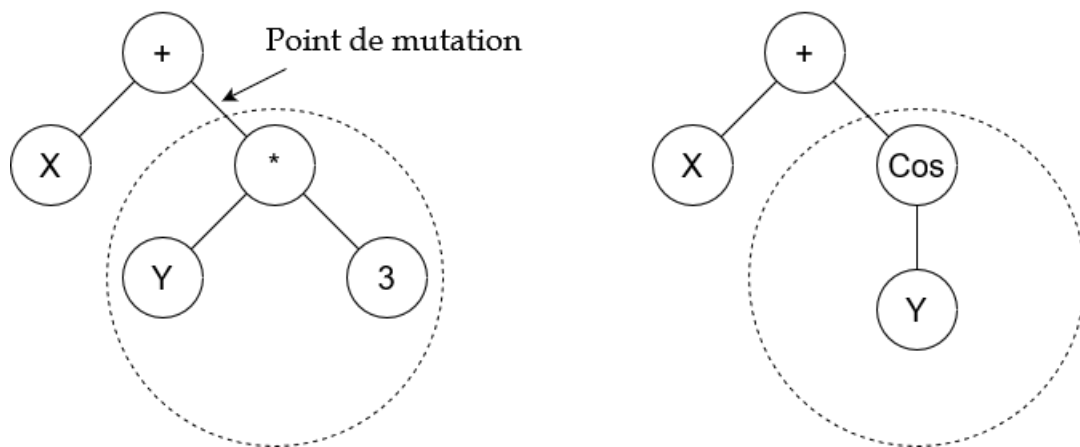


Figure 3.3. - Une opération de mutation en GP

En se référant à la littérature, de nombreuses différentes variantes de GP ont été proposées pour améliorer ces performances ; on trouve par exemple : STGP (Strongly-typed Genetic Programming) (Montana, 1995), GGGP (Grammar Guided Genetic Programming) (Wong and Leung, 1995), GEP (Gene Expression Programming) (Ferreira, 2001), GE (Grammatical Evolution) (O'Neill and Ryan, 2001), et GSGP (Geometric Semantic Genetic Programming) (Moraglio et al., 2012), et MGGP (Multi-Gene Genetic Program) (Gandomi and Alavi, 2012).

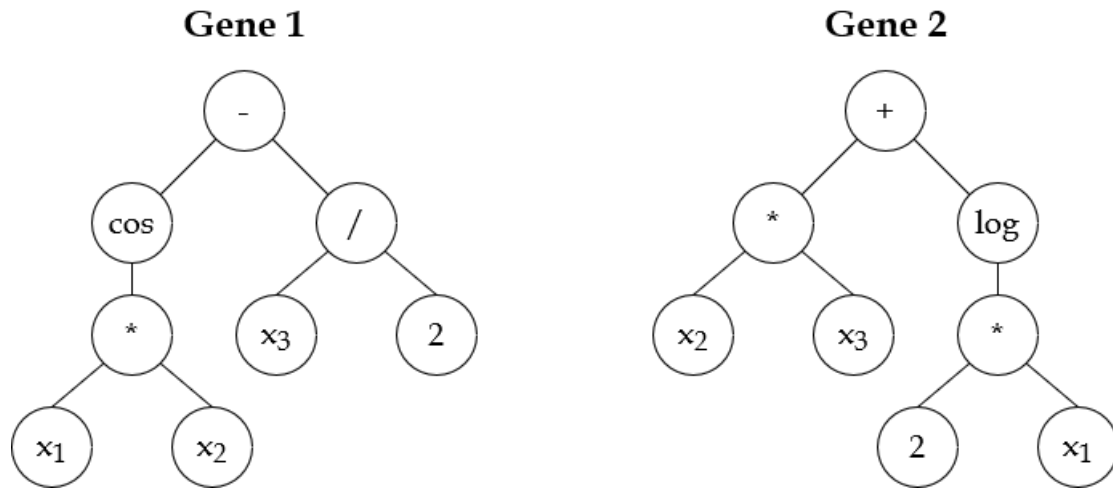
Cette dernière version (MGGP) est considérée comme une variante GP puissante qui accumule efficacement les capacités des solutions GP standard pour renforcer les compétences de prédiction du modèle de régression. Par conséquent, il est très approprié pour résoudre des problèmes de régression et de modélisation (Ari and Alagöz, 2021). Dans ce qui suit on va détailler sur la variante MGGP.

3.4 Programmation Génétique Multigène (MGGP)

MGGP est une variante de GP et est conçu pour développer un modèle mathématique empirique, qui est une combinaison linéaire pondérée d'un certain nombre d'arbres de GP et est appelé «gènes». «Multi-gène» fait référence à la combinaison linéaire de ces gènes (Gandomi and Alavi, 2012).

En pratique, le nombre maximal autorisé de gènes pour un modèle (G_{max}) et la profondeur maximale de l'arbre pour tout gène (D_{max}) peut avoir être spécifié par l'utilisateur. Un contrôle remarquable peut être exercé sur la complexité maximale des modèles développés par MGGP par rapport au GP standard. Cela se traduit généralement par l'évolution de modèles relativement compacts. Les modèles évolués sont des combinaisons linéaires de transformations non linéaires d'ordre inférieur des variables prédictives. Les coefficients linéaires pour chacun des gènes évolués d'un individu sont estimés à l'aide de la méthode des moindres carrés ordinaires. Des descriptions complètes de l'algorithme MGGP peuvent être trouvées dans (Searson, 2015).

La Figure 3.4 montre un exemple de modèle MGGP dans lequel le résultat est représenté par une combinaison linéaire de deux gènes (Gene1 et Gene2) développés à l'aide de trois variables d'entrée (x_1, x_2, x_3). Chaque gène est un modèle non linéaire car il contient des termes non linéaires ($\cos(\cdot), \log(\cdot)$). Les coefficients linéaires (poids) de Gene1 et Gene2 (d_1 et d_2) et le biais (d_0) du modèle sont obtenus à partir des données d'apprentissage à l'aide d'une analyse de régression (méthode des moindres carrés ordinaires).



$$y = d_0 + d_1[\cos(x_1 * x_2) - (x_3/2)] + d_2[(x_2 * x_3) + \log(x_1 * 2)]$$

Figure 3.4. - Exemple typique du modèle MGGP

La Figure 3.5 présente un organigramme typique d'une procédure MGGP dans laquelle N_{gen} est le nombre de générations, Pr , Pc et Pm sont les probabilités de reproduction, de croisement et de mutation, respectivement.

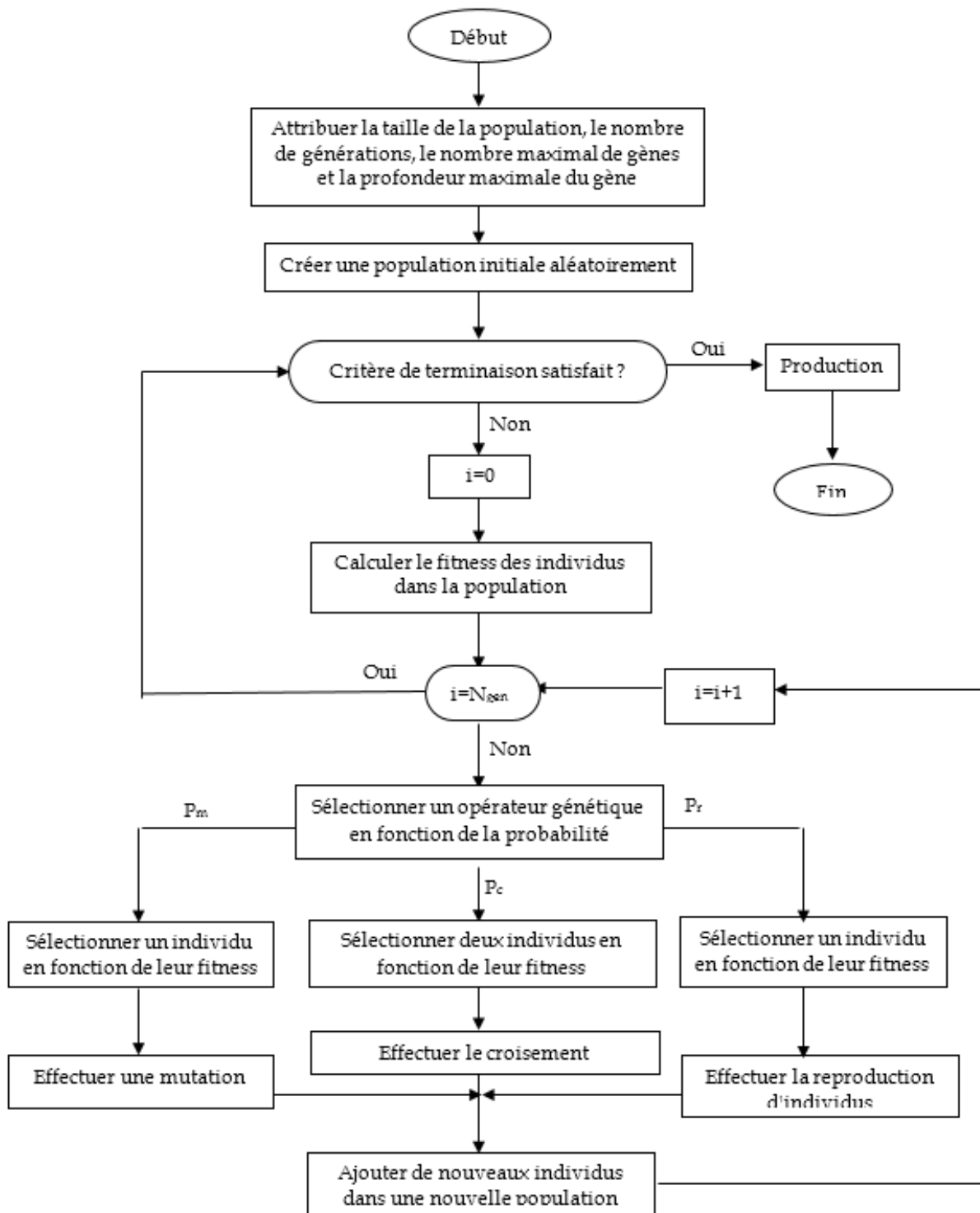


Figure 3.5.- Un organigramme typique pour une procédure de programmation génétique multi-gène

Dans la procédure MGGP, la population initiale est générée en créant des individus contenant des gènes évolutifs de manière aléatoire à partir des fonctions et des variables définies par l'utilisateur. Outre les mécanismes d'évolution de la GP standard décrits ci-dessus, il existe des mécanismes de croisement MGGP spéciaux (Searson, 2015), qui permettent

l'échange de gènes entre individus. Une brève description de ceux-ci est présentée ci-dessous.

3.4.1 Croisement de haut niveau

Une opération de croisement de haut niveau permet l'échange de gènes entre deux individus parents du pool d'accouplement et peut être expliquée à l'aide d'un exemple où le premier individu possède quatre gènes [G1, G2, G3, G4] et le second contient trois gènes [G5, G6, G7], avec G_{\max} égal à 5. Deux points de croisement sont sélectionnés aléatoirement pour chaque parent et les gènes délimités par des points de croisement sont désignés par {...}.

[G1, {G2, G3, G4}], [G5, G6, {G7}]

Les gènes entourés par les points de croisement sont échangés et ainsi, deux individus de progéniture sont créés comme indiqué ci-dessous.

[G1, {G7}], [G5, G6, {G2, G3, G4}]

Si l'échange de gènes a pour résultat qu'un individu contient plus de gènes que G_{\max} , les gènes sont sélectionnés de manière aléatoire et supprimés jusqu'à ce que l'individu contienne les gènes G_{\max} .

3.4.2 Croisement de bas niveau

Le croisement de bas niveau est appelé aussi le croisement de sous-arbre GP standard. Dans cette opération, un gène est d'abord sélectionné de manière aléatoire parmi chacun des individus parents dans le pool d'accouplement, puis un échange de sous-arbres sous des nœuds choisis arbitrairement de chaque gène est effectué. Les arbres résultants remplacent les arbres parents chez les individus parents par ailleurs inchangés, lesquels produisent ensuite des individus descendants de la génération suivante sans aucune suppression de gènes.

Les probabilités de chacun des processus de recombinaison (mécanismes évolutifs) peuvent être définies par les utilisateurs pour obtenir le meilleur modèle MGGP. Ces processus sont regroupés en catégories appelées événements. Par conséquent, la probabilité de croisement, de mutation et l'événement de reproduction directe doivent être spécifiés par l'utilisateur de manière à ce que la somme de ces probabilités soit égale à un. Les probabilités des sous-types d'événement peuvent également être spécifiées par l'utilisateur.

Par exemple, une fois que la probabilité d'événement de croisement est sélectionnée, il est possible de définir les probabilités d'un croisement de haut niveau et d'un croisement de bas niveau en gardant à l'esprit que la somme de ces probabilités de sous-type d'événement doit être égale à un.

3.5 Application de la programmation génétique en traitement d'images

Les techniques GP ont été appliqués dans divers domaines de traitement d'image tels que l'amélioration d'image, la restauration, l'enregistrement, la compression, la segmentation, la récupération, la filigrane, la détection d'objets, la détection de mouvement. Récemment, les auteurs dans l'article (Khan et al., 2021) ont fait une étude détaillée sur les applications de la programmation génétique dans le traitement d'images.

En raison de la nature différente des tâches de traitement d'image, aucune règle stricte et rapide ne peut être définie. Cependant, dans la plupart des applications de GP dans le traitement d'images, les auteurs dans (Khan et al., 2021) sont parvenus à un certain nombre d'observations, notamment :

- La grande population au sein de chaque génération contribue à accroître la diversité et augmente donc les chances d'obtenir un meilleur individu en moins de générations.
- La méthode d'initialisation de population couramment utilisée est la méthode de rampe moitié-moitié. Cette méthode produit l'arbre initial de longueur variable et contribue ainsi à augmenter la diversité de la population initiale.
- Une probabilité de croisement plus élevée est utilisée par rapport à la probabilité de mutation car des valeurs plus élevées de probabilité de mutation augmentent la zone de recherche dans l'espace de recherche et l'algorithme peut rester bloqué dans les minima locaux.
- La méthode Tournoi est utilisée comme méthode de sélection. L'avantage d'utiliser la méthode de sélection des tournois est qu'elle aide à maintenir une diminution de sélection constante.
- La sélection de la fonction de reproduction pour une application de traitement d'image particulière est la partie la plus importante.
- Le réglage des paramètres est également une étape importante dans l'application de GP dans l'une des tâches liées au traitement d'image. Avant de régler les paramètres GP, un chercheur doit étudier et analyser les réglages des paramètres GP dans les applications de traitement d'image connexes. Cela peut aider à gagner du temps, chaque fois que des paramètres de GP doivent être définis pour toute application liée au traitement d'image.

- Plusieurs essais sont effectués dans la plupart des travaux rapportés pour montrer l'efficacité des méthodes proposées.
- En raison du traitement lourd impliqué dans les tâches de traitement d'image, les algorithmes nécessitent un temps de formation important. Le temps de formation peut être considérablement réduit en exploitant les GPU (Graphics Processing Unit) pour des algorithmes améliorés.

Dans la littérature, GP a montré d'excellentes performances pour les problèmes liés à l'optimisation. L'utilisation de GP a énormément d'avantages à savoir : la compréhension, la flexibilité, la diversité de l'espace de recherche, et la flexibilité de la fonction de Fitness. Cependant, elle nécessite de grandes données de formation, et en raison de la nature stochastique de la GP, elle ne garantit pas une solution exacte.

3.6 Conclusion

Dans ce chapitre, nous avons présenté les concepts fondamentaux de la programmation génétique et plus particulièrement sa variante robuste MGGP. La MGGP combine la capacité du GP standard à construire la structure du modèle avec la capacité de régression traditionnelle dans l'estimation des paramètres.

Pour être plus précis, la structure et les paramètres d'un modèle de prédiction sont obtenus par MGGP ; ils ne nécessitent pas d'être supposés à l'avance. En conséquence, cet avantage permet de développer un modèle de prédiction sans limitation de forme, tandis que l'utilisateur peut décider du compromis entre la précision et la complexité du modèle de prédiction en contrôlant le maximum de gènes autorisés dans chaque individu et la profondeur des arbres dans MGGP, ce qui permet de faire un compromis entre la simplicité et la précision de la méthode.

Par conséquent, dans le chapitre suivant, nous présenterons notre contribution qui se manifeste dans le développement des modèles d'évaluation de qualité d'image en utilisant la variante robuste MGGP.

Chapitre 4

MGGP pour l'évaluation de qualité d'image

4.1 Introduction

Dans ce chapitre, nous exposerons notre contribution sur la proposition de nouvelle métrique pour l'évaluation de la qualité d'images naturelles tout en utilisant la programmation génétique multigène (MGGP).

La raison principale pour utiliser un tel algorithme est sa propriété d'exploiter un vaste espace de recherche composé de toutes les combinaisons possibles de plusieurs scores objectifs des métriques de qualité d'image (IQMs : Image Quality Metrics) et trouver le meilleur modèle de régression, compte tenu de sa simplicité ainsi que de ses performances sur les données d'apprentissage, qui décrit la somme pondérée des scores objectifs IQMs pour prédire les scores subjective des images dans des bases de données, sans faire d'hypothèses préalables sur la structure du modèle.

Dans ce qui suit, nous expliquons notre nouvelle approche et nous décrivons en détail chaque étape. Ensuite nous rapportons nos résultats, accompagnés d'une analyse et une discussion des constatations obtenues.

4.2 Principe de l'approche

Dans notre approche, nous avons choisi 21 métriques d'évaluation de qualité d'image sur la base de leur utilisation ou construction prévue et du fait que le code de chacun d'eux est accessible au public. Les métriques choisies : VSI (Zhang et al., 2014), FSIM (Zhang et al., 2011b), FSIMc (Zhang et al., 2011b), GSM (Liu et al., 2012), IFC (Sheikh et al., 2005), IW-SSIM (Wang and Li, 2011), MAD (Chandler, 2010), MSSIM (Wang et al.,

2003), NQM (Damera-Venkata et al., 2000), PSNR (Sheikh et al., 2006), RFSIM (Lin et al., 2010), SR-SIM (Zhang and Li, 2012), VIF (Sheikh and Bovik, 2006), IFS (Chang et al., 2015), SFF (Chang et al., 2013), SSIM (Wang et al., 2004), COHERENSI (Hegazy and Al Regib, 2014), UNIQUE (Temel et al., 2016), MSUNIQUE (Prabhushankar et al., 2017), PerSIM (Temel and AlRegib, 2015), RVSIM (Yang et al., 2018).

Pour faciliter la mise en œuvre de la programmation génétique multigène, nous avons choisi la plate-forme technologique largement utilisée pour la régression symbolique via MGGP «GPTIPS», qui est une boîte à outils MATLAB open source gratuite développée par Dominic Searson (Hii et al., 2011). Une version améliorée GPTIPS2 (Searson, 2015) est disponible depuis mai 2015. (GPTIPS est un acronyme de : Genetic Programming Toolbox for the Identification of Physical Systems).

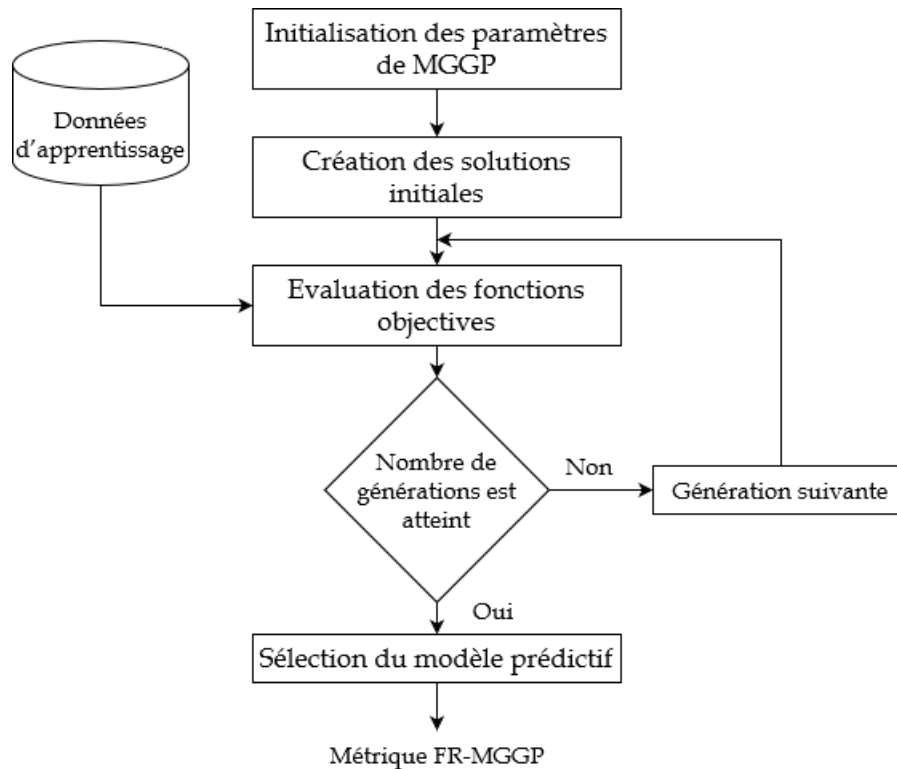


Figure 4.1. - Organigramme de l'approche proposée

Un organigramme de la méthode proposée, appelée «Full Reference Multi Gene Genetic Program» (FR-MGGP), est illustré à la Figure 4.1. La description de ses caractéristiques importantes (ensemble de données d'entraînement, solution, fonctions objectives et paramètres MGGP liés) est présentée dans ce qui suit :

4.2.1 Ensemble de données d’entraînement

Afin de construire les modèles MGGP et de tester leur efficacité et la généralisation de leur capacité, nous avons utilisé six benchmarks d’images populaires ; Laboratoire d’ingénierie de l’image et de la vidéo (LIVE (Wang et al., 2004)), Qualité d’image subjective catégorique (CSIQ (Chandler, 2010)), Tampere Image Database 2008 (TID2008 (Ponomarenko et al., 2009)), Image and Video Communication (IVC (Ninassi et al., 2006)), Tampere Image Database 2013 (TID2013 (Ponomarenko et al., 2015)) et Multiply Distorted Image Database (MDID (Sun et al., 2017)). Chacun contient des images de référence, des images déformées, et le score subjectif correspondant (soit le score d’opinion moyen (MOS) ou le score d’opinion moyen différentiel (DMOS)). Cependant, elles sont distinctes en terme, nombre d’images de référence et déformées, types de distorsion (les images contiennent simultanément plusieurs types de distorsions MTD, ou un seul type de distorsion (STD)), le nombre et les niveaux de distorsions dans l’image déformée. La description et les caractéristiques de ces bases ont été discutées dans le premier chapitre.

Généralement les méthodes d’apprentissage souffrent du problème de l’overfitting (le surajustement) ; il survient lorsque notre modèle essaye de trop coller aux données d’entraînements. Pour éviter ce problème, on va utiliser la méthode la plus efficace et la plus utilisée à savoir le cross validation (validation croisée en français).

Donc, pour assurer une robustesse des résultats, et selon la stratégie de validation croisée k-fold, plusieurs ensembles d’entraînement ont été construits ; chaque ensemble de données d’images a été divisé en 5 plis (plis=sous-sections) de taille approximativement égale (c’est-à-dire la méthode 5-fold est utilisée) ; 20% des images totales de chaque base de données sont sélectionnées comme ensemble de test, et les 80% restants ont été utilisées pour la formation. Cela a été faite 5 fois, où chaque pli a été utilisé comme ensemble de test une fois ; chaque image doit être testée exactement une fois et est utilisée dans l’entraînement 4 fois. La précision moyenne, des tests sur 5 plis, est prise comme mesure de performance.

L’ensemble d’apprentissage se compose d’un ensemble d’images utilisées uniquement pour l’apprentissage et il joue un rôle dans la construction des modèles MGGP. L’ensemble de test n’est pas utilisé pour faire évoluer les modèles et sert à donner une indication de la façon dont les modèles se généralisent aux nouvelles données.

Chaque image de l’ensemble d’apprentissage est décrite par des scores subjectifs y (MOS ou DMOS), et ses évaluations par 21 IQM utilisés présentées sous forme de scores objectifs $x_i : i = 1, \dots, 21$ (voir Tableau 4.1). La structure de l’échantillon, dans l’ensemble

d'apprentissage, est illustrée à la Figure 4.2.

Le score subjectif y est la variable de sortie/réponse, les scores objectifs, $x_i : i = 1, \dots, 21$, fournis par les IQM utilisés, sont les variables d'entrée utilisées pour prédire le score de qualité. Le modèle prédictif de y est une combinaison linéaire pondérée de transformations non linéaires d'ordre inférieur des variables d'entrée.

Tableau 4.1.- Les métriques FR-IQA utilisées

x_1 : VSI	x_{12} : SR-IM
x_2 : FSIM	x_{13} : VIF
x_3 : FSIMc	x_{14} : IFS
x_4 : GSM	x_{15} : SFF
x_5 : IFC	x_{16} : SSIM
x_6 : IW-SSIM	x_{17} : COHERENSI
x_7 : MAD	x_{18} : UNIQUE
x_8 : MSSIM	x_{19} : MSUNIQUE
x_9 : NQM	x_{20} : PerSIM
x_{10} : PSNR	x_{21} : RVSIM
x_{11} : RFSIM	

Image index	y (MOD/DMOS)	(1) : x_1	(2) : x_2	...	(21) : x_{21}
-------------	----------------	-------------	-------------	-----	-----------------

Figure 4.2.- Structure simple dans l'ensemble de formation

4.2.2 Solutions

Dans MGGP, chaque solution est généralement exprimée sous la forme d'une structure d'arbres, également appelée gènes. Il contient aléatoirement entre 1 et G_{\max} les gènes (arbres) avec la profondeur de l'arbre choisie aléatoirement entre 1 et D_{\max} ; G_{\max} et D_{\max} sont des paramètres définis par l'utilisateur. Chaque arbre est composé du terminal et de l'ensemble de fonctions. Les éléments de l'ensemble terminal T sont l'indice des IQM, qui sont choisis au hasard entre 1 et 21. L'ensemble de fonctions $F = \{ "+", "-" \}$ sont les nœuds internes de l'arbre.

Chaque solution codifie une équation mathématique, qui décrit un modèle de régression pour le score prédit (qualité de prédiction) \hat{y} de chaque image dans l'ensemble d'apprentissage. Les scores prédits \hat{y} peuvent être écrits comme la combinaison des scores objectifs pondérés x_i des IQM dont l'indice $i, i \in \{1, 2, \dots, 21\}$, sont présentés dans l'ensemble ter-

minal de la solution , plus un terme de biais ;

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_Gx_G; \quad (1)$$

$$G \in \{1, 2, \dots, 21\}$$

Où b_0 est le terme de biais (bruit) et b_G sont les coefficients de régression (c'est-à-dire les poids des gènes). Généralement, les coefficients sont déterminés par la méthode des moindres carrés ordinaires pour chaque individu MGGP (Searson, 2015).

4.2.3 Les Fonctions Objective

MGGP crée une séquence de populations et des générations, en appliquant des opérateurs génétiques (sélection, croisement et mutation) aux individus. Dans la première génération, une population d'individus aléatoires est générée. Chaque individu contient aléatoirement entre 1 et G_{max} de gènes. Un arbre représentant chaque gène est formé (avec la profondeur de l'arbre choisie aléatoirement entre 1 et D_{max}) en combinant aléatoirement les éléments de l'ensemble fonctionnel (+, -) et quelques éléments de l'ensemble terminal (l'indice des IQM).

A chaque génération, les individus sont évalués simultanément, à l'aide de deux fonctions objectives, l'une exprimant la complexité de la structure individuelle codant l'équation mathématique du modèle et l'autre est sa précision.

La restriction du nombre maximum d'arbres (gènes), de la structure de l'individue et de la profondeur du gène exerce un contrôle sur la complexité du modèle et aboutit à un modèle précis et compact. La minimisation de la somme des nœuds, dans la structure des arbres individuels, est utilisée comme fonction objectif pour évaluer la complexité du modèle.

La précision du modèle présentée par une fonction objective sera généralement la maximisation ou la minimisation d'une fonction d'agrégat qui combine les résultats de l'application de chaque échantillon de l'ensemble d'apprentissage à l'individu qui codifie l'équation mathématique de ce modèle. Maximiser le coefficient de corrélation R^2 (ou minimiser $1-R^2$) est utilisé comme fonction objective pour évaluer la précision du modèle.

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2)$$

où y_i et \hat{y}_i sont des scores subjectifs et prédits pour la $i^{\text{ème}}$ image dans l'ensemble d'apprentissage, respectivement. \bar{y} est la moyenne des scores prédits et n est le nombre d'images (échantillons) dans l'ensemble d'apprentissage.

Comme il existe, plus qu'une fonction objectif qui doivent être considérées simultanément et qui sont contradictoire, le problème est connu sous le nom de problème d'optimisation multi-objectif. Par conséquent, pour satisfaire ces fonctions objectives, un ensemble de solutions optimales est requis au lieu d'une solution optimale. La raison de l'optimalité de nombreuses solutions est qu'aucune fonction objective ne peut être considérée comme meilleure qu'une autre. Ces solutions sont des solutions «Pareto-optimales» ou des solutions non dominées. L'image de l'ensemble optimal de Pareto dans l'espace objectif est appelée le front de Pareto.

Dans l'algorithme MGGP, la détermination des solutions Pareto-optimales se fait selon le principe de l'algorithme NSGA-II (Deb et al., 2002), qui est incorporé dans GPTIPS (Searson, 2015). L'algorithme NSGA-II est utilisé à la fin de chaque génération de l'algorithme MGGP. Deux mesures sont utilisées pour comparer les individus (dans la sélection et la reproduction) : le rang de non-domination et le facteur d'encombrement. Le rang de non-domination mesure la proximité d'un individu par rapport au front non dominé, ce rang permet de classer les individus de la nouvelle et de l'ancienne population en fonction de leur position sur le front de Pareto (un individu de rang inférieur est toujours préféré à un individu de rang supérieur). Ensuite, le calcul d'un facteur d'encombrement pour chaque individu, il s'agit de la distance moyenne d'une solution aux solutions les plus proches (de chaque côté) sur le même front de Pareto.

Les solutions sont alors classées selon leur position et les solutions qui sont sur le même front de Pareto sont classées selon leur facteur d'encombrement. Les meilleures 50 % de la population survivent jusqu'à la génération suivante tandis que les autres sont rejetés.

En apprenant à partir d'un ensemble d'apprentissage, la méthode proposée donne lieu à un certain nombre de solutions optimales, appelées solutions Pareto-optimales; chacun décrit une équation précise du modèle de régression pour prédire le score de qualité. Cependant, ces modèles sont de complexité et de performances variables. La Figure 4.3 affiche des exemples de la population de modèles évolués en termes de complexité ainsi que de leurs performances, pour les cinq plis d'entraînement (validation croisée 5 fois) dans le cas du benchmark LIVE. Là où les cercles bleus montrent les résultats de tous les modèles évolués, les cercles verts comprennent les modèles Pareto-optimaux dans la population.

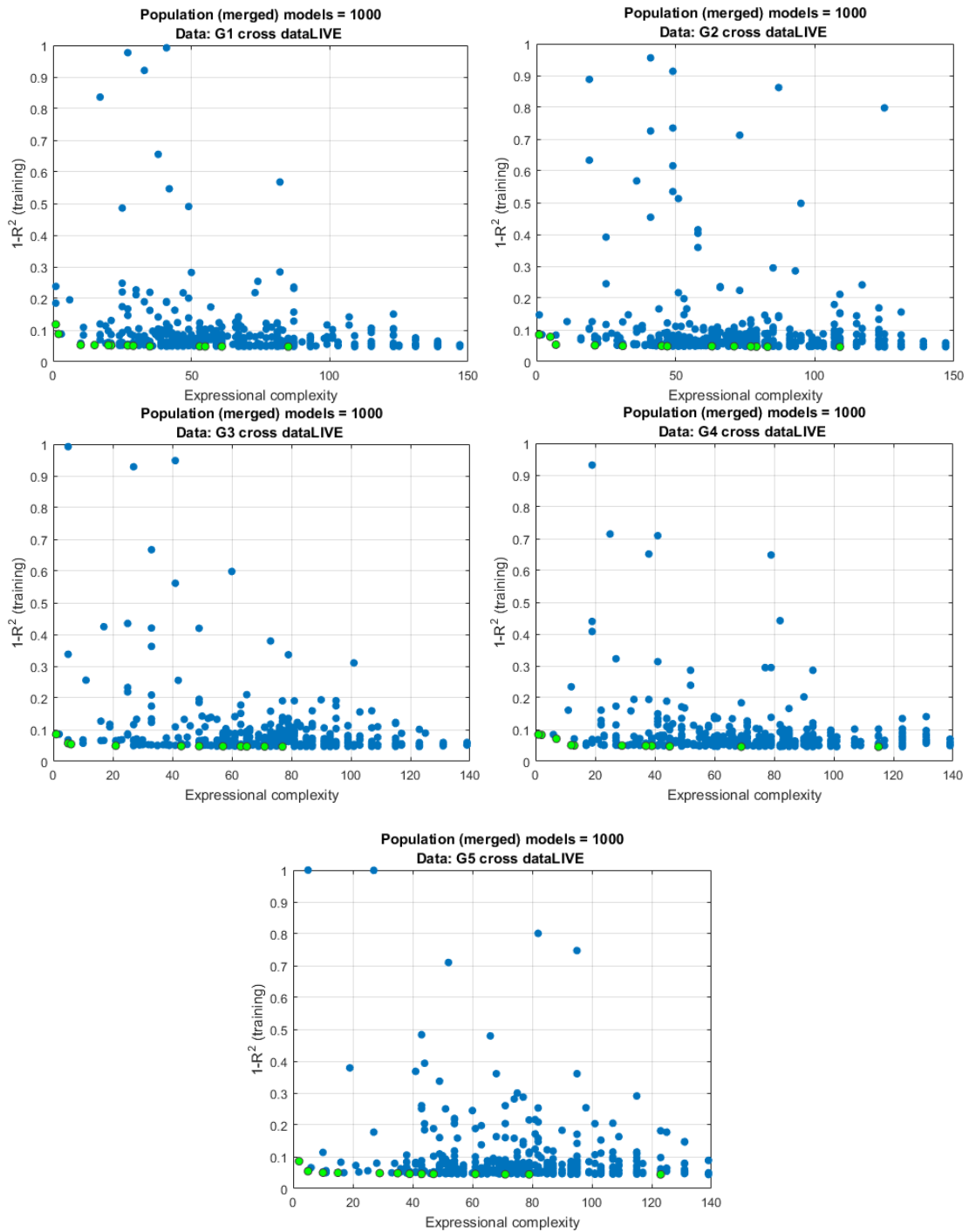


Figure 4.3. - Population de modèle évoluée, en termes de performances et de complexité du modèle, pour les cinq ensembles d'entraînement (cas : LIVE)

4.2.4 Paramètres MGGP

Un ensemble de paramètres doit être défini pour le processus évolutif de MGGP. Le nombre d'individus (solutions ou programmes) dans la population est déterminé par la taille de la population (TP). Le nombre de générations que l'algorithme utilisera avant la fin de l'exécution est défini par le nombre de générations (NG). La taille et les diverses

formes du modèle à rechercher dans l'espace global des solutions, sont définies par le nombre maximum de gènes (arbres) autorisés chez un individu (Gmax) et la profondeur maximale de l'arbre (Dmax). Le taux de croisement à haut niveau à utiliser pour générer de nouveaux gènes pour les individus ainsi que pour réduire le nombre total de gènes pour un modèle et augmenter le nombre total de gènes pour l'autre est déterminé par le paramètre de taux de croisement. Des croisements dispersés, des mutations gaussiennes et des règles de sélection de tournoi ont été utilisés. Les valeurs de ces paramètres sont déterminées sur la base des valeurs suggérées précédemment qui peuvent être trouvées dans la littérature (Hii et al., 2011; Searson, 2015; Searson et al., 2007) et par des expériences.

Le tableau 4.2 montre les paramètres utilisés pour la mise en œuvre de MGGP dans cette étude.

Tableau 4.2. - Paramètres de l'algorithme MGGP

Paramètre	Valeurs
Taille de la Population (TP)	100
Nombre de générations (NG)	100
Fonctions	+, -
Nombre Maximum de gènes (Gmax)	3
Nombre Maximum de depth (Dmax)	5
Taille de Tournoi	2
Fraction d'Elite	0.05
Evènement de Croisement	0,85
Evènement de Mutation	0,1

4.2.5 Sélection du modèle prédictif

Afin de sélectionner le modèle prédictif, à partir d'un modèle Pareto-optimal, un compromis doit être fait entre la complexité et les performances du modèle. Nous avons remarqué que les modèles de Pareto en bas à gauche de la population (haute précision et faible complexité) sont généralement ceux où une solution satisfaisante peut être trouvée.

Pour chaque ensemble de données de référence, nous obtenons cinq modèles prédictifs pour l'évaluation de la qualité d'image de référence complète (M_i , $i=1..5$), comme résultat de l'application de la stratégie de validation croisée à 5 plis. Par conséquent, la moyenne de ces cinq modèles est le modèle prédictif précis pour prédire les scores subjectives d'images dans les ensembles de données considérés, il est appelé une métrique de référence complète basée sur la programmation génétique multigène (FR-MGGP : Full Reference metric based

4.3 Évaluation expérimentale

Pour évaluer les performances de la méthode proposée, plusieurs expériences menées avec différents aspects, sont analysées dans les paragraphes suivants.

4.3.1 Évaluation des performances de prédiction

La relation entre les scores de qualité d'image prédits par les métriques IQA objectives et les scores subjectifs (généralement exprimés en MOS ou DMOS) peut être examinée par un graphe de nuages de points. La Figure 4.4 présente, pour chaque ensemble de données de référence, les nuages de points pour le modèle proposé (FR-MGGPs) et les trois meilleures métriques IQA de pointe. De plus, un ajustement avec une fonction logistique comme suggéré dans (Sheikh et al., 2006) et (MRohaly, 2000) est montré pour une comparaison plus facile. Chaque point représente une image dans le benchmark, l'axe horizontal correspond au score métrique (valeur objective) et l'axe vertical correspond aux score subjectif pour cette image. On remarque que tout les modèles FR-MGGPs présentent une meilleure linéarité et corrélation par rapport aux autres métriques de pointe ; où le pourcentage de valeurs aberrantes est diminué et une tendance à un comportement monotone est augmentée. Alors, on conclut que le FR-MGGP résultant est adéquat pour la perception humaine sur tous les ensembles de données de référence.

Pour évaluer les performances des six modèles FR-MGGPs développés, nous avons utilisé les indices de performance (concernant la précision, la monotonie et la cohérence de la prédiction) (MRohaly, 2000; Sheikh et al., 2006) suivant : l'erreur quadratique moyenne (RMSE), le coefficient de corrélation linéaire de Pearson (PCC), le coefficient de corrélation d'ordre de Spearman (SRCC) et le coefficient de corrélation d'ordre de Kendall (KRCC). Compte tenu de la relation non linéaire entre les distorsions d'image et leur qualité perçue, ces indices de performance sont calculés après une fonction de mappage non linéaire (Sheikh et al., 2006; Wang et al., 2004) entre un vecteur de score objectif S et des scores subjectifs, en utilisant la fonction de mappage suivante pour la régression non linéaire (comme recommandé par le Group des Experts de Quality Video (VQEG) (Barri et al., 2014)) :

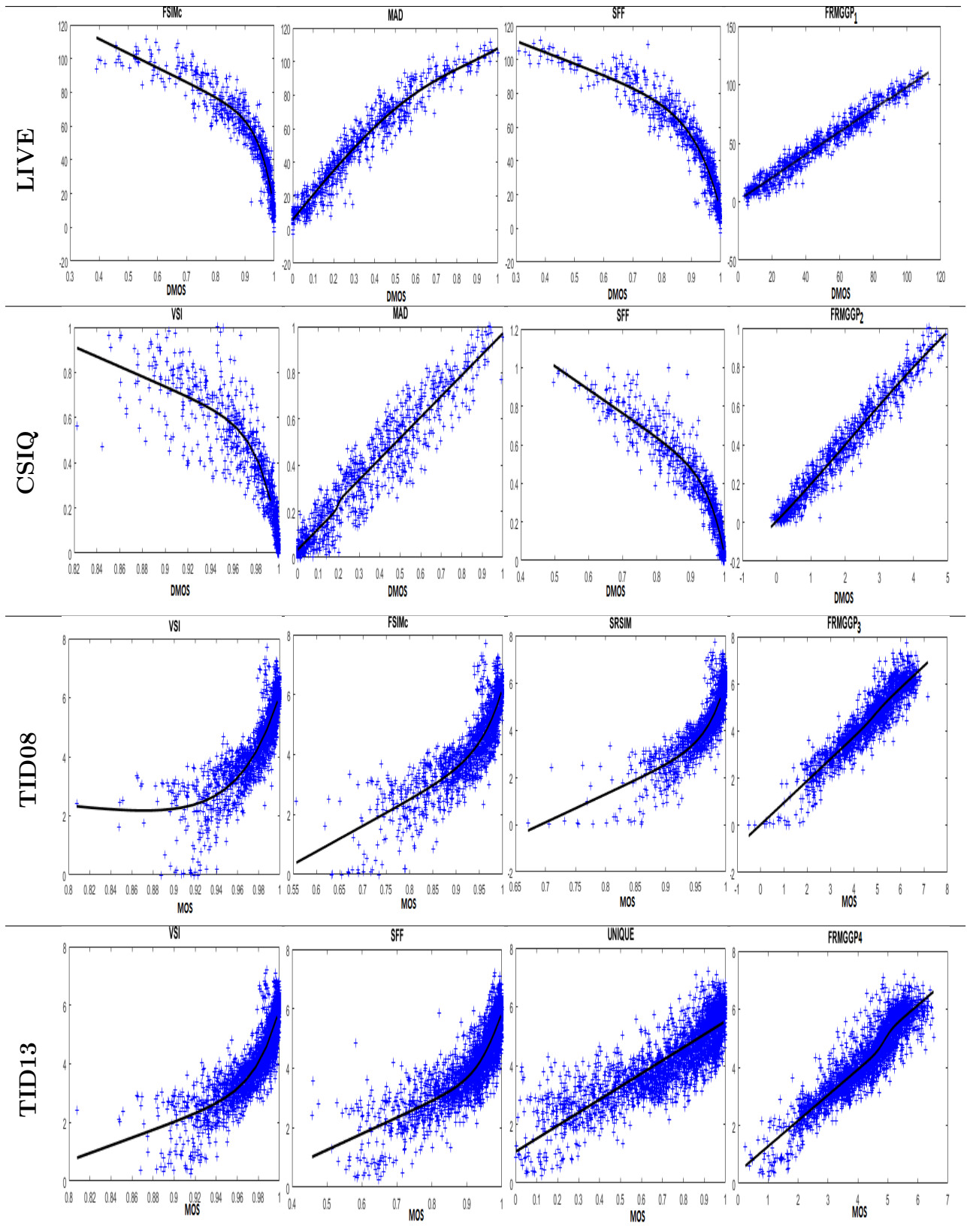
$$S_m = \beta_1 \frac{1}{2} - \frac{1}{1 + \exp(\beta_2(S - \beta_3))} + \beta_4 S + \beta_5 \quad (11)$$

Où $\{\beta_1, \beta_2, \beta_3, \beta_4, \beta_5\}$ sont des paramètres à ajuster, et S_m le mappage non linéaire de S . Des valeurs différentes peuvent conduire à des valeurs différentes de PCC et RMSE, mais elles n'affectent pas SRCC et KRCC. Cette non-linéarité est appliquée au score de l'algorithme, qui donne un meilleur ajustement pour toutes les données.

Le Tableau 4.3 présente les résultats de l'évaluation, pour les dix meilleurs modèles (parmi les 21 métriques utilisées) et les FR-MGGPs obtenus. Les deux premiers modèles pour chaque critère sont indiqués en gras. Selon (MRohaly, 2000; Sheikh et al., 2006), des valeurs plus élevées de SRCC, KRCC et PCC sont considérées comme meilleures, contrairement aux valeurs de RMSE. Nous observons par exemple, dans la base de données LIVE, les valeurs RMSE de toutes les métriques appartiennent à l'intervalle [5,913 – 12,022], où les valeurs les plus petites sont respectivement 5,913 et 6,816 pour FRMGGP1 et FRMGGP3 et la valeur la plus élevée est 12,022 pour MS-UNIQUE. En même temps ces méthodes (FRMGGP1 et FRMGGP3) ont les valeurs les plus élevées pour les trois coefficients (SRCC, KRCC et PCC) ; ce qui confirme qu'ils s'agissent des deux meilleures méthodes pour cette base de données.

En résumé, les résultats du Tableau 4.3 montrent que toutes les mesures FR-MGGP sont clairement les modèles les plus performants par rapport aux autres mesures dans les autres bases de données ; les métriques FR-MGGP1 et FR-MGGP3 sont nettement surperformées par rapport aux mesures sur LIVE. Pour le CSIQ, le FR-MGGP2 est le meilleur. FR-MGGP3 et FR-MGGP4 ont donné les meilleurs résultats au TID2008. Et pour TID2013, le FR-MGGP4 est meilleur que toutes les métriques comparées. FR-MGGP5 et FR-MGGP6 offrent respectivement les meilleures performances sur IVC et MDID.

Le tableau 4.3 contient également des moyennes directes et pondérées des valeurs obtenues. Pour les moyennes pondérées, le nombre d'images dans la base de données est utilisé comme son poids. Étant donné que les résultats globaux pour RMSE ne prennent pas en compte les ensembles de données LIVE et MDID en raison de la différence de portée. Les résultats montrent des performances exceptionnelles de la mesure FR-MGGP4.



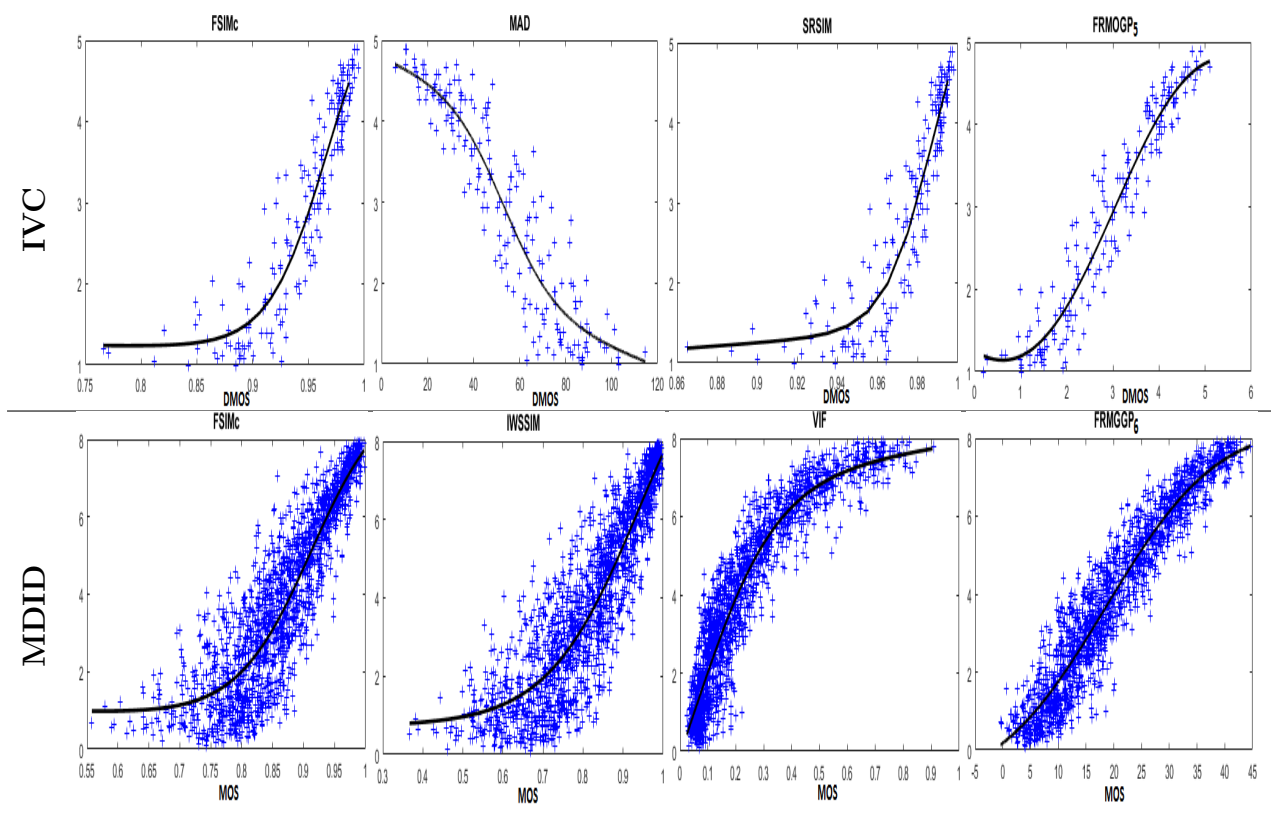


Figure 4.4. - Corrélations des métriques objectives obtenues et les trois meilleures mesures IQA de pointe avec les valeurs subjectives pour chaque ensemble de données (Chaque point de données représente une image de test)

Tableau 4.3.- Comparaison des performances de l'approche proposée avec les dix meilleurs modèles d'IQA

	VSI	SRSIM	IWSSIM	FSIMc	SFF	MAD	MSSIM	VIF	MS- UNIQUE	PerSIM	FR- MGGP1	FR- MGGP2	FR- MGGP3	FR- MGGP4	FR- MGGP5	FR- MGGP6
LIVE																
SRCC	0,9524	0,9618	0,9567	0,9645	0,9649	0,9669	0,9513	0,9636	0,9452	0,9435	0,9754	0,9259	0,9705	0,966	0,9397	0,965
KRCC	0,8058	0,8299	0,8175	0,8363	0,8365	0,8421	0,8045	0,8282	0,7958	0,7957	0,8609	0,7675	0,8469	0,8347	0,7904	0,8317
PCC	0,9482	0,9553	0,9522	0,9613	0,9632	0,9675	0,9489	0,9411	0,898	0,9209	0,9763	0,9166	0,9684	0,9651	0,9231	0,9624
RMSE	8,6816	8,0813	8,3473	7,5297	7,346	6,9073	8,6188	9,2402	12,0225	10,8607	5,9131	10,9248	6,8165	7,1565	10,5046	7,4251
CSIQ																
SRCC	0,9423	0,9319	0,9213	0,931	0,9627	0,9466	0,9133	0,9195	0,9292	0,9295	0,966	0,9753	0,9653	0,9527	0,9205	0,9487
KRCC	0,7857	0,7725	0,7529	0,769	0,8288	0,797	0,7393	0,7537	0,7594	0,7686	0,8398	0,8622	0,8373	0,8091	0,7551	0,8022
PCC	0,9279	0,925	0,9144	0,9192	0,9643	0,95	0,8991	0,9277	0,9286	0,8982	0,9723	0,9791	0,9651	0,9516	0,9312	0,9566
RMSE	0,0979	0,0997	0,1063	0,1034	0,0695	0,082	0,1149	0,098	0,0974	0,1154	0,0613	0,0534	0,0688	0,0807	0,0957	0,0765
TID08																
SRCC	0,8979	0,8913	0,8559	0,884	0,8767	0,834	0,8542	0,7491	0,8696	0,8568	0,8913	0,9023	0,9146	0,9146	0,7843	0,8385
KRCC	0,7123	0,7149	0,6636	0,6991	0,6882	0,6445	0,6568	0,586	0,6816	0,6799	0,7073	0,7276	0,7427	0,7483	0,6114	0,6539
PCC	0,8762	0,8866	0,8579	0,8762	0,8817	0,8306	0,8451	0,8084	0,8459	0,8376	0,8961	0,9041	0,9176	0,9128	0,8297	0,8625
RMSE	0,6466	0,6206	0,6895	0,6468	0,6333	0,7473	0,7173	0,7899	0,7157	0,733	0,5957	0,5734	0,5334	0,548	0,7491	0,679
TID13																
SRCC	0,8965	0,7999	0,7779	0,851	0,8513	0,7807	0,7859	0,6769	0,8708	0,8536	0,8292	0,8137	0,8462	0,9072	0,7407	0,7988
KRCC	0,7183	0,6314	0,5977	0,6665	0,6581	0,6035	0,6047	0,5147	0,6872	0,6776	0,6431	0,6409	0,6681	0,7421	0,5658	0,6117
PCC	0,9	0,859	0,8319	0,8769	0,8706	0,8267	0,8329	0,7735	0,8543	0,8549	0,8649	0,8658	0,8806	0,9205	0,8084	0,8326
RMSE	0,5404	0,6347	0,688	0,5959	0,6099	0,6975	0,6861	0,7856	0,6463	0,7594	0,6222	0,6202	0,5874	0,4844	0,7297	0,6866
IVC																
SRCC	0,8993	0,9268	0,9125	0,9293	0,9249	0,9146	0,898	0,8964	0,9128	0,8947	0,9141	0,9146	0,9132	0,9185	0,9544	0,9163
KRCC	0,7217	0,7562	0,7339	0,7636	0,7553	0,7406	0,7203	0,7158	0,7451	0,7132	0,7399	0,741	0,7361	0,7475	0,8127	0,742
PCC	0,912	0,9361	0,9231	0,9392	0,9324	0,921	0,9108	0,9028	0,9241	0,9002	0,9153	0,9214	0,9194	0,9188	0,9599	0,9159
RMSE	0,4999	0,4284	0,4686	0,4183	0,4404	0,4746	0,5029	0,5239	0,4657	0,5305	0,4907	0,4736	0,4793	0,481	0,3417	0,4892
MDID																
SRCC	0,8569	0,8521	0,8911	0,8904	0,8396	0,7249	0,8296	0,9306	0,8712	0,8193	0,0132	0,804	0,6633	0,8332	0,8703	0,9505
KRCC	0,6707	0,6685	0,7092	0,7122	0,6599	0,5331	0,6365	0,7714	0,689	0,6285	0,0184	0,6111	0,4782	0,6445	0,6919	0,8051
PCC	0,8703	0,8685	0,8983	0,8998	0,859	0,7552	0,8419	0,9367	0,8807	0,8311	0,1333	0,8231	0,6677	0,8443	0,8863	0,9525
RMSE	1,0853	1,0922	0,9682	0,9612	1,128	1,4442	1,1891	0,7717	1,0437	1,2254	2,1838	1,2512	1,6403	1,1807	1,0204	0,6707
Moyenne directe																
SRCC	0,9076	0,8940	0,8859	0,9084	0,9034	0,8613	0,8721	0,8560	0,8907	0,8829	0,7649	0,8893	0,8789	0,9154	0,8683	0,9030
KRCC	0,7358	0,7289	0,7125	0,7411	0,7378	0,6935	0,6937	0,6950	0,7264	0,7106	0,6349	0,7251	0,7182	0,7544	0,7046	0,7411
PCC	0,9058	0,9051	0,8963	0,9121	0,9119	0,8752	0,8798	0,8817	0,8886	0,8738	0,7930	0,9017	0,8865	0,9189	0,8898	0,9138
RMSE	0,4462	0,4459	0,4881	0,4411	0,4383	0,5004	0,5053	0,5494	0,4813	0,5346	0,4425	0,4302	0,4172	0,3985	0,4791	0,4828
Moyenne pondérée																
SRCC	0,8996	0,8625	0,8527	0,8873	0,8794	0,8205	0,8415	0,8014	0,8853	0,8657	0,7140	0,8614	0,8515	0,9053	0,8194	0,8722
KRCC	0,7238	0,6939	0,6750	0,7124	0,7033	0,6461	0,6588	0,6414	0,7063	0,6909	0,5796	0,6923	0,6842	0,7409	0,6486	0,7038
PCC	0,8973	0,8851	0,8733	0,8956	0,8915	0,8431	0,8577	0,8491	0,8718	0,8589	0,7523	0,8842	0,8655	0,9117	0,8564	0,8906
RMSE	0,5012	0,5401	0,5902	0,5280	0,5267	0,6085	0,5999	0,6707	0,5749	0,6434	0,5224	0,5130	0,4869	0,4399	0,6234	0,5824

4.3.2 Évaluation de la signification statistique

Afin d'évaluer la signification statistique des modèles IQA obtenus, un test F statistique a été utilisé (Sheikh et al., 2006) ; Il est basé sur les résidus de prédiction (différence numérique entre les scores subjectives et les scores prédits d'une métrique) de chaque mesure après mappage non linéaire, où la plus petite variance résiduelle dénotait la meilleure prédiction.

Le tableau 4.4 présente les résultats de ces tests ; où un symbole «1», «0» ou «-1» indique que la métrique de la ligne est statistiquement meilleur avec une confiance supérieure à 95%, impossible à distinguer ou pire que la métrique de la colonne, respectivement.

Tableau 4.4. - Evaluation de la signification statistique des base (a) LIVE, (b) CSIQ, (c) TID2008, (d) TID2013, (e) IVC and (f) MDID

(a) LIVE	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1
M2	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	0	-1	1	1	-1
M3	1	0	1	0	0	-1	1	1	1	1	-1	1	0	1	1	0
M4	1	0	0	-1	-1	-1	0	1	1	1	1	1	-1	0	1	-1
M5	-1	-1	-1	-1	-1	-1	-1	-1	0	0	-1	-1	-1	-1	0	-1
M6	1	1	1	0	1	-1	1	1	1	1	-1	1	0	1	1	0

(b) CSIQ	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	1	1	1	1	1	1	1	1	1	1	0	-1	1	1	1	1
M2	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1
M3	1	1	1	1	0	1	1	1	1	1	-1	-1	0	1	1	0
M4	1	1	1	1	-1	0	1	1	1	1	1	1	-1	0	1	0
M5	0	0	1	1	-1	-1	1	1	0	1	-1	-1	-1	-1	0	-1
M6	1	1	1	1	0	1	1	1	1	1	-1	-1	0	0	1	0

(c) TID 08	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	1	1	1	1	1	1	1	1	1	1	0	-1	-1	-1	1	1
M2	1	1	1	1	1	1	1	1	1	1	1	0	-1	-1	1	1
M3	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1
M4	1	1	1	1	1	1	1	1	1	1	1	1	-1	0	1	1
M5	-1	-1	-1	-1	-1	0	-1	-1	-1	0	-1	-1	-1	-1	0	-1
M6	-1	-1	0	-1	-1	-1	1	1	0	1	-1	-1	-1	-1	1	0

(d) TID 13	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	-1	0	1	-1	-1	1	1	1	1	1	0	-1	-1	-1	1	1
M2	-1	1	1	-1	-1	1	1	1	1	1	1	0	-1	-1	1	1
M3	-1	1	1	1	1	1	1	1	1	1	1	1	0	-1	1	1
M4	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0
M5	-1	-1	-1	-1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	0	-1
M6	-1	-1	0	-1	-1	0	0	1	1	1	-1	-1	-1	-1	1	0

(e) IVC	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	0	0	0	-1	0	0	0	0	0	0	0	0	0	-1	-1	0
M2	0	0	0	-1	0	1	0	0	0	0	0	0	0	-1	-1	0
M3	0	-1	0	-1	0	0	0	0	0	0	0	0	0	0	-1	0
M4	0	0	0	0	0	0	0	1	0	1	1	1	0	0	-1	0
M5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
M6	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	-1	0

(f) MDID	VSI	SRSIM	IFS	FSIMc	SFF	MAD	MSSIM	VIF	UNIQUÉ	Persim	M1	M2	M3	M4	M5	M6
M1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	-1	-1	-1	-1	-1
M2	-1	-1	-1	-1	-1	1	-1	-1	-1	0	-1	0	-1	-1	-1	-1
M3	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	-1	0	-1	-1	-1
M4	-1	-1	-1	-1	-1	1	-1	-1	1	1	1	1	0	-1	-1	-1
M5	1	1	-1	-1	1	1	1	-1	0	1	1	1	1	1	0	-1
M6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0

Où :

M1=FR-MGGP1, M2=FR-MGGP2, M3=FR-MGGP3, M4=FR-MGGP4, M5=FR-MGGP5, M6=FR-MGGP6

En résumé, les tests de signification confirment d'excellentes performances de la famille développée de FR-MGGP, en particulier dans LIVE, CSIQ, TID2008 et TID2013. Et

dans IVC (MDID) la métrique FR-MGGP5 (FR-MGGP6) est nettement meilleurs que les autres.

4.3.3 Comparaisons des performances des mesures obtenues avec les mesures de fusion IQA

Il serait agréable de comparer les mesures FR-MGGPs avec d'autres mesures IQA de fusion connexes des algorithmes IQA de pointe en termes de la mesure de performance SRCC habituellement utilisée (étant donné que les valeurs de PCC, KRCC et RMSE ne sont souvent pas disponibles). Nous avons choisi quatre bases de données : TID2013, TID2008, CSIQ et LIVE, considérant les plus utilisées pour l'évaluation des mesures de fusion IQA. Le tableau 4.5 contient cette évaluation comparative.

Les deux meilleurs résultats pour un ensemble de données de référence donné sont écrits en gras. Certains résultats n'ont pas été rapportés dans les travaux référés; par conséquent, ils sont désignés par «-». Les résultats des méthodes sans référence (sont écrits en italique) ont été exclus de la comparaison, par exemple NR-SVR et NR-ANN, car notre méthode est avec une référence complète.

Les résultats de l'évaluation montrent que pour la base de données LIVE, FR-MGGP1 a surpassé les autres approches. FR-MGGP2 est le meilleur sur la base de données CSIQ, et FR-MGGP3 et FR-MGGP4 se sont bien comportés sur la base de données TID2008. Sur la base de données TID2013, FR-MGGP4 a fourni des résultats supérieurs aux autres approches et est comparable à CNNM (mais cette méthode est entraînée et testée sur une seule base de données).

On peut voir que les mesures formées sur des bases les plus petites ont tendance à mal fonctionner sur les grandes bases, par exemple, FR-MGGP1, FR-MGGP2 sur TID2013 et FR-MGGP2 sur TID2008. Le principal facteur qui affecte la performance des techniques évaluées est le nombre de distorsions présentées dans les bases de données de référence. Les approches entraînées sur de grandes bases, qui contiennent une variété de distorsions, fonctionnent généralement mieux que les techniques entraînées sur des bases plus petits, par exemple, FR-MGGP4 sur le LIVE, CSIQ et également dans TID2008.

De plus, les moyennes (directe et pondérée) ont été calculées hors TID2013 car certaines mesures n'avaient pas été évaluées dessus, et tiennent compte des mesures IQA pour lesquelles des résultats sont disponibles pour les autres bases de données. Ils sont donc comparés aux cinq mesures d'IQA de fusion FR dont les résultats sont connus. Les résultats des moyennes ont confirmé les performances exceptionnelles des FR-MGGP3 et

FR-MGGP4. La mesure de fusion introduite dans (Oszust, 2016b) a également obtenu de bons résultats, mais était moins bonne que FR-MGGP3 pour la Moyenne pondérée.

Tableau 4.5.- Comparaison des mesures obtenues avec d'autres mesures de fusion IQA basées sur les valeurs SRCC

IQA measure	TID13	TID08	CSIQ	LIVE	Moy. Dir.	Moy. Pond.
CNNM (Lukin et al., 2015)	0,93	-	-	-	-	-
<i>NR-SVR</i> (Borhen Eddine et al., 2016)	<i>0,859</i>	-	<i>0,920</i>	<i>0,967</i>	-	-
<i>NR-ANN</i> (Ieremeiev et al., 2019)	<i>0.772</i>	-	-	<i>0.982</i>	-	-
VCGS (Shi and Lin, 2020)	0.892	0.897	0.944	0.976	0.939	0.928
IGN (Wu et al., 2013)	-	0,890	0,940	0,958	0,929	0,919
LAF (Barri et al., 2014)	-	0,810	0,963	0,957	0,910	0,885
ESIM (Oszust, 2016a)	0,880	0,902	0,962	0,942	0,935	0,927
LCSIM (Oszust, 2016b)	0,830	0,910	0,973	0,972	0,952	0,939
RLCM (Ieremeiev et al., 2018)	0,887	-	0,945	0,963	-	-
FR-MGGP1	0,829	0,891	0,966	0,975	0,944	0,930
FR-MGGP2	0,813	0,902	0,975	0,925	0,934	0,927
FR-MGGP3	0,834	0,926	0,959	0,960	0,948	0,943
FR-MGGP4	0,907	0,914	0,952	0,959	0,942	0,935

En ce qui concerne la base de données MDID récemment publiée, une évaluation complète de 32 métriques FR-IQA de pointe a été présentée dans (Varga, 2019), où les auteurs ont démontré que la seule métrique HaarPSI (Reisenhofer et al., 2018) était capable de produire des valeurs SRCC supérieures à 0,9. Néanmoins, nous avons réussi avec notre nouvelle approche à obtenir une valeur plus élevée que cette méthode avec une valeur SRCC=0.9505 (où : SRCC(HaarPSI)=0.9028).

4.3.4 Test croisé de bases de données

De plus, pour évaluer la généralisation de la méthode FR-MGGP proposée, nous avons formé le système basé sur une base de données et nous l'avons testé sur les cinq autres bases de données. Les résultats sont présentés dans le Tableau 4.6.

Tableau 4.6. - Performances SRCC de FR-MGGP entre bases de données

Train \ Test	Test					
	LIVE	CSIQ	TID2008	TID2013	IVC	MDID
LIVE	-	0.96	0.90	0.80	0.91	0.81
CSIQ	0.92	-	0.90	0.78	0.91	0.79
TID2008	0.96	0.95	-	0.82	0.91	0.77
TID2013	0.96	0.96	0.90	-	0.91	0.86
IVC	0.86	0.90	0.73	0.66	-	0.67
MDID	0.96	0.94	0.83	0.76	0.92	-

On peut voir que les valeurs SRCC sont supérieures à 0,901 pour la plupart des cas, sauf lorsque le système est formé sur IVC ou testé sur TID2013 et MDID. Cette exception peut s'expliquer par le fait que l'ensemble testé implique quelques types de distorsion, qui ne sont pas abordés dans l'ensemble d'apprentissage, par exemple, TID13 et MDID sont beaucoup plus grands que les autres ensembles de données en termes de nombre d'images, de nombre de types de distorsion, et les images déformées avec multiple distorsions. D'autre part, IVC est la plus petite base, qui ne peut pas couvrir complètement tous les types de distorsion.

Les résultats montrent que la métrique proposée fonctionne bien en termes de corrélation SRCC ce qui confirme la généralisation de notre métrique.

4.3.5 Influence des métriques IQA sur la fusion

Les résultats présentés précédemment confirment la bonne performance des mesures de fusion IQA obtenues en comparaison avec les métriques de fusion de pointe et les métriques IQA simples. Cependant, il serait souhaitable de préciser pourquoi certaines métriques sont plus impliquées que d'autres dans la fusion ; alors on va estimer l'importance des intrants.

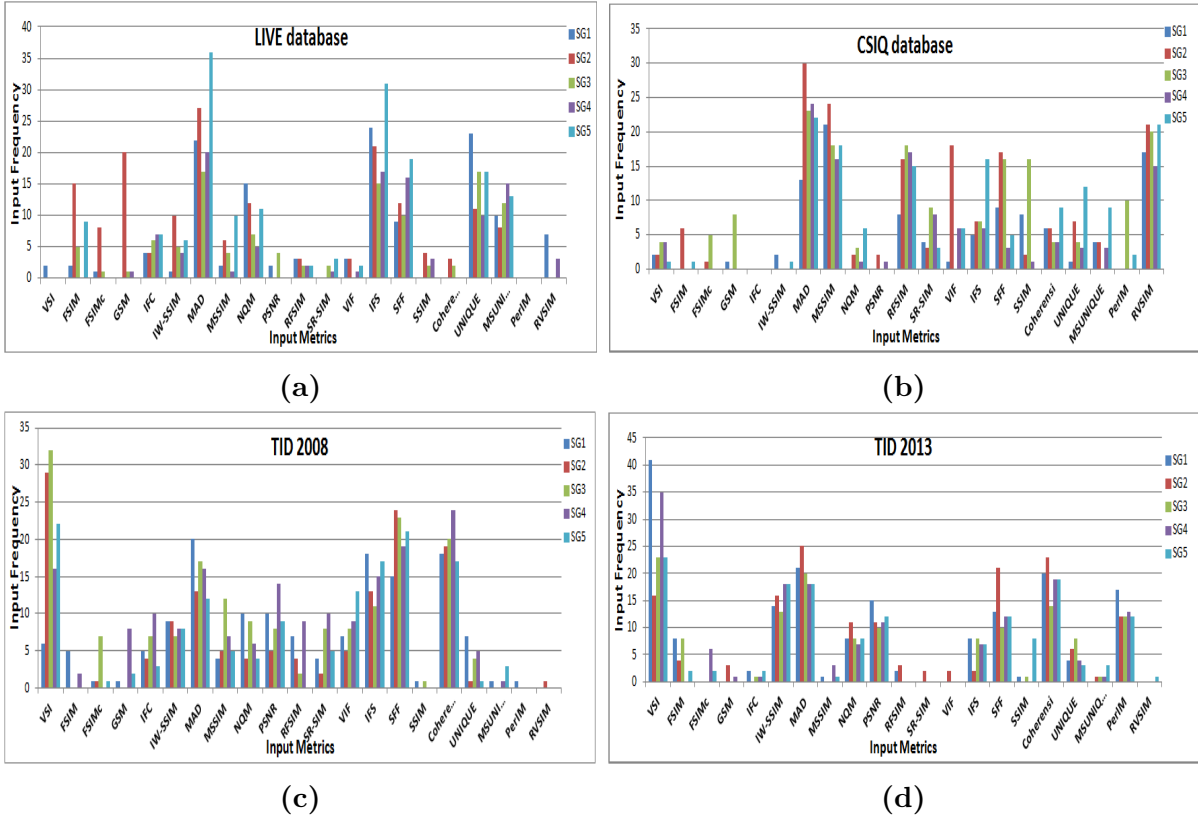


Figure 4.5. - Fréquence d'entrée dans les modèles Pareto pour chaque sous-groupe

La Figure 4.5 affiche un graphique à barres de fréquence d'entrée pour les modèles de Pareto dans chaque sous-groupe pour chaque ensemble de données de référence. Par exemple dans la base de données LIVE (Figure 4.5a), on peut voir l'importance des métriques : IFC, MAD, NQM, IFS, SFF, UNIQUE, MSUNIQUE ; ce qui est en accord avec notre prédiction dans la section précédente (FR-MGGP1). Également dans la base de données CSIQ (Figure 4.5b), les mesures MAD, MSSIM, RFSIM, SRSIM, IFS, SFF et MSUNIQUE ont une grande importance. Des observations similaires ont été faites pour les autres bases de données.

4.4 Conclusion

Au cours de ce chapitre, nous avons présenté notre approche afin de générer automatiquement des modèles prédictifs pour l'évaluation de la qualité d'image de référence complète, en utilisant la méthode de programmation génétique multigène. Le score d'image prédit est obtenu par la somme pondérée des scores objectifs d'un petit nombre de métriques de qualité d'image (IQMs). En apprenant à partir des bases de données d'images de référence que la MGGP peut déterminer les métriques de qualité d'image appropriées, à partir de 21 métriques utilisées, dont les scores objectifs sont utilisés comme prédicateurs, dans la régression symbolique, en optimisant simultanément deux objectifs concurrents

de «bonne adéquation» du modèle aux données et de la «complexité».

Pour évaluer les performances de la méthode proposée, plusieurs expérimentations ont été menées, avec des aspects différents, en utilisant six différentes benchmarks d'images (LIVE, CSIQ, TID2008, TID2013, IVC et MDID) et quatre indices de performance (SRCC, PCC, KROCC, RMSE). Pour chaque base de référence, nous avons obtenu un modèle prédit, à savoir une métrique de référence complète basée sur la programmation génétique multigène (FR-MGGP). Les résultats présentés confirment les performances supérieures des FR-MGGP obtenus par rapport aux mesures IQA de pointe, y compris d'autres approches de fusion récemment publiées.

Chapitre 5

Evaluation de la qualité des images du contenu de l'écran

5.1 Introduction

L'étude de l'évaluation de la qualité d'image du contenu de l'écran (Screen Content Image SCI) est un nouveau et intéressant sujet en raison de sa bonne capacité d'enseignement et d'optimisation dans divers systèmes de traitement. D'après nos modestes connaissances, il n'y a pas de méthode de combinaison de mesures dans le cas d'images de contenu de l'écran. Et comme nous avons obtenu d'excellents résultats par notre approche sur les images naturels; nous allons l'étendre pour le cas des images de contenu d'écran.

Dans ce chapitre, nous présenterons les caractéristiques des images de contenu d'écran, les bases d'images spécifiées pour ce type d'image et les différentes métriques développées pour évaluer les SCIs. Ensuite nous exposerons notre nouvelle méthode, et nous rapporterons nos résultats, accompagnés d'une analyse et une discussion des performances de la méthode proposée par rapport à quelques travaux récents de la littérature.

5.2 Image de contenu d'écran SCI

Avec le développement rapide du multimédia et des réseaux sociaux, de nombreuses applications typiques, telles que les actualités et la publicité en ligne, l'éducation en ligne, les brochures électroniques, Facebook, Twitter, etc., impliquent des images de contenu d'écran généralisées (SCI) (Wang et al., 2018, 2016; Yang et al., 2015). Les SCI typiques contiennent un mélange de sources qui incluent des images naturelles, du texte, des graphiques, des cartes et des logos, etc. Par conséquent, dans notre vie quotidienne, les images de contenu d'écran sont devenues très populaires. De plus, récemment une évaluation de la qualité de SCI est également apparue comme un sujet brûlant en raison de l'augmen-

tation exponentielle de la grande demande de ces services graphiquement riches.

les résultats de (Gu et al., 2016b) montrent que les SCI présentent plusieurs caractéristiques distinctes des images naturelles (NI). Ces NI ont des variations de couleurs riches, des bords lisses et un contenu de texture complexe avec des lignes épaisses, tandis que les images SCI ont généralement des lignes plus fines, des bords nets, avec des variations de couleurs limitées et des formes simples. Ces différences rendent les théories de conception développées pour évaluer la qualité des images de scènes naturelles moins fiables et non trivial pour le cas des SCIs.

5.3 Bases d'image de contenu d'écran

Dans la littérature, il existe peu de base d'images spécialisées pour les images de contenu d'écran. Et les plus utilisées pour la validation des méthodes de qualité d'images sont :

- La base de données d'évaluation de la qualité des images d'écran (SIQAD) (Yang et al., 2015). Cette base de données, la première créée, est constituée de 20 images de contenu d'écran de référence, et de 980 images correspondantes déformées par sept types de distorsions, chacune à sept niveaux de distorsions. Les types de distorsion sont spécifiquement le bruit gaussien (GB), le flou gaussien (GN), flou de mouvement (MB), changement de contraste (CC), compression JPEG (JPEG), compression JPEG2000 (J2K) et codage basé sur la segmentation des couches (LSC). Dans cette base de données, les SCI ont des résolutions différentes et les évaluations subjectives (en termes de valeurs DMOS) ont été collectées à l'aide d'une méthodologie à stimulus unique.
- La base de données de qualité d'image du contenu de l'écran (SCID) (Ni et al., 2017) est composée de 40 SCI de référence et de 1800 SCI déformés qui ont été subi par neuf types de distorsions à cinq niveaux. Les types de distorsion dans cette base de données incluent le bruit gaussien (GB), le flou gaussien (GN), flou de mouvement (MB), changement de contraste (CC), compression JPEG (JPEG), compression JPEG2000 (J2K), changement de saturation des couleurs (CSC), quantification des couleurs avec tramage (CQD) et codage de contenu d'écran à haute efficacité (HEVC-SCC). La résolution est fixée à 1280×720 . Les évaluations subjectives (en termes de valeurs MOS) ont été recueillies à l'aide d'une méthode d'échelle de dégradation à double stimulus.

Notez que dans les deux bases de données, les SCI d'origine ont été collectées avec diverses combinaisons de texte, de graphiques et d'images à travers des captures d'écran de pages Web, de diapositives PowerPoint, de fichiers PDF, de magazines numériques, etc.

Les images de référence des deux bases SCID et SIQAD sont présentées dans les Figures 5.1 et 5.2.

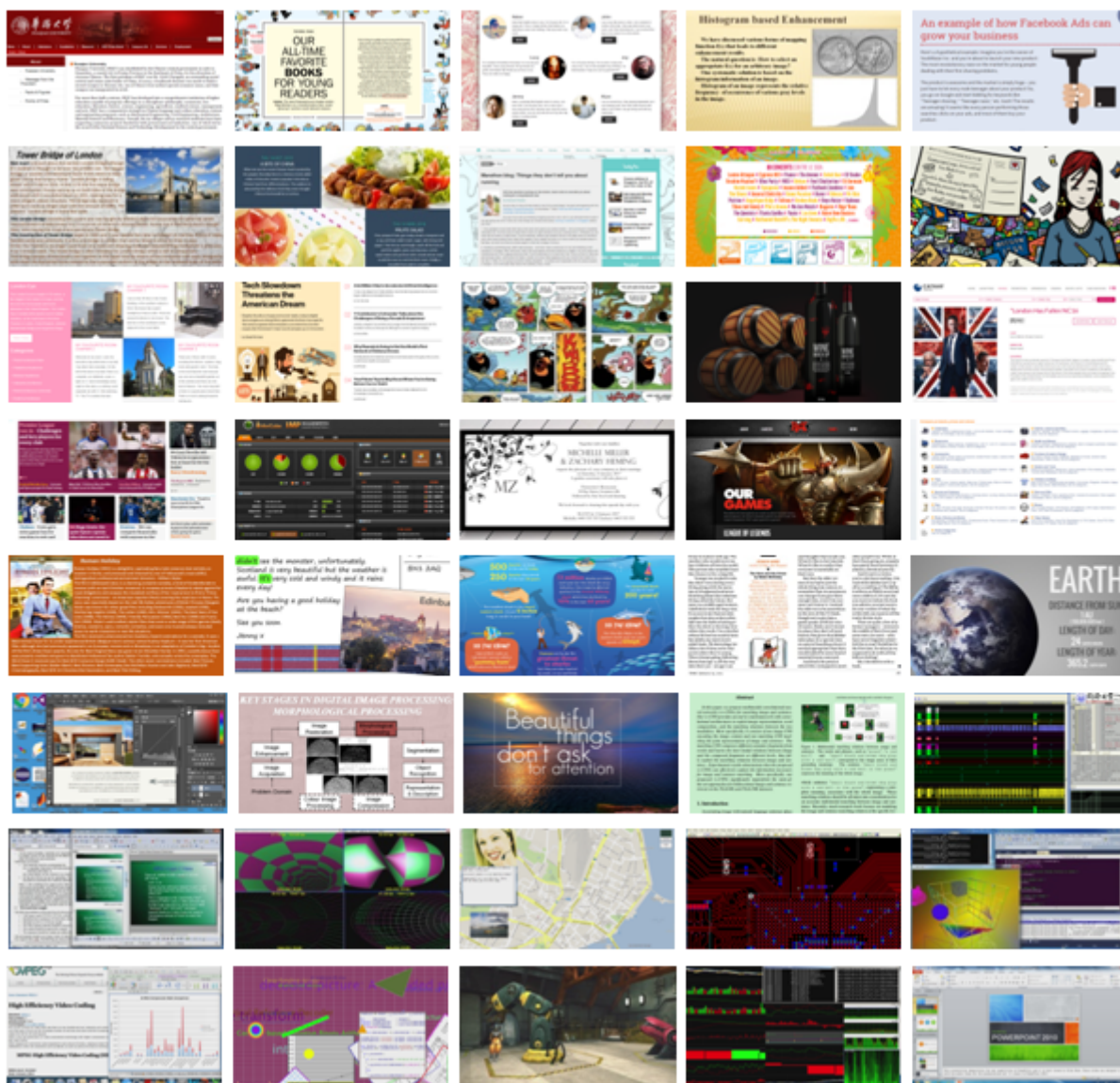


Figure 5.1.- Images de référence de la base SCID

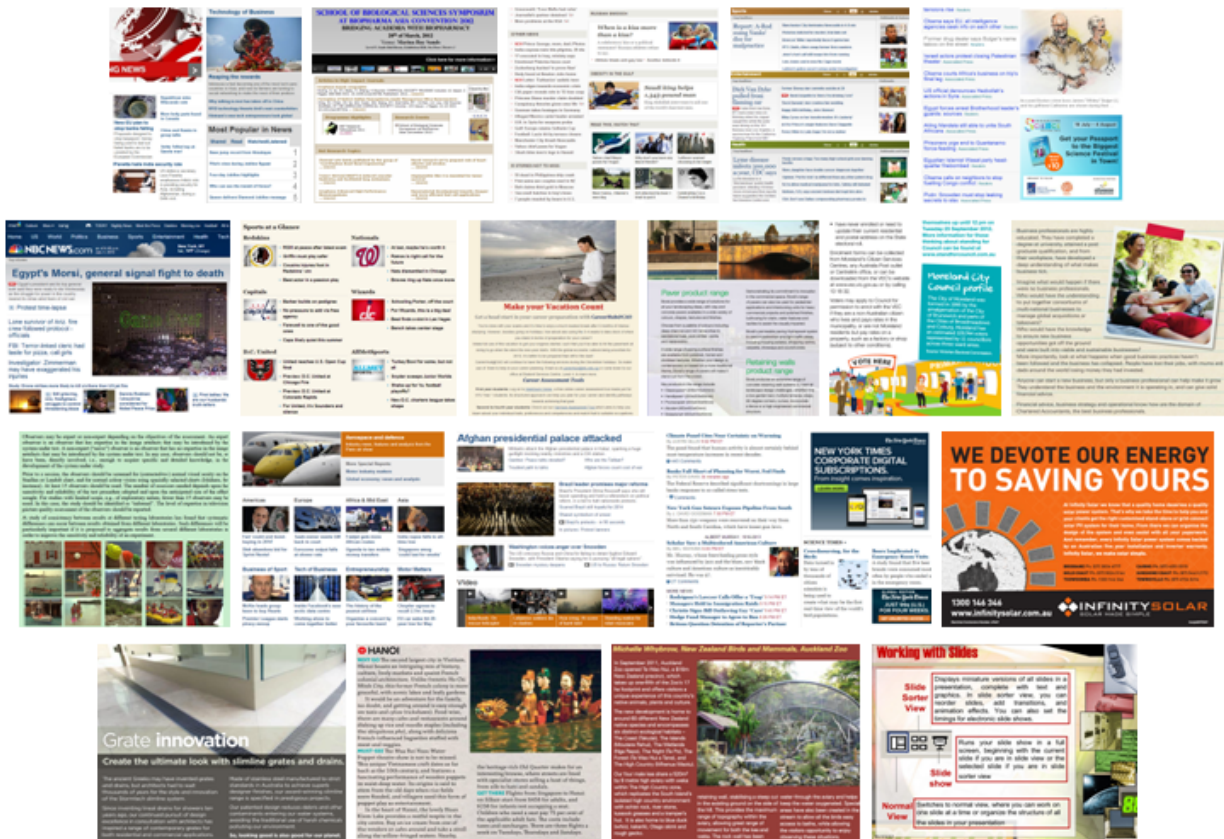


Figure 5.2.- Images de référence de la base SIQAD

5.4 Etat de l'art sur les métriques d'IQA des SCIs

Plusieurs algorithmes d'évaluation de la qualité d'image du contenu de l'écran (SCI-QA) ont été proposés. Parmi eux GFM (Ni et al., 2018), GSS (Ni et al., 2016), SCDM_new, SCQI_new (Bae and Kim, 2016), SQMS (Gu et al., 2016a), SVQI (Gu et al., 2018), ESIM (Ni et al., 2017). Bien que ces méthodes conçues pour les SCI aient obtenu des résultats encourageants. Une description générale de ces méthodes est dessous :

- Gu et al. (Gu et al., 2016a) ont proposé une mesure de la qualité des SCI, nommée SQMS (Saliency-guided Quality Measure of SCIs), guidée par l'identification des régions saillantes. Cette méthode repose principalement sur un simple opérateur de convolution, elle détecte les régions importantes de distorsion qui attirent généralement le plus attention, et ceci après avoir montré la dégradation de la structure causée par différents types de distorsion.
- S.Bae et M.Kim (Bae and Kim, 2016) ont proposé une nouvelle méthode, appelée SCQI (Structural Contrast-Quality Index), ils utilisent un indice de contraste structurel, qui pourrait bien caractériser les perceptions de la qualité visuelle locale et globale pour différentes caractéristiques d'image avec différent types de distorsion. En outre, ils ont développé une métrique de distorsion de contraste structurel

(SCDM : Structural Contrast Distortion Metric), une variante de SCQI avec les propriétés mathématiques souhaitables de capacité métrique de distance valide et de quasi-convexité.

- Ni et al. (Ni et al., 2016) ont proposé une approche appelée GSS (Gradient Similarity for SCI) qui utilise des similitudes dans la direction et l’amplitude du gradient entre les SCI de référence et déformés. Ensuite, une technique de mise en commun basée sur la variance est utilisée en combinant les deux cartes de similarité pour obtenir l’estimation globale de la qualité.
- Ni et al. (Ni et al., 2017) ont utilisé trois attributs de bord saillants (informations sur le contraste des bords, la largeur des bords et la direction des bords) extraits des SCI puis un processus de pondération approprié a été envisagé dans leur modèle IQA développé, appelé mesure de similarité des bords (ESIM : Edge SIMilarity). Il convient de mentionner qu’ils ont également développé et présenté une nouvelle base de données SCI, appelée SCID.
- Ni et al. (Ni et al., 2018) ont proposé un modèle de caractéristiques de Gabor (GFM : Gabor Feature-based Model). Ils ont utilisé la partie imaginaire horizontale et verticale du Gabor pour le développer. Cette approche est similaire à l’interprétation HVS que d’autres approches, et la complexité de la mesure est donc plus petite.
- Gu et al. (Gu et al., 2018) ont proposé aussi un indice de qualité basé sur la variation structurelle (SVQI : Structural Variation based Quality Index). Et selon la perception, ils divisent la structure en structures globales pour la perception de base et en structures locales pour la perception détaillée des êtres humains. Enfin, pour prédire le score final de qualité de SCI, une mesure des variations des structures globales et locales est systématiquement combinée.

Étant donné que le nombre de métriques d’IQA établies a augmenté ces dernières années et que chacune a sa propre utilisation ou construction prévue. Ce serait idéal s’il existe une métrique universelle qui exploite les avantages de certaines métriques et diminue l’influence de leurs inconvénients; ce qui peut être réalisé en adoptant la méthode de combinaison de métriques.

5.5 La programmation génétique et les images de contenu d’écran (GP-SCIQ)

La même idée, de l’approche FRMGGP discuté dans le chapitre précédant, est utilisée pour mesurer la qualité des images de contenu d’écran.

Pour les données d’entraînement, nous utilisons les sept mesures IQA de référence complète GFM (Ni et al., 2018), GSS (Ni et al., 2016), SCDM, SCQI (Bae and Kim,

2016), SQMS (Gu et al., 2016a), SVQI (Gu et al., 2018), ESIM (Ni et al., 2017)) où leurs codes sources sont librement disponibles, et les valeurs de sortie correspondantes sont les valeurs subjectives.

Les performances du modèle proposé sont évaluées sur deux ensembles de données d’images de référence qui sont souvent utilisés comme bases de données de validation SCI ; ils sont succinctement décrits au-dessus.

Pour toutes les expériences, nous appliquons la stratégie de validation croisée k-fold ($K=5$) pour sélectionner nos ensembles d’entraînement et de test. Sur la base d’images SIQAD, 980 SCI déformés sont répartis aléatoirement en deux sous-ensembles en fonction du contenu de l’image. L’ensemble d’apprentissage contient 784 SCI déformées associées à 16 images de référence (80% des données pour l’entraînement) et les autres 196 SCI déformées associées à 4 images de référence sont utilisées comme ensemble de test (20% des données pour les tests). Des images de références sont incluses soit dans l’ensemble de données d’entraînement ou bien dans l’ensemble de données de test ; cette procédure peut éviter le problème de sur ajustement.

Après 50 exécutions avec des fractionnements aléatoires d’ensembles de données d’entraînement/de test, un modèle ajusté peut être écrit sous la forme :

$$GP-SCIQ = 0,23 * GFM + 0,43 * SVQI + 0,37 * ESIM - 1,9 \quad (1)$$

On peut voir que le modèle obtenu n’a utilisé que trois métriques parmi les sept utilisées.

5.6 Évaluation expérimentale

5.6.1 Évaluation des performances de prédiction

Pour montrer l’allure des scores objectifs vis-à-vis des scores subjectifs, on utilise le graphe de nuage de points entre ces deux scores pour l’ensemble des images de la base SIQAD. La Figure 17 montre le graphe pour notre nouvelle méthode et aussi pour les sept métriques utilisées dans la fusion.

De plus, un ajustement avec une fonction logistique comme suggéré dans (Sheikh et al., 2006) et (MRohaly, 2000) est montré pour une comparaison plus facile. Chaque point du tracé représente une image dans le benchmark, l’axe horizontal correspond au score objectif pour cette image et l’axe vertical correspond à son score subjectif.

On peut voir, dans le plot de notre méthode GPSCIQ, que le pourcentage de valeurs aberrantes est diminué et une tendance à un comportement monotone est augmentée contrairement aux autres méthodes par exemple SCQI, SCDM et GSS. Nous constatons que les scores objectifs prédits par la méthode proposée ont une forte cohérence avec les scores subjectifs donc le résultat de notre méthode est adéquat à la perception humaine.

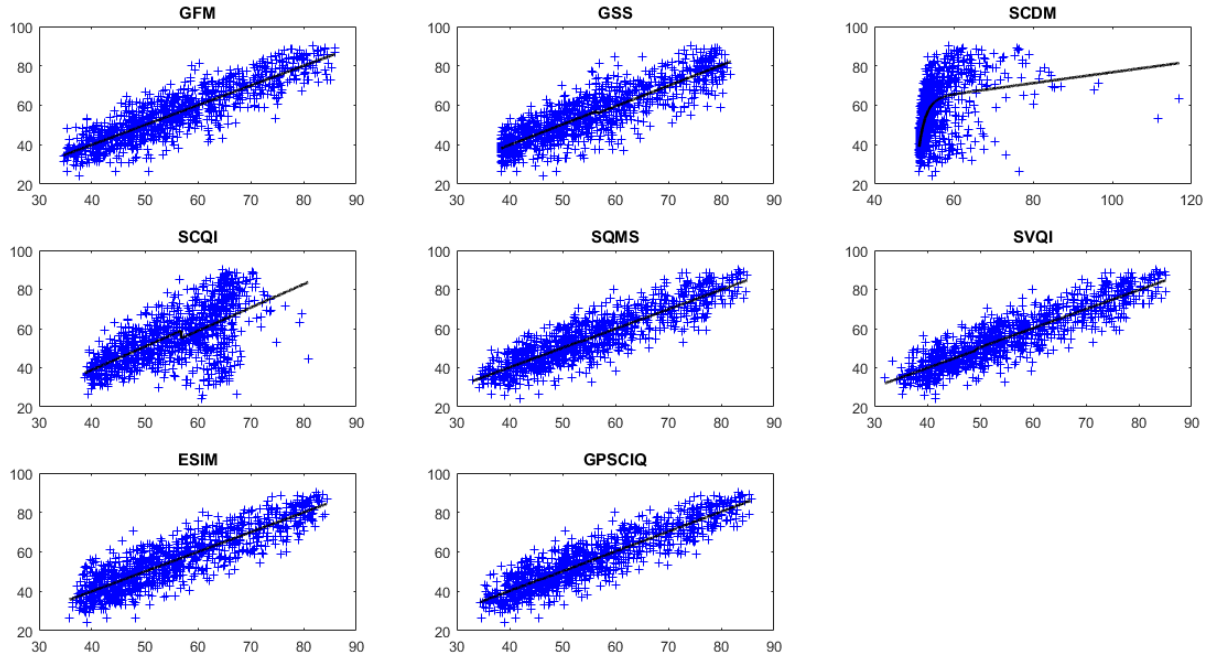


Figure 5.3. - Corrélations entre l'évaluation subjective et le GPSCIQ obtenu et les sept métriques IQA de pointe pour l'ensemble des images de SIQAD

Pour l'évaluation et le développement du modèle développé, selon le protocole largement utilisé (MRohaly, 2000; Sheikh et al., 2006), nous utilisons les quatre indices suivants de précision de prédiction, de monotonie et de cohérence pour mesurer les performances du modèle IQA. L'erreur quadratique moyenne (RMSE), le coefficient de corrélation linéaire de Pearson (PCC), le coefficient de corrélation d'ordre de Spearman (SRCC) et le coefficient de corrélation d'ordre de Kendall (KRCC).

Afin d'obtenir l'authenticité et la validité de la comparaison, il est nécessaire (avant de calculer les quatre indices), de s'assurer que les scores de prédiction des différents modèles IQA se situent sur la même échelle. Par conséquent, selon la suggestion du groupe d'experts en qualité vidéo (VQEG) (Sheikh et al., 2006), nous adoptons une fonction de régression non linéaire entre un vecteur de scores objectifs, et des scores subjectifs.

Le Tableau 5.1 montre les résultats de l'évaluation dans les deux benchmarks, entre notre nouvelle méthode et les méthodes utilisées pour la fusion. Le meilleur modèle pour chaque critère est indiqué en gras. Les valeurs de RMSE plus petites sont considérées comme meilleures, contrairement à SRCC, KRCC et PCC dont les valeurs devraient être

plus grandes. Les résultats montrent que notre méthode (GP-SCIQ) est clairement un modèle très performant par rapport aux métriques individuelles.

Tableau 5.1.- Comparaison des performances du modèle proposé avec les sept métriques utilisées dans la fusion

	GFM	GSS	SCDM	SCQI	SQMS	SVQI	ESIM	GP-SCIQ
SIQAD (980 images)								
SPCC	0,8734	0,8436	0,6124	0,6112	0,8803	0,8835	0,8632	0,8917
KCC	0,6875	0,6486	0,4565	0,4546	0,6935	0,6985	0,6742	0,7094
PCC	0,8831	0,8515	0,4517	0,6138	0,8872	0,891	0,8791	0,9005
RMSE	6,7158	7,5051	12,7703	11,3	6,6039	6,4961	6,8220	6,2239
SCID (1800 images)								
SPCC	0,8758	0,749	0,7715	0,7814	0,832	0,8385	0,8486	0,8802
KCC	0,6843	0,5422	0,5737	0,5826	0,6429	0,6501	0,6517	0,6946
IPCC	0,876	0,7609	0,7781	0,7854	0,8563	0,861	0,8642	0,8875
RMSE	6,8296	9,189	8,8956	8,7646	7,3134	7,2026	7,1259	6,5252
Overall direct								
SPCC	0,8746	0,7963	0,6919	0,6963	0,8561	0,861	0,8559	0,8860
KCC	0,6859	0,5954	0,5151	0,5186	0,6682	0,6743	0,6630	0,7020
PCC	0,8795	0,8062	0,6149	0,6996	0,8717	0,876	0,8717	0,8940
RMSE	6,7727	8,3470	10,8329	10,0323	6,9586	6,8493	6,9740	6,3746
Overall weighted								
SPCC	0,8750	0,7823	0,7154	0,7214	0,8490	0,8544	0,8537	0,8843
KCC	0,6854	0,5797	0,5324	0,5375	0,6607	0,6672	0,6596	0,6998
PCC	0,8785	0,7928	0,6630	0,7249	0,8672	0,8716	0,8695	0,8921
RMSE	6,7895	8,5954	10,2615	9,6584	7,0633	6,9535	7,0188	6,4190

Le tableau contient également des moyennes directes et pondérées des valeurs obtenues. Pour la moyenne pondérée, le nombre d’images dans la base de données est utilisé comme son poids. Les résultats montrent les performances exceptionnelles de notre nouveau modèle.

5.6.2 Comparaisons de performances avec des méthodes SCI sur SIQAD

Ici, le GP-SCIQ proposé est comparé, sur la base de données SIQAD, avec d’autres récentes approches développées spécifiquement pour les SCI à savoir : SFUW (Fang et al., 2017), PICNN (Chen et al., 2018), MDOGS (Fu et al., 2018), FQI (Rahul and Tiwari, 2019), SIQADFII (Jiang et al., 2020a), Xia19 (Xia et al., 2020), SPSIM (Yang et al., 2021), WS-HV (Cheraaqee et al., 2021). Les résultats des différents critères de corrélation sont présentés sur le Tableau 5.2.

Nous observons que le GP-SCIQ proposé surpasse toutes les autres méthodes, ce qui est démontré par les valeurs PLCC et SRCC les plus élevées et la valeur RMSE la plus petite. Bien que le SRCC de notre GP-SCIQ soit très proche au SRCC de PICNN, mais il le surpasse en termes de PLCC et RMSE.

Tableau 5.2. - Résultats expérimentaux de la méthode proposée et d’autres méthodes actuelles sur la base de données SIQAD

	SFUW (2017)	PICNN (2018)	MDOGS (2018)	FQI (2019)	SIQADFII (2020)	Xia19 (2020)	SPSIM (2021)	WS-HV (2021)	GPSCIQ
SRCC	0.8800	0.897	0.8822	0.7729	0.8880	0.8213	0.8810	0.8851	0.8917
PLCC	0.8910	0.896	0.8839	0.8018	0.9000	0.8343	0.8920	0.8876	0.9005
RMSE	6.4990	6.790	6.6951	8.7312	6.2422	7.8924	6.4713	6.5925	6.2239

5.6.3 Évaluation croisée

Par ailleurs, pour valider la robustesse et la capacité de généralisation de notre modèle, nous avons entraîné le système sur SIQAD, et le testé sur le sous-ensemble de SCID. Considérant que les deux bases d’images contiennent différents types de distorsion, les résultats expérimentaux de notre modèle sont donnés sur 6 types de distorsion communs constitués de GN, GB, MB, CC, JPEG et JPEG2000. Tous les SCI déformés de six types de distorsion dans SIQAD sont utilisés comme ensemble d’apprentissage.

L’évaluation croisée (en termes de SPCC) de notre modèle proposé est comparée aux nouveaux modèles de pointe suivants : QODCNN-FR (Jiang et al., 2019), RIQA (Jiang

et al., 2020b), SIQA-DFII (Jiang et al., 2020a) et MTDL-SCI2021 (Gao et al., 2021), les valeurs de SPCC sont tirées des articles publiés à l'origine.

À partir du Tableau 5.3, on peut observer que notre modèle obtient une augmentation de performance exceptionnelle, ce qui implique que le modèle proposé possède une plus grande capacité de généralisation.

Tableau 5.3. - Évaluation croisée (Apprentissage sur SIQAD et test sur SCID)

	GP-SCIQ	QODCNN- FR 2019	RIQA 2020	SIQA-DFII 2020	MTDL-SCI 2021
SPCC	0,9126	0,876	0.8771	0,8507	0.8186

5.6.4 Importance statistique

Pour quantifier si la différence numérique entre les performances de l'algorithme IQA était statistiquement significative, une méthode F-test (Sheikh et al., 2006) est réalisée avec un niveau de signification de 0,05.

F-test est un test statistique typique qui suit l'hypothèse basée sur la variance et révèle des informations supplémentaires sur les performances relatives de différents modèles d'évaluation de la qualité d'image. Plus précisément, F-test est calculé sur la base des différences résiduelles des notes de qualité subjectives et des prédictions de qualité objectives associées (après un mappage non linéaire) (Sheikh et al., 2006).

Les résultats de la performance statistique de GP-SCIQ par rapport à d'autres algorithmes IQA sur les deux bases de données SIQAD et SCID sont présentés dans le Tableau 5.4. Chaque entrée du tableau est un mot de code composé de deux symboles ; le premier symbole désigne le résultat sur la base SIQAD et le deuxième c'est le résultat sur la base SCID.

Une valeur de «1» indique que la mesure de la ligne est significativement supérieure à celle de la colonne, une valeur de «0» indique que la mesure de la ligne est significativement comparable à celle de la colonne.

Pour les deux bases d'images, les résultats de notre modèle sont tous «1». En résumé, les tests de signification confirment les excellentes performances du modèle développé par rapport à toutes les méthodes répertoriées.

Tableau 5.4. - Tests de signification statistique sur SIQAD et SCID

SIQAD/SCID	GFM	GSS	SCQI	SQMS	SVQI	ESIM	GP-SCIQ
GFM	0 0	1 1	1 1	0 1	0 1	0 0	0 0
GSS	0 0	0 0	0 1	0 0	0 0	0 0	0 0
SCQI	0 0	0 0	0 0	0 0	0 0	0 0	0 0
SQMS	0 0	1 1	1 1	0 0	0 0	1 0	0 0
SVQI	0 0	1 1	1 1	0 0	0 0	1 0	0 0
ESIM	0 0	0 0	1 1	0 0	0 0	0 0	0 0
GP-SCIQ	1 1	1 1	1 1	1 1	1 1	1 1	0 0

5.7 Conclusion

Dans ce chapitre nous avons appliqué une technique de régression symbolique utilisant la programmation génétique multigène (MGGP) pour le cas des images de contenu d'écran (SCI). La MGGP peut déterminer les métriques de qualité d'image appropriées, à partir de sept métriques, dont les scores objectifs sont utilisés comme prédicateurs.

Les performances du modèle proposé sont évaluées à l'aide de deux bases d'images SCI les plus utilisées (SIQAD et SCID). Les résultats atteignent des excellentes performances par rapport à de nombreuses méthodes.

Conclusion Générale

Les progrès technologiques ont permis une connectivité toujours plus grande et abordable, ainsi que le développement d'une multitude d'appareils mobiles, menant à un monde bien connecté. Une proportion de plus en plus importante de la population mondiale accède désormais au contenu visuel via Internet à diverses fins telles que la communication, le divertissement, l'éducation, le sport, le partage sur les réseaux sociaux, etc.

Cependant, le contenu visuel subit un certain nombre de distorsions au cours des processus d'acquisition, de stockage, de transmission sous contraintes de bande passante et d'affichage, dont chacune peut dégrader sa qualité perçue. Compte tenu du rôle important qu'un tel contenu est venu jouer dans notre vie, l'évaluation de la qualité de l'image et de la vidéo perceptive, visant à évaluer la qualité du contenu visuel telle qu'un humain le percevrait, est devenue un problème fondamental pour la conception, l'optimisation et l'évaluation de divers algorithmes et systèmes de traitement d'images et de vidéos.

Ainsi, le développement d'algorithmes objectifs d'évaluation de la qualité qui sont connus pour avoir une bonne corrélation avec la qualité perceptive du contenu, été la cible d'intenses recherches. Au cours des deux dernières décennies, de nombreux travaux ont été menés sur le terrain. Par conséquent, il est devenu nécessaire de faire une revue sur les différentes métriques pour combler les lacunes des métriques existantes.

Pour cette fin, nous avons consacré le premier chapitre de notre thèse à la présentation des axes de recherche du domaine d'évaluation de qualité d'image et de vidéo ainsi que les méthodes d'évaluation existantes dans la littérature.

Le problème de prédiction de qualité d'image est souvent considéré comme un problème d'optimisation qui est traité par les techniques d'apprentissage et les métaheuristiques d'optimisation. De ce fait nous avons présenté dans le deuxième chapitre un état de l'art sur le problème d'optimisation et notamment les métaheuristiques pour bien clarifier leurs

caractéristiques. Cette étude nous a permis de choisir la méthode la plus appropriée pour notre cas de recherche.

Nous avons, lors de notre réalisation de cette thèse, exploité la flexibilité de la programmation génétique multigène (MGGP) pour proposer de nouvelles métriques de qualité d'image (IQM). La MGGP peut exploiter un vaste espace de recherche composé de toutes les combinaisons possibles de scores objectifs de plusieurs métriques IQM et trouver le meilleur modèle de régression, compte tenu de sa simplicité ainsi que de ses performances sur les données d'apprentissage (multi-objectives), qui décrit la somme pondérée des scores objectifs IQM pour prédire les scores subjectifs des images dans des ensembles de données, sans faire d'hypothèses a priori sur la structure du modèle.

Pour évaluer les performances de la méthode proposée, plusieurs expérimentations ont été menées, avec des aspects différents, en utilisant six plus grands benchmarks d'images (LIVE, CSIQ, TID2008, TID2013, IVC et MDID) et quatre indices de performance (SRCC, PCC, KROCC, RMSE). Pour chaque base de référence, nous avons obtenu un modèle prédit, à savoir la programmation génétique multigène basée sur une métrique de référence complète (FR-MGGP). Les résultats présentés confirment les performances supérieures des FR-MGGP obtenus par rapport aux mesures IQA existantes, y compris d'autres approches de fusion de métriques récemment publiées.

Nous avons aussi étendu notre approche aux images de contenu d'écran, ce type d'image est récemment très utilisé dans divers systèmes de traitement. Les résultats obtenus ont été comparés à d'autres approches sous différents aspects et leurs excellentes performances sont prouvées.

Les travaux présentés dans cette thèse peuvent être élargis de différentes manières, telle que l'élaboration d'une :

- Extension pour des images Hybrides (Images naturelles et images du contenu d'écran) sans distinguer les divers types d'images évaluées.
- Extension à une méthode d'évaluation de la qualité d'image 3D.
- Extension de la méthode proposée à une mesure d'évaluation de qualité de vidéo.

Références Bibliographiques

- Abdel-Basset, M., Abdel-Fatah, L., and Sangaiah, A. (2018). Chapter 10 - metaheuristic algorithms : A comprehensive review. In Sangaiah, A., Sheng, M., and Zhang, Z., editors, *Computational Intelligence for Multimedia Big Data on the Cloud with Engineering Applications*, pages 185–231. Academic Press.
- Ahmed, I., Der, C., and Hammad, B. (2016). A survey of recent approaches on no-reference image quality assessment with multiscale geometric analysis transforms. *International Journal of Scientific & Engineering Research*, 7(10).
- Ahvanooey, M., Li, Q., Wu, M., and Wang, S. (2019). A survey of genetic programming and its applications. *KSII Transactions on Internet and Information Systems (TIIS)*, 13 :1765–94.
- Akhtar, A. (2019). Evolution of ant colony optimization algorithm—a brief literature review. *arXiv preprint arXiv :1908.08007*.
- Al-Bandawi, H. and Deng, G. (2018). Blind image quality assessment based on benford’s law. *IET Image Processing*, 12 :1983–93.
- Appina, B. and Channappayya, S. (2018). Full-reference 3-d video quality assessment using scene component statistical dependencies. *IEEE Signal Processing Letters*, 25 :823–7.
- Ari, D. and Alagöz, B. (2021). A review of genetic programming : Popular techniques, fundamental aspects, software tools and applications. *Sakarya University Journal of Science*, 25 :397–416.
- Azencott, R. (1992). *Simulated annealing : parallelization techniques*, volume 27. Wiley-Interscience.
- Bae, S.-H. and Kim, M. (2016). A novel image quality assessment with globally and locally consistent visual quality perception. *IEEE Trans on Image Process*, 25 :2392–406.
- Bampis, C., Li, Z., and Bovik, A. (2018). Spatiotemporal feature integration and model fusion for full reference video quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, 29 :2256–70.

- Barri, A., Dooms, A., Jansen, B., and Schelkens, P. (2014). A locally adaptive system for the fusion of objective quality measures. *IEEE Transactions on Image Processing*, 23 :2446–58.
- Becker, S., Müller, K.-R., Wiegand, T., and Bosse, S. (2019). A neural network model of spatial distortion sensitivity for video quality estimation. In *2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, pages 1–6.
- Blum, C. and Roli, A. (2003). Metaheuristics in combinatorial optimization : Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35 :268–308.
- Borhen Eddine, D., Fella, H., and Azeddine, B. (2016). Blind image quality assessment method based on a particle swarm optimization support vector regression fusion scheme. *J Electron Imaging*, 25(61623).
- BT, R.-R. (2002). *Methodology for the subjective assessment of the quality of television pictures*. International Telecommunication Union.
- Burnet, F. (1959). *The clonal selection theory of acquired immunity*.(1959. Cambridge Univer-Sity Press.
- Chandler, D. (2010). Most apparent distortion : full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(11006).
- Chandler, D. (2013). Seven challenges in image quality assessment : Past, present, and future research. *ISRN Signal Processing*, 2013 :1–53.
- Chang, H., Zhang, Q., Wu, Q., and Gan, Y. (2015). Perceptual image quality assessment by independent feature detector. *Neurocomputing*, 151 :1142–52.
- Chang, H.-W., Yang, H., Gan, Y., and Wang, M.-H. (2013). Sparse feature fidelity for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 22 :4007–18.
- Chen, C., Choi, L., De Veciana, G., Caramanis, C., Heath, R., and Bovik, A. (2014). Modeling the time—varying subjective quality of HTTP video streams with rate adaptations. *IEEE Transactions on Image Processing*, 23 :2206–21.
- Chen, J., Shen, L., Zheng, L., and Jiang, X. (2018). Naturalization module in neural networks for screen content image quality assessment. *IEEE Signal Process Lett*, 25 :1685–9.
- Cheraaqee, P., Maviz, Z., Mansouri, A., and Mahmoudi-Aznaveh, A. (2021). Quality assessment of screen content images in wavelet domain. *IEEE Trans Circuits Syst Video Technol*, pages 1–1.

- Chetouani, A., Beghdadi, A., and Deriche, M. (2010). A universal full reference image quality metric based on a neural fusion approach. In *2010 IEEE International Conference on Image Processing*, pages 2517–20. IEEE.
- Chikkerur, S., Sundaram, V., Reisslein, M., and Karam, L. (2011). Objective video quality assessment methods : A classification, review, and performance comparison. *IEEE Transactions on Broadcasting*, 57 :165–82.
- Chong, E. and Zak, S. (2004). *An introduction to optimization*. John Wiley & Sons.
- Chong, E. and Zak, S. (2013). *An introduction to optimization*. John Wiley & Sons, fourth edition.
- Chopard, B. and Tomassini, M. (2018). *An introduction to metaheuristics for optimization*. Springer.
- Courat, J.-P., Raynaud, G., Mrad, I., and Siarry, P. (1994). Electronic component model minimization based on log simulated annealing. *IEEE Transactions on Circuits and Systems I : Fundamental Theory and Applications*, 41 :790–5.
- Damera-Venkata, N., Kite, T., Geisler, W., Evans, B., and Bovik, A. (2000). Image quality assessment based on a degradation model. *IEEE Transactions on Image Processing*, 9 :636–50.
- Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm : NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6 :182–97.
- Deneubourg, J.-L., Aron, S., Goss, S., and Pasteels, J. (1990). The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior*, 3 :159–68.
- Dihin, R., Hamza, N., and Toman, Z. (2020). Full-reference facial image quality assessment and identification by two proposed measures. *Journal of Southwest Jiaotong University*, 55(12).
- Dorigo, M. and Gambardella, L. (1997). Ant colony system : a cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1 :53–66.
- Eberhart, R. and Kennedy, J. (1995). A new optimizer using particle swarm theory. In *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*, pages 39–43. Ieee.

- Ebrahimi, S., Salahshour, E., Malekzadeh, M., and Gordillo, F. (2019). Parameters identification of PV solar cells and modules using flexible particle swarm optimization algorithm. *Energy*, 179 :358–72.
- Elsheikh, A. and Abd Elaziz, M. (2019). Review on applications of particle swarm optimization in solar energy systems. *International Journal of Environmental Science and Technology*, 16 :1159–70.
- Emmerich, M. and Deutz, A. (2018). A tutorial on multiobjective optimization : fundamentals and evolutionary methods. *Natural Computing*, 17 :585–609.
- Fang, Y., Yan, J., Liu, J., Wang, S., Li, Q., and Guo, Z. (2017). Objective quality assessment of screen content images by uncertainty weighting. *IEEE Trans on Image Process*, 26 :2016–27.
- Fastowicz, J. and Okarma, K. (2018). Fast quality assessment of 3d printed surfaces based on structural similarity of image regions. In *International Interdisciplinary PhD Workshop (IIPhDW)*, *IEEE*, pages 401–6.
- Feo, T. and Resende, M. (1989). A probabilistic heuristic for a computationally difficult set covering problem. *Operations Research Letters*, 8 :67–71.
- Feo, T. and Resende, M. (1995). Greedy randomized adaptive search procedures. *Journal of Global Optimization*, 6 :109–33.
- Ferreira, C. (2001). Gene expression programming : a new adaptive algorithm for solving problems. *arXiv preprint cs/0102027*.
- Festa, P. and Resende, M. (2009a). An annotated bibliography of GRASP–part i : Algorithms. *International Transactions in Operational Research*, 16 :1–24.
- Festa, P. and Resende, M. (2009b). An annotated bibliography of GRASP–part II : Applications. *International Transactions in Operational Research*, 16 :131–72.
- Forrest, S. and Mitchell, M. (1993). Relative building-block fitness and the building-block hypothesis. *Foundations of genetic algorithms*, 2 :109–26. Publisher : Elsevier.
- Fu, Y., Zeng, H., Ma, L., Ni, Z., Zhu, J., and Ma, K.-K. (2018). Screen content image quality assessment using multi-scale difference of gaussian. *IEEE Trans Circuits Syst Video Technol*, 28 :2428–32.
- Gandomi, A. and Alavi, A. (2012). A new multi-gene genetic programming approach to nonlinear system modeling. part i : materials and structural engineering problems. *Neural Comput & Applic*, 21 :171–87.

- Ganesh, K. and Patil, C. (2022). Video quality assessment using optimization algorithms. *International Journal*, 7.
- Gao, R., Huang, Z., and Liu, S. (2021). Multi-task deep learning for no-reference screen content image quality assessment. In Lokoč, J., Skopal, T., Schoeffmann, K., Mezaris, V., Li, X., and Vrochidis, S., editors, *MultiMedia Modeling*, pages 213–26. Springer International Publishing.
- Gendreau, M. and Potvin, J.-Y. (2005). Metaheuristics in combinatorial optimization. *Annals of Operations Research*, 140(1) :189–213.
- Ghani, R. and Ajrash, A. (2018). Quality of experience metric of streaming video : A survey. *Iraqi Journal of Science*, pages 1531–7.
- Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computers & Operations Research*, 13 :533–49.
- Glover, F. (1990). Artificial intelligence, heuristic frameworks and tabu search. *Managerial and Decision Economics*, 11 :365–75.
- Glover, F. and Laguna, M. (1997). Tabu search kluwer academic. *Boston, Texas*.
- Gu, K., Qiao, J., Min, X., Yue, G., Lin, W., and Thalmann, D. (2018). Evaluating quality of screen content images via structural variation analysis. *IEEE Trans Visual Comput Graphics*, 24 :2689–701.
- Gu, K., Wang, S., Yang, H., Lin, W., Zhai, G., and Yang, X. (2016a). Saliency-guided quality assessment of screen content images. *IEEE Trans Multimedia*, 18 :1098–110.
- Gu, K., Zhai, G., Lin, W., Yang, X., and Zhang, W. (2016b). Learning a blind quality evaluation engine of screen content images. *Neurocomputing*, 196 :140–9.
- Gustafson, S., Burke, E., and Krasnogor, N. (2005). On improving genetic programming for symbolic regression. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 912–9. IEEE.
- Hart, J. and Shogan, A. (1987). Semi-greedy heuristics : An empirical study. *Operations Research Letters*, 6 :107–14.
- Hastings, W. (1970). Monte carlo sampling methods using markov chains and their applications.
- Hegazy, T. and Al Regib, G. (2014). COHERENSi : A new full-reference IQA index using error spectrum chaos. In *2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 965–9. IEEE.

- Hemami, S. and Reibman, A. (2010). No reference image and video quality estimation : applications and human motivated design. *Signal Processing : Image Communication*, 25 :469–481.
- Hii, C., Searson, D., and Willis, M. (2011). Evolving toxicity models using MultigeneSymbolic regression and multiple objectives. *International Journal of Machine Learning and Computing*, pages 30–5.
- Ieremeiev, O., Lukin, V., Ponomarenko, N., and Egiazarian, K. (2018). Robust linearized combined metrics of image visual quality. *Electronic Imaging*, 2018 :260–1.
- Ieremeiev, O., Lukin, V., Ponomarenko, N., and Egiazarian, K. (2019). Combined no-reference IQA metric and its performance analysis. *Electronic Imaging*, 2019 :260–1.
- Injeti, S. and Thunuguntla, V. (2020). Optimal integration of DGs into radial distribution network in the presence of plug-in electric vehicles to minimize daily active power losses and to improve the voltage profile of the system using bio-inspired optimization algorithms. *Protection and Control of Modern Power Systems*, 5 :1–15.
- Izima, O., de, F. R., and Malik, A. (2021). A survey of machine learning techniques for video quality prediction from quality of delivery metrics. *Electronics*, 10(2851).
- Jiang, X., Shen, L., Ding, Q., Zheng, L., and An, P. (2020a). Screen content image quality assessment based on convolutional neural networks. *Journal of Visual Communication and Image Representation*, 67(102745).
- Jiang, X., Shen, L., Feng, G., Yu, L., and An, P. (2019). Deep optimization model for screen content image quality assessment using neural networks. *arXiv preprint arXiv :1903.00705*.
- Jiang, X., Shen, L., Yu, L., Jiang, M., and Feng, G. (2020b). No-reference screen content image quality assessment based on multi-region features. *Neurocomputing*, 386 :30–41.
- Jordehi, A. (2016). Time varying acceleration coefficients particle swarm optimisation (TVACPSO) : A new optimisation algorithm for estimating parameters of PV cells and modules. *Energy Conversion and Management*, 129 :262–74.
- Jordehi, A. (2018). Enhanced leader particle swarm optimisation (ELPSO) : An efficient algorithm for parameter estimation of photovoltaic (PV) cells and modules. *Solar Energy*, 159 :78–87.
- Khaing, Y., Sugiura, Y., and Shimamura, T. (2019). Combination of dissimilar feature scores for image quality assessment using particle swarm optimization algorithm. *Journal of Signal Processing*, 23 :205–14.

- Khan, A., Qureshi, A., Wahab, N., Hussain, M., and Hamza, M. (2021). A recent survey on the applications of genetic programming in image processing. *Computational Intelligence*, 37 :1745–78.
- Kim, W., Jongyoo, K., Ahn, S., Jinwoo, K., and Lee, S. (2018). Deep video quality assessor : From spatio-temporal visual sensitivity to a convolutional neural aggregation network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 219–34.
- Kirkpatrick, S., Gelatt, C., and Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220 :671–80.
- Koza, J. (1992). *Genetic programming : on the programming of computers by means of natural selection*. MIT Press.
- Lahouhou, A., Viennet, E., and Beghdadi, A. (2010). Selecting low-level features for image quality assessment by statistical methods. *Journal of Computing and Information Technology*, 18(183).
- Lei, F., Li, S., Xie, S., and Liu, J. (2022). Subjective and objective quality assessment of swimming pool images. *Front Neurosci*, 15(766762).
- Leszczuk, M. and Janowski, L. (2021). New recommendation for subjective video quality assessment methods for recognition tasks. *ENGINEERING*.
- LI, D., JIANG, T., and JIANG, M. (2019). Recent advances and challenges in video quality assessment. *ZTE COMMUNICATIONS*, 17.
- Li, Z., Aaron, A., Katsavounidis, I., Moorthy, A., and Manohara, M. (2016). Toward a practical perceptual video quality metric. *The Netflix Tech Blog*, 6.
- Lin, Z., Lei, Z., and X, M. (2010). RFSIM : A feature based image quality assessment metric using riesz transforms. In *2010 IEEE International Conference on Image Processing*, pages 321–4. IEEE.
- Ling, S. and Le Callet, P. (2017). Image quality assessment for free viewpoint video based on mid-level contours feature. In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, pages 79–84.
- Liu, A., Lin, W., and M, N. (2012). Image quality assessment based on gradient similarity. *IEEE Transactions on Image Processing*, 21 :1500–12.
- Liu, M. and Yang, X. (2008). A new image quality approach based on decision fusion. In *2008 Fifth International Conference on Fuzzy Systems and Knowledge Discovery*, pages 10–4. IEEE.

- LN, D. C. and FJ, V. Z. (2002). Learning and optimization using the clonal selection principle. *IEEE Transactions on Evolutionary Computation*, 6 :239–51.
- Lourenço, H., Martin, O., and Stützle, T. (2003). *Iterated local search. Handbook of metaheuristics*. Springer.
- Lukin, VV, P., NN, I., OI, E., KO, A., and J. (2015). Combining full-reference image visual quality metrics by neural network. In H, R., editor, *Rogowitz BE, Pappas TN*, page 93940.
- Lévêque, L., Zhang, W., Cavarro-Ménard, C., Le Callet, P., and Liu, H. (2017). Study of video quality assessment for telesurgery. *IEEE Access*, 5 :9990–9.
- Ma, J., Man, K., Guan, S.-U., Ting, T., and Wong, P. (2016). Parameter estimation of photovoltaic model via parallel particle swarm optimization algorithm. *International Journal of Energy Research*, 40 :343–52.
- Ma, J., Nakarmi, U., Kin, C., Sandino, C., Cheng, J., and Syed, A. (2020). Diagnostic image quality assessment and classification in medical imaging : Opportunities and challenges. In *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI*, pages 337–40. IEEE.
- Madhusudana, P., Birkbeck, N., Wang, Y., Adsumilli, B., and Bovik, A. (2021a). High frame rate video quality assessment using VMAF and entropic differences. In *2021 Picture Coding Symposium (PCS), IEEE; 2021b*, pages 1–5.
- Madhusudana, P., Birkbeck, N., Wang, Y., Adsumilli, B., and Bovik, A. (2021b). ST-GREED : Space-time generalized entropic differences for frame rate dependent video quality prediction. *IEEE Transactions on Image Processing*, 30 :7446–57.
- Malik, H., Iqbal, A., Joshi, P., Agrawal, S., and Bakhsh, F. (2021). *Metaheuristic and evolutionary computation : algorithms and applications*. Springer.
- Manasa, K. and Channappayya, S. (2016). An optical flow-based full reference video quality assessment algorithm. *IEEE Transactions on Image Processing*, 25 :2480–92.
- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21 :1087–92.
- Mittal, A., Moorthy, A., and Bovik, A. (2012). No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing*, 21 :4695–708.
- Mohammadi, P., Ebrahimi-Moghadam, A., and Shirani, S. (2014). Subjective and objective quality assessment of image : A survey.

- Montana, D. (1995). Strongly typed genetic programming. *Evolutionary Computation*, 3 :199–230.
- Moorthy, A. and Bovik, A. (2011). Blind image quality assessment : From natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, 20 :3350–64.
- Moorthy, A., Choi, L., Bovik, A., and De Veciana, G. (2012). Video quality assessment on mobile devices : Subjective, behavioral and objective studies. *IEEE Journal of Selected Topics in Signal Processing*, 6 :652–71.
- Moraglio, A., Krawiec, K., and Johnson, C. (2012). Geometric semantic genetic programming. In *International Conference on Parallel Problem Solving from Nature*, pages 21–31. Springer.
- MRohaly (2000). *VQEG : Final Report from the video quality experts group on the Validation of objective models of video quality assessment*. //Www.Vqeg.Org.
- Nazari-Heris, M., Mohammadi-Ivatloo, B., and Gharehpetian, G. (2018). A comprehensive review of heuristic optimization algorithms for optimal combined heat and power dispatch from economic and environmental perspectives. *Renewable and Sustainable Energy Reviews*, 81 :2128–43.
- Ni, Z., Ma, L., Zeng, H., Cai, C., and Ma, K.-K. (2016). Gradient direction for screen content image quality assessment. In *IEEE SIGNAL PROCESSING LETTERS 2016 ;23* :5.
- Ni, Z., Ma, L., Zeng, H., Chen, J., and Cai, C. (2017). ESIM : Edge similarity for screen content image quality assessment. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 26(14).
- Ni, Z., Zeng, H., Ma, L., Hou, J., Chen, J., and Ma, K.-K. (2018). A gabor feature-based quality assessment model for the screen content images. *IEEE Trans on Image Process*, 27 :4516–28.
- Ninassi, A., Le Callet, P., and Autrusseau, F. (2006). Pseudo no reference image quality metric using perceptual data hiding. Pages : 60570 Place : San Jose, CA.
- Ninassi, A., Le Meur, O., Le Callet, P., and Barba, D. (2008). On the performance of human visual system based image quality assessment metric using wavelet domain. *SPIE Conference Human Vision and Electronic Imaging*, XIII, vol. 6806 :680610 1–680610 12.
- Nunes, H., Pombo, J., Mariano, S., Calado, M., and De Souza, J. (2018). A new high performance method for determining the parameters of PV cells and modules based on guaranteed convergence particle swarm optimization. *Applied Energy*, 211 :774–91.

- Okarma, K. (2010). Combined full-reference image quality metric linearly correlated with subjective assessment. In Rutkowski, L., Scherer, R., Tadeusiewicz, R., Zadeh, L., and Zurada, J., editors, *Artificial Intelligence and Soft Computing*, volume 6113, pages 539–46. Springer.
- Okarma, K. (2012). Combined image similarity index. *Optical Review*, 19 :349–54.
- Okarma, K. (2013). Extended hybrid image similarity – combined full-reference image quality metric linearly correlated with subjective scores. *Electronics and Electrical Engineering*, 19.
- Okarma, K. (2014). Quality assessment of images with multiple distortions using combined metrics. *ELAE*, 20 :128–31.
- Okarma, K. (2019). Current trends and advances in image quality assessment. *ELAE*, 25 :77–84.
- Ortiz-Jaramillo, B., Kumcu, A., Platisa, L., and Philips, W. (2014). A full reference video quality measure based on motion differences and saliency maps evaluation. In *2014 International Conference on Computer Vision Theory and Applications (VISAPP)*, volume 2, pages 714–22. IEEE.
- Osman, I. and Laporte, G. (1996). *Metaheuristics : A bibliography*. Springer.
- Oszust, M. (2016a). Decision fusion for image quality assessment using an optimization approach. *IEEE Signal Processing Letters*, 23 :65–9.
- Oszust, M. (2016b). Full-reference image quality assessment with linear combination of genetically selected quality measures. Type : PLOS ONE 2016b ;11 :e0158333.
- Oszust, M. (2016c). A regression-based family of measures for full-reference image quality assessment. *Measurement Science Review*, 16 :316–25.
- O’Neill, M. and Ryan, C. (2001). Grammatical evolution. *IEEE Transactions on Evolutionary Computation*, 5 :349–58.
- Papadimitriou, C. (1976). *The complexity of combinatorial optimization problems*. Princeton University.
- Patel, V., Savsani, V., and Tawhid, M. (2019). *Thermal System Optimization : A Population-Based Metaheuristic Approach*. Springer International Publishing.
- Pei, S.-C. and Chen, L.-H. (2015). Image quality assessment using human visual DOG model fused with random forest. *IEEE Transactions on Image Processing*, 24 :3282–92.

- Peng, P. and Li, Z.-N. (2012). A mixture of experts approach to multi-strategy image quality assessment. In Campilho, A. and Kamel, M., editors, *Image Analysis and Recognition*, volume 7324, pages 123–30. Springer.
- Phadikar, B., Maity, G., and Phadikar, A. (2018). Full reference image quality assessment : A survey. In Bhattacharyya, S., Sen, S., Dutta, M., Biswas, P., and Chattopadhyay, H., editors, *Industry Interactive Innovations in Science, Engineering and Technology*, volume 11, pages 197–208. Springer Singapore.
- Pinson, M., Choi, L., and Bovik, A. (2014). Temporal video quality model accounting for variable frame delay distortions. *IEEE Transactions on Broadcasting*, 60 :637–49.
- Pinson, M. and Wolf, S. (2004). A new standardized method for objectively measuring video quality. *IEEE Transactions on Broadcasting*, 50 :312–22.
- Piórkowski, R., Mantiuk, R., and Siekawa, A. (2017). Automatic detection of game engine artifacts using full reference image quality metrics.
- Ponomarenko, N., Jin, L., Ieremeiev, O., Lukin, V., Egiazarian, K., and Astola, J. (2015). Image database TID2013 : Peculiarities, results and perspectives. *Signal Processing : Image Communication*, 30 :57–77.
- Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Astola, J., and Carli, M. (2009). TID2008 – a database for evaluation of full- reference visual quality assessment metrics. *Advances of Modern Radioelectronics*, 2009 :30–45.
- Prabhushankar, M., Temel, D., and AlRegib, G. (2017). MS-UNIQUE : Multi-model and sharpness-weighted unsupervised image quality estimation. *Electronic Imaging*, 2017 :30–5.
- Prajapati, V., Jain, M., and Chouhan, L. (2020). Tabu search algorithm (TSA) : A comprehensive survey. In *3rd International Conference on Emerging Technologies in Computer Engineering : Machine Learning and Internet of Things (ICETCE)*, IEEE, pages 1–8.
- Punchihewa, A. and Bailey, D. (2002). Artefacts in image and video systems : classification and mitigation. In *Proceedings of International Conference on Image and Vision Computing*, pages 197–202.
- Rahul, K. and Tiwari, A. (2019). FQI : feature-based reduced-reference image quality assessment method for screen content images. *IET Image Processing*, 13 :1170–80.
- Rehman, A. and Wang, Z. (2012). Reduced-reference image quality assessment by structural similarity estimation. *IEEE Transactions on Image Processing*, 21 :3378–89.

- Reisenhofer, R., Bosse, S., Kutyniok, G., and Wiegand, T. (2018). A haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing : Image Communication*, 61 :33–43.
- Resende, M. and Ribeiro, C. (2019). Greedy randomized adaptive search procedures : advances and extensions. In *Handbook of metaheuristics*, pages 169–220. Springer.
- Saad, M., Bovik, A., and Charrier, C. (2012). Blind image quality assessment : A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing*, 21 :3339–52.
- Searson (2015). *GPTIPS 2 : an open-source software platform for symbolic data mining. Handbook of genetic programming applications*. Springer.
- Searson, D., Willis, M., and Montague, G. (2007). Co-evolution of non-linear PLS model components. *J Chemometrics*, 21 :592–603.
- Seshadrinathan, K. and Bovik, A. (2007). A structural similarity metric for video based on motion models. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 1, pages –869– –872. IEEE.
- Seshadrinathan, K. and Bovik, A. (2009). Motion tuned spatio-temporal quality assessment of natural videos. *IEEE Transactions on Image Processing*, 19 :335–50.
- Seshadrinathan, K., Soundararajan, R., Bovik, A., and Cormack, L. (2010). Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing*, 19 :1427–41.
- Sheikh, H. and Bovik, A. (2006). Image information and visual quality. *IEEE Transactions on Image Processing*, 15 :430–44.
- Sheikh, H., Bovik, A., and Veciana, G. (2005). An information fidelity criterion for image quality assessment using natural scene statistics. *IEEE Transactions on Image Processing*, 14 :2117–28.
- Sheikh, H., Sabir, M., and Bovik, A. (2006). A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, 15 :3440–51.
- Shi, C. and Lin, Y. (2020). Full reference image quality assessment based on visual salience with color appearance and gradient similarity. *IEEE Access*, 8 :97310–20.
- Silberholz, J., Golden, B., Gupta, S., and Wang, X. (2019). *Computational comparison of metaheuristics. Handbook of metaheuristics*. Springer.

- Silver, E. (2004). An overview of heuristic solution methods. *Journal of the Operational Research Society*, 55 :936–56.
- Simone, D., M, T., M, N., S, T., and H, E. T. (2010). 264/AVC video database for the evaluation of quality metrics. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing, IEEE*, pages 2430–3.
- Sinno, Z. and Bovik, A. (2018). Large-scale study of perceptual video quality. *IEEE Transactions on Image Processing*, 28 :612–27.
- Sotoudeh-Anvari, A. and Hafezalkotob, A. (2018). A bibliography of metaheuristics-review from 2009 to 2015. *International Journal Of Knowledge-Based And Intelligent Engineering Systems*, 22 :83–95.
- Soundararajan, R. and Bovik, A. (2011). RRED indices : Reduced reference entropic differencing for image quality assessment. *IEEE Transactions on Image Processing*, 21 :517–26.
- Sun, W., Zhou, F., and Liao, Q. (2017). MDID : A multiply distorted image database for image quality assessment. *Pattern Recognition*, 61 :153–68.
- Talbi, E.-G. (2009). *Metaheuristics : from design to implementation*, volume 74. John Wiley & Sons.
- Temel, D. and AlRegib, G. (2015). PerSIM : Multi-resolution image quality assessment in the perceptually uniform color domain. In *2015 IEEE International Conference on Image Processing (ICIP)*, pages 1682–6. IEEE.
- Temel, D., Prabhushankar, M., and UNIQUE, A. G. (2016). Unsupervised image quality estimation. *IEEE Signal Processing Letters*, 23 :1414–8.
- Tu, Z., Wang, Y., Birkbeck, N., Adsumilli, B., and Bovik, A. C. (2021). Ugc-vqa : Benchmarking blind video quality assessment for user generated contentz. *IEEE Transactions on Image Processing*, 30 :4449–4464.
- Uhrina, M., Holesova, A., Bienik, J., and Sevcik, L. (2021). Impact of scene content on high resolution video quality. *Sensors*, 21(2872).
- Varga, D. (2019). Empirical evaluation of full-reference image quality metrics on MDID database.
- Varga, D. (2020). A combined full-reference image quality assessment method based on convolutional activation maps. *Algorithms*, 13(313).

- Vu, P. and Chandler, D. (2014). ViS3 : An algorithm for video quality assessment via analysis of spatial and spatiotemporal slices. *Journal of Electronic Imaging*, 23(13016).
- Wang, S., Gu, K., Zeng, K., Wang, Z., and Lin, W. (2018). Objective quality assessment and perceptual compression of screen content images. *IEEE Comput Grap Appl*, 38 :47–58.
- Wang, S., Ma, L., Fang, Y., Lin, W., Ma, S., and Gao, W. (2016). Just noticeable difference estimation for screen content images. *IEEE TRANSACTIONS ON IMAGE PROCESSING*, 25(14).
- Wang, Y., Ke, J., Talebi, H., Yim, J., Birkbeck, N., and Adsumilli, B. (2021). Rich features for perceptual quality assessment of UGC videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13435–44.
- Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment : From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13 :600–12.
- Wang, Z. and Li, Q. (2011). Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 20 :1185–98.
- Wang, Z., Simoncelli, E., and Bovik, A. (2003). Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, pages 1398–402. IEEE.
- Wang, Z., Wu, G., HR, S., EP, S., Yang, E.-H., and AC, B. (2006). Quality-aware images. *IEEE Trans on Image Process*, 15 :1680–9.
- Winkler, S. (2005). *Digital Video Quality : Vision Models and Metrics*. John Wiley & Sons.
- Wong, M. and Leung, K.-S. (1995). Inducing logic programs with genetic algorithms : The genetic logic programming system. *IEEE Expert*, 10 :68–76.
- Wu, J., Lin, W., Shi, G., and Liu, A. (2013). Perceptual quality metric with internal generative mechanism. *IEEE Transactions on Image Processing*, 22 :43–54.
- Xia, Z., Gu, K., Wang, S., Liu, H., and Kwong, S. (2020). Toward accurate quality estimation of screen content pictures with very sparse reference information. *IEEE Transactions on Industrial Electronics*, 67 :2251–61.
- Xu, M., Chen, J., Wang, H., Liu, S., Li, G., and Bai, Z. (2020a). C3dvqa : Full-reference video quality assessment with 3d convolutional neural network. In *ICASSP 2020-2020*

- IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, pages 4447–51.
- Xu, M., Li, C., Zhang, S., and Le Callet, P. (2020b). State-of-the-art in 360 video/image processing : Perception, assessment and compression. *IEEE Journal of Selected Topics in Signal Processing*, 14(1) :5–26.
- Yamsang, N. and Udomhunsakul, S. (2009). Image quality scale (IQS) for compressed images quality measurement. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 1, pages 789–94.
- Yang, G., Li, D., Lu, F., Liao, Y., and Yang, W. (2018). RVSIM : a feature similarity method for full-reference image quality assessment. *EURASIP Journal on Image and Video Processing*, 2018.
- Yang, H., Fang, Y., and Lin, W. (2015). Perceptual quality assessment of screen content images. *IEEE Trans on Image Process*, 24 :4408–21.
- Yang, J., Bian, Z., Zhao, Y., Lu, W., and Gao, X. (2021). Full-reference quality assessment for screen content images based on the concept of global-guidance and local-adjustment. *IEEE Trans on Broadcast*, 67 :696–709.
- Ye, M., Wang, X., and Xu, Y. (2009). Parameter extraction of solar cells using particle swarm optimization. *Journal of Applied Physics*, 105(94502).
- Zavala, G., Nebro, A., Luna, F., and Coello Coello, C. (2014). A survey of multi-objective metaheuristics applied to structural optimization. *Structural and Multidisciplinary Optimization*, 49 :537–58.
- Zhai, G. and Min, X. (2020). Perceptual image quality assessment : a survey. *Sci China Inf Sci*, 63(211301).
- Zhang, F., Li, S., Ma, L., Wong, Y., and Ngan, K. (2011a). *IVP subjective quality video database*. The Chinese University of Hong Kong, [Http ://Ivp Ee Cuhk Edu Hk/Research/Database/Subjective](http://Ivp.Ee.Cuhk.Edu.Hk/Research/Database/Subjective).
- Zhang, L. and Li, H. (2012). SR-SIM : A fast and high performance IQA index based on spectral residual. In *2012 19th IEEE International Conference on Image Processing*, pages 1473–6. IEEE.
- Zhang, L., Shen, Y., and Li, H. (2014). VSI : A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, 23 :4270–81.
- Zhang, L., Zhang, L., Mou, X., and D, Z. (2011b). FSIM : A feature similarity index for image quality assessment. *IEEE Transactions on Image Processing*, 20 :2378–86.

- Zhang, X. and Hou, J. (2019). Quality assessment towards cell diffraction image based on multi-channel feature fusion. *Journal of Visual Communication and Image Representation*, 64(102632).
- Zhang, Y., Wang, Y., Liu, F., Liu, Z., Li, Y., and Yang, D. (2018). Subjective panoramic video quality assessment database for coding applications. *IEEE Trans on Broadcast*, 64 :461–73.
- Černý, V. (1985). Thermodynamical approach to the traveling salesman problem : An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, 45 :41–51.